Institutions, Roles, and Agency

Dissertation

zur Erlangung des akademischen Grades

Dr. phil.

eingereicht

am Institut für Philosophie und Politikwissenschaft,
Fakultät Humanwissenschaften und Theologie
der Technischen Universität Dortmund

von Max Magnus Gab, M.A.
geboren am 10.01.1994 in Achern

Abstract:

This thesis provides an analysis of institutional group agency, i.e. the capacity of institutional groups to perform actions. I explore different accounts of institutional group agency and consequently argue that it is best explained by so called role-based accounts. According to such role-based accounts of institutional agency, the actions of institutional groups consist of the contributory actions of the groups' members, who act according to the tasks and functions of their institutional roles. I will, however, argue that such role-based theories of institutional action face certain problems. Most importantly, they under-theorize the relation between institutional roles and the individuals who occupy them. In order to illuminate this relation, I develop my own, novel account of *Role Agency*. Role Agency describes a form of agency that individuals engage in when they act within their institutional roles. It is best thought of as a form of *perspective taking*, which includes both the *internalization* and *idealization* of institutional roles.

Acknowledgements:

It would be an odd result to write a whole thesis about our human capacity to cooperate, only to end up believing that this was something that I did on my own. Therefore, I want to thank the many people who contributed to this thesis. First and foremost, I want to thank my supervisor Katja Crone. Throughout the years, Katja has been a true role-model as well as an invaluable source of guidance and encouragement. I am deeply grateful for her enduring and intensive help, and for providing both the intellectual freedom and rigor that made this thesis possible. My deep gratitude also goes to Eva Schmidt, who took not only immense effort as the second supervisor of this thesis, but also helped to shape it to its final form. I also want to thank Peter Königs for his great effort concerning my defence.

Next, I want to thank all the participants of the Philosophical Colloquia at TU Dortmund, as well the engaging audiences at conferences in Berlin, Stockholm and Leeds. I want to thank the Graduate School on Political Cohesion (GSPC) for providing traveling funds. And of course, I am grateful for my dear colleagues at the Institute for Philosophy and Political Science, who helped me through the ups and downs of this project. Especially, I want to thank Nora Olbrisch, Stefano Vincini, Dick Timmer and Martina Herrmann for their academic guidance. I want to thank Daniel Beck, Niklas Dummer, and Nora Olbrisch for providing substantive feedback for my thesis, which contributed vitally to its refinement. I also want to thank all my friends and family, and especially my parents, Gisela and Thomas, as well as my sister Franziska. Finally, I want to thank my grandparents, Werner and Maria. So much, including my remaining sanity, is owed to Pauline (and our dog Lala). A mere "Thank You" could not hope to do justice to her support.

Im planmäßigen Zusammenwirken mit andern streift der Arbeiter seine individuellen Schranken ab und entwickelt sein Gattungsvermögen.

Karl Marx: Das Kapital. Band I

Preface

I

Humans began to do philosophy, Aristotle said, "at first because they wondered about the strange things right in front of them, and then later, advancing little by little, because they came to find greater things puzzling" (*Met.* 982b12). Right in front of me, as I write this introduction, is a #2 pencil. At first glance, this pencil seems to be a rather ordinary, mundane object. You can buy such pencils in the store, grab a handful of them at IKEA, or steal one from your coworker's desk. Sometimes you lose them, and sometimes they break. They also come in different colors. Otherwise, there aren't that many interesting things to say, nor to wonder about #2 pencils.

But consider that one day, for some reason, you were given the task to produce a #2 pencil on your own. This, on the other hand, would be a lot more interesting story to tell. It surely would take a lot of work. You would have to cut down a tree, saw its wood into pieces, somehow manage to locate, mine and process graphite, build a machine to mill-cut the wood into hexagonal shape, drill a hole in it, etc. It would probably take years of work for you to produce just one, ordinary #2 pencil. Some say that you wouldn't be able to produce one at all.[1]

When we're cooperating, however, human individuals excel at producing pencils. In fact, we're so good at producing them that we've grown accustomed to their existence and don't think much about them whatsoever. The ways in which we cooperate and the things we accomplish this way, like producing #2 pencils, come so naturally to us that we tend to forget about them. My thesis, in part, is about the ways in which our capacity to cooperate with one another empowers us to do great things. Things which, alone by ourselves, we would not be able to do. And it's a reminder that even the small and mundane things in our lives, like #2 pencils, are actually great achievements that stem out of our capacity to cooperate with one another. This, I think, is the strange thing right in front of us, and here I come to find greater things puzzling.

II

Our capacity to cooperate with one another shapes our human life in myriad ways. And one of the ways in which our lives are shaped by this capacity is through the existence of institutions. But institutions, or institutional groups, do not only shape the small and mundane aspects of our lives. Like the control of fire,

[1] In his essay "I, Pencil: My Family Tree", Leonard Read (1958) gives a more detailed explanation of the necessary steps involved in order to produce a pencil. He states that "millions of human beings have had a hand in [its] creation, no one of whom even knows more than a very few of the others" (Read 1958, 9) and that "[t]here isn't a single person in all these millions, including the president of the pencil company, who contributes more than a tiny, infinitesimal bit of know-how" (ibid) in producing a pencil. Read suggests that "not a single person on the face of this earth" (Read 1958, 7) knows how to make a pencil on her own. The lesson to be drawn here, according to Read, speaks in favor of libertarian, free-market capitalism. I do not wish to argue that this lesson is to be drawn from our capacity to produce pencils. But I agree with Read that the division of labour is a quite astonishing aspect of our modern lives.

agriculture, or writing systems, institutional groups are a technology that humans invented which ultimately changed their historical trajectory.

As such, institutional groups come with great power. The establishment of modern science, medicine, education, industrial production, and cultural innovation all are connected, in some form or the other, to institutional groups. Institutional groups put mankind in space and on the moon. They developed the vaccines that helped to end the pandemic. The computer, the internet, airplanes, and railways all exist in virtue of institutional groups. But the great power of institutions is, in and of itself, not neutral. Institutional groups have wielded their power in both the most positive, but also in the most devastating forms. They played their part in the most tragical and abhorrent moments of modern life. The second World War, the Shoa, and dropping the first atomic bomb reveal not only what humans can do to each other, but they also bare witness to what *institutions* are capable of. In a perverted, yet important sense, these tragical events are the product of our human capacity to cooperate.[2]

III

It is therefore crucial to consider this capacity with great care. A theory of institutional group agency should do justice to the fact that institutions exert immense influence, both positive and negative, over our modern lives. It should carefully consider the relations that hold between the actions of individuals and the actions of institutional groups. For the way we think about the actions of institutional groups has implications for how to understand attributions of control, rationality, culpability and ultimately moral responsibility to such groups. Some argue that institutional groups should be viewed as moral agents in their own right, and on this basis to be the subject of moral responsibility. In cases of harmful, and morally apprehensible actions of institutional groups, the subject of moral responsibility should be the group itself, in addition to, or even instead of its members. Others argue that groups are never the proper agents of such moral assessment. According to some, to say that an institutional group is morally responsible for a wrongful act is only an "elliptical (and somewhat dangerous) way of saying that certain human individuals are morally responsible for that act" (Velasquez 1983, 1). Although I will not address these normative questions directly, my thesis is not just a theoretical exercise. Answering questions concerning the moral responsibility of institutions asks for a deepened understanding of institutional actions, collective agency and institutional groups. And I hold my thesis to be an attempt of providing such an understanding.

[2] See for organizational analyses of the Holocaust focussing on the functioning of the involved institutions: Kogon 2006; Kühl 2014; Sofsky 1996.

Table of Contents

1. Introduction

In the broadest possible terms, my thesis is about our capacity to cooperate with one another; our capacity to *do things together.* Our human capacity for cooperation is unmatched in both its scope and ubiquity. And it is also a heterogenous and multi-facetted phenomenon. It can range from two individuals dancing a *Pas de deux,* or a group of friends going on a camping trip, to kitchen brigades preparing meals, or lawyers sitting in their offices, filing papers on behalf of their corporation. So let me be a bit more precise as to what exactly my thesis is about.

In my thesis, I want to explain the agency of *institutional groups*, i.e. the capacity of institutional groups to perform actions. Cases of such institutional group action include, e.g., a trade-union negotiating a collective bargaining agreement, a corporation declaring bankruptcy, a court of law sentencing a defendant, or a military unit fighting in a battle. Talk about such institutional group actions pervades our daily lives. Pick up any newspaper, or turn on any radio and you will read and hear about the triumphs, struggles and misdemeanors of institutional groups.

Cases of institutional group action allow for interesting questions that have occupied philosophers for quite some time: Can we rightfully attribute such actions to the groups themselves? Are institutional groups *agents*? If so, by which way (or ways) do individuals come to constitute such agents? Is there a single mechanism that allows groups to bear the capacity for agency? Or can the phenomenon come about in multiple, different ways? And if groups *are* agents, are they *minded*? What other capacities do institutional groups possess? Or should we instead think of such actions as simply the "sum" of individual actions? But if that was true: how are such actions "summed up" then, i.e. how are the actions of individual members related to the actions of institutional groups?

Within the field of research that grapples with such questions, my thesis will focus on what I call *role-based* explanations of institutional group agency. By role-based explanations of institutional agency, I refer to those theories which explain the agency of institutional groups by pointing to the individual members' capacity to perform the tasks and functions of their institutional roles. To say that an institutional group acts, according to such theories, is to say that the individual members carry out the tasks and functions of their institutional roles. A group agent, to paraphrase Kirk Ludwig, is a *group of agents* (cf. Ludwig 2017a, 296-298). Ultimately, my thesis wants to *advance on* and *supplement* such role-based explanations of institutional group agency by providing an account of what I christened *Role Agency*. Part of my account of Role Agency is motivated by the fact that such role-based explanations of institutional agency *under-theorize* the notion of what it means to *act qua one's institutional role*. My goal here is to identify, and consequently fill this lacuna in the existing literature. And because this relation between an individual agent and her institutional role is neglected in the literature, I aim to explore what it consists of, and how we could further analyze it. Role Agency captures a form of agency that individuals engage in when they act in their institutional roles. I will argue that it should be thought of as a form of *perspective taking*, which includes both the *internalization* and *idealization* of institutional roles. My account of Role Agency should be understood as an attempt to explain the ways in which individual agents connect to the institutional roles they come to occupy. I hold my account to offer a deepened understanding about the way in which

individuals exercise influence over their institutional roles, and how their institutional roles, in turn, influence them.

By writing these introductory remarks alone, I have already mentioned a lot of complicated, and contested concepts. So I will continue by trying to clarify them first. I'll do so by asking some basic, consecutive questions through which we can progressively advance in our endeavor: What are *actions*? What are *collective* actions? What are *social groups* to begin with? What are *institutional* groups? How could we understand claims about (institutional) *group agency*? Answering these questions will provide the conceptual clarity we need in order to move on. It will also introduce the reader to the relevant fields of research that grapple with these questions. I want to advise the reader that the following sections constitute a *tour de force* through a vast amount of literature and that a comprehensive analysis of the depicted concepts would be beyond the scope of this thesis. Instead, I will only try to clarify some of the basic ideas at play, which are subject to a controversial and ongoing discussion. So let's start with some basic clarifications and initial definitions.

Actions, agents and agency

I will only try to give *minimalistic* definitions of the concepts of action, agency, and intentional states (see: Box 1). For one, this is because I do not want to get sidetracked in debates which would lie way beyond the scope of my thesis.[1] I am therefore not going to give a comprehensive discussion of the existing literature on the philosophy of action, a topic that I could not hope to do justice to. The second, and more pressing reason I only aim to provide a minimal definition of these concepts is twofold: First, I want to explore the agency of *institutional groups*. Institutional groups, however, are not the paradigmatic examples by which contemporary action theory (and the philosophy of mind more generally) operates. The classical concepts of "action", "agent", "agency", or "intentional states" were developed in light of *individuals*, and not in terms of social groups, or institutional groups. This is not to say that we cannot use such theories to study institutional groups. But it runs risk of relying too heavily on concepts that are inapplicable to groups from

[1] The *Philosophy of Action* cuts through a whole bunch of questions and problems that puzzle philosophers since centuries, e.g., metaphysical questions regarding causal laws, determinism and free will, the problem of mental causation, or the question of how mind and body are related to one another. Then there are linguistic questions regarding the status of speech in action, but also practical philosophical questions, e.g., under which circumstances one could be *responsible* for one's actions, what makes actions morally significant, or how the concept of agency relates to our concept of (human) *persons,* etc. Then there are also questions regarding our political and social *practices*, e.g., whether criminal law tracks the right way to attribute culpability, etc. For an overview I recommend: Piñeros Glasscock & Tenenbaum's (2023) SEP-entry on *Action*, as well as Paul's "*Philosophy of Action: A Contemporary Introduction*" (2020).

the get-go.[2] Second, I will present different accounts of *institutional group agency* during the course of this thesis. These accounts argue for different views of what it means to be an agent, and they provide different arguments for the claim that an institutional group has agency, i.e., the capacity to perform actions. I want those accounts to have their fair trial, and so I will treat lightly on these concepts.

Minimally, agency refers to the capacity to perform actions, i.e., to bring about change in one's environment on the basis of intentional states such as beliefs, intentions, and desires. A preliminary way to understand this claim is to compare (or juxtapose) *actions* and *events.*[3] As Tollefsen handily explains, there is a difference between one's tripping (an event) and one's performing a prat fall (an action) (cf. Tollefsen 2015, 27). Events are usually characterized as things that passively *happen*, whereas actions are something which are *actively done*, or *done intentionally*.[4] In contemporary action theory, the work of Donald Davidson (2001a-h) has provided us with a concept of agency that draws on this distinction between intentional actions and events, and tries to explain what this difference amounts to. Both proponents and opposing theorists are inclined to call Davidson's theory the "standard view" of intentional action and agency (see, e.g., Aguilar & Buckareff 2010; Mele 2003; Chant 2017 Sec. 1-2; Paul 2020). This "standard view" of agency is embedded within the *causal theory of action* (CTA). The CTA should not be understood as a unified theory, but rather as a family of resembling theories. To summarize, this "standard view" ultimately subscribes to the idea that intentional *actions are events that have appropriate mental items such as beliefs, desires, and intentions as their causes.* According to CTA, an event is an intentional action in virtue of its causal connection to certain mental states (see also: Piñeros Glasscock & Tenenbaum 2023; Davis 2010). Here's the definition of the causal theory of action provided by Aguilar and Buckareff (2010, 1):

> **CTA:** Any behavioral event *A* of an agent *S* is an action if and only if *S*'s *A*-ing is caused in the right way and causally explained by some appropriate nonactional mental item(s) that mediate or constitute *S*'s reasons for *A*-ing.

---

[2] Take the example of explicating the nature of mental states. Are functionalist, behavioral, or computational theories of mental states the proper way to analyze them? What, if anything, realizes mental states? Are they realized in individual human brains? Can only individual biological animals with sophisticated neural systems realize them? Can they extend to the outside of the skull of such animals? Are they *realized* by brain states, or are they to be equated with brain states altogether? Now, by settling on any preliminary understanding and a stipulative definition of the concept of *mental states*, the possibility of ruling them out to be realizable by a group arises. If, for example, my working definition of mental states sees them as to be necessarily "caused by neurobiological processes in the brain and […] realized in the brain, […] the neuronal processes themselves [being] manifestations of and dependent on even more fundamental processes at the molecular, atomic and subatomic levels" (Searle 1995, 4), the possibility of groups to realize mental states is jeopardized right from the beginning. Why? Because if mental states are realized only within individual biological brains, then groups will not possess mental states, simply because groups do not have individual biological brains.

[3] Ludwig analyses events as changes, and changes as "higher-order property instantiations, the having or losing of a property by an object" (2017a, 20). See also: Casati & Varzi 2020; Lombard 1986.

[4] As to avoid confusion, I will proceed to use the terms "actions" and "intentional actions" interchangeably unless otherwise indicated.

This view, according to the authors, is the "nearest approximation in the field to a theoretical orthodoxy" (Aguilar & Buckareff 2010).[5] According to this minimal definition of action, the term "*agency"* describes the *capacity* to *bring about, to produce*, or *to perform actions.* Hence, I take the term "agent" to - minimally - refer to an (or any) entity that possess the capacity for agency.

On such a minimalistic account of agency, and according to the so called "standard view", explaining intentional action invokes the central concept of *intentional states*. Both the standard view on agency, the *causal theory of action* (CTA) and its main competitors, the *agent causal theory of action* (ACTA) and the *volitionist theory of action* (VTA) rely on some notion of *intentional states* (see, e.g., Aguilar & Buckareff 2010; Paul 2020). Often claimed to be the *hallmark of the mental* (cf. Jacob 2019) and, as mentioned, crucial for our understanding of action and agency, I will try to give a brief analysis of the terms "intentionality" and "intentional states".

Being a philosophical shibboleth, the term "intentional states" refers to those mental states that have the property of *intentionality*.[6] In turn, I take the term "intentionality" to refer to the specific feature of mental states to be *about* something, i.e. about (fictional or non-fictional) objects, properties, or states of affairs (see: Brentano 2009; Jacob 2019; Searle 1983, 1992, 1994; Perry 1994; Pettit 1996, Ch.1; Tollefsen 2015, Ch. 2; Jankovic & Ludwig 2017b). So minimally, intentionality concerns the *directed*- or *aboutness* of mental states. Not all mental states have this feature and so not all mental states are *intentional states*. Nevertheless, intentionality is involved in many *different* kinds of mental states, including propositional attitudes (e.g., *believing*, *desiring*, *intending*, *hoping*, *wishing*, *doubting*, etc.), perceptional states (e.g. *seeing* a pigeon on one's balcony), or emotions (e.g. *being happy* about the pigeon flying away).

Paradigmatic examples of intentional states, i.e., propositional attitudes such as *believing*, *desiring*, or *intending* reveal the "aboutness" of intentional states. But they also allow us to learn something about their structure: Take my belief that it is sunny outside. My belief *is about something*, i.e., it has as its *content* that it's sunny outside. Apart from having content, it also is a state attributable to a *subject*, i.e. it is *me* having this believe and not, e.g., my grandmother. Further, we can specify the *mode* of this intentional state by contrasting it, on a rainy day, to my *desire* for it to be sunny outside. The conditions under which my *belief* that it is sunny outside would be satisfied are different from the conditions under which my *desire* that it is sunny outside would be satisfied. As for my belief to be satisfied, my state of mind would have to "fit" the state of the world. For my desire to be satisfied, it would be the other way around. So a *subject* can represent one and the same *content* in different *modes*. These modes, in turn, can be individuated by specifying their *conditions of satisfaction* (see Searle 1983).

---

[5] My presentation of the CTA as the "standard" theory should be taken with a grain of salt and primarily reflects the focus of the subfield of the philosophy of collective action. According to some, the CTA has lost this status in recent years and current trends tend to see action as a manifestation of competences to act. I want to thank Eva Schmidt for pointing this out to me.

[6] Intentionality, in turn, is not to be confounded with the above mentioned intentions: While intentions can be said to be intentional states, not all intentional states are intentions (as there are lots more). To navigate around yet another similarly sounding term, intentionality is not to be equated with intensionality (or as some call it "intensionality-with-an-s" Searle 1994, 385) which describes the phenomenon of sentences satisfying (or failing to satisfy) certain semantical tests for extensionality. While more could be said about the connection between those two concepts, we can summarize - rather bluntly - that intentionality is about mental states, and intensionality is about reports of those states.

ACTIONS: Actions are events that have appropriate mental items such as beliefs, desires, and intentions as their causes.

AGENCY: Agency is the capacity to act, to bring about change in one's environment on the basis of states such as belief, intention, and desire

INTENTIONALITY: intentionality is the feature of mental states to be about, or to represent something, e.g., things, properties and states of affairs.

INTENTIONAL STATES: Intentional states are those mental states that have the feature of intentionality, i.e., the feature of being about, or of representing something, e.g., things, properties and states of affairs.

Box 1: minimal definitions of actions, agency, intentionality and intentional states

Collective action

Up to this point, I provided only a minimalistic definition of the concepts of agents and agency. Rather than engaging in an attempt to further specify (perhaps necessary or sufficient) conditions for agency, I invite the reader, at least for now, to follow Deborah Tollefsen's approach and view agency as occurring on a *spectrum* "with very simple agents at one end and very complex agents at the other" (Tollefsen 2015, 53).

Now many of the things that agents do involve other agents. Driving on a busy road, or visiting a concert is something that an individual agent does which involves other agents. Some, but not all of the actions that involve other agents are *collective actions.* What do I mean by *collective,* or *joint actions*?[7]

A first way to demarcate the phenomenon of *collective action* is to distinguish it from the broader category of *social action*. Social actions are actions that "conceptually presuppose the existence of other agents and of various social institutions" (Tuomela & Miller 1988, 369; also: Miller 2001, Ch.1.). Miller elaborates that "most human action is in fact at some level, or to some extent, or in some sense, social action. Even these actions of eating, drinking, eating ice cream, individually walking down the road, or having a shower typically *presuppose* social forms or objects, such as farms, ice cream parlors, cups, roads, and shower rooms" (Miller 2001, 1). All collective actions are social actions in that they presuppose the existence of *other agents.* The opposite, however, does not hold: Not all social actions are collective actions, i.e., actions performed in *cooperation with* other agents. One time, someone stole my phone. This was a social action performed by the thief, i.e., it was an action that was depending on the existence of someone to steal the

[7] A problem in the existing literature is that the terms used are not used consistently, and that different terms are used by different scholars to describe the same phenomenon in question. What I will call "collective action" has been called "shared cooperative activity" (Bratman 1992), "joint action" (Gilbert 1990), "jointly social action" (Miller & Tuomela 1988), or "intentional joint agency" (Pacherie 2013) by others. This makes it rather difficult to see the similarities (and differences) of these terms. I will try to be as consistent with the use of my own definitions of the term as possible.

phone from. But it would be false to say that this was something that the thief and I did *together*. Luckily, my dear colleague Niklas jumped into action and we managed to catch the thief later on. This, in turn, was something that Niklas and I did *together*. So how did *we* do it?

Recall my stipulative definition of actions: they are *events that have mental items such as beliefs, desires, and intentions as their causes*. Naturally, *collective action* could therefore simply describe events that have mental items such as beliefs, desires, and intentions *of more than one individual* as their causes. But this does not adequately capture the targeted phenomenon. An event can be caused by more than one individual, where it would be false to say that these individuals *acted together*, or that there was a *collective, or shared intentional action.*

Here's an example of what I mean by that: After the holidays, there was a big pile of Christmas trees on the corner of the street where I live. It started with one, and by the end of the week, all the neighbors had thrown their Christmas trees onto the other trees, thereby creating the pile on the corner. None of my neighbors produced this pile of Christmas trees *individually*, i.e. none of them produced this pile on their own. But although all my neighbors were involved in producing the pile, they did not produce it *together*. My neighbors didn't wake up one morning and said "Hey fellow neighbors, let's all form a big pile of Christmas trees, so that Max will know that the holidays are over!". They surely could have done so. But this is not what happened. Rather, the pile was the *aggregated result of their parallel individual actions*.[8]

To introduce yet another distinction, consider the ambiguity of sentences like "Anne and Betty sang the national anthem". On a *distributive* reading, this simply means that Anne sang the national anthem and that Betty sang the national anthem. The distributive reading does not imply that Anne and Betty *together* sang the national anthem. The stronger, non-distributive reading of "Anne and Betty sang the national anthem" then suggests that there was *one* event that both Anne and Betty were the agents of, i.e., their singing of the national anthem together (see especially: Ludwig 2017a).

So how can we explicate the difference between cases of *parallel individual action* on the one hand, and cases of *collective,* or *joint action* on the other? One way to analyze collective action is proposed by Chant, who called the way of applying theories of *individual* action to collective cases the "wash, rinse and repeat" (Chant 2017, 13) approach to collective action theory:

> "take the very best accounts of individual action and agency and simply ‚collectivize' them. Since the very best accounts of individual action theory depend on an agent's intentions and other mental states, when ‚collectivized', the result is a collective action theory that depends on *collective* agents that have *collective* intentions and other *collective* mental states" (ibid).

[8] As mentioned, I use the terms "actions" and "intentional actions" interchangeably. However, some authors (e.g. Ludwig 2007) would describe this example as a case of collective action, but not as a case of collective *intentional* action. Roth (2017) explains that one example of such a distributed collective action, which is not a collective *intentional* action "would be our jointly bringing about some severe environmental damage. This might come about as a side effect of each of us pursuing our own projects. No subject intends the severe environmental damage, under any description: no single individual has enough of an impact to intend anything that would count as severe environmental damage, and as a collective the polluters seem not to be sufficiently integrated to count as a subject of intention" (Roth 2017; see also: Chant 2007).

This is where the standard story (e.g., Schmid & Schweikard 2009; Schweikard & Schmid 2021; Searle 1990; Bratman 1992; Roth 2017; Pettit & Schweikard 2006; Pettit 2018) of *collective intentionality* begins to tell its tale. Perhaps the most prominently discussed example that is supposed to illuminate this difference is provided by Searle (cf. Searle 1990, 402). Searle asks us to imagine a group of individuals scattered in a park. Suddenly rain sets in, and each individual runs towards a nearby shelter. In this scenario, their running towards this shelter is not something that they do *together*. Rather, they all *individually* seek refuge from the rain. The second scenario has the same individuals behaving in the exact same way, i.e., they too all run towards the nearby shelter. But in this case, they are all part of a *corps du ballet* and they *together perform* a (somewhat weird) dance choreography. Because there is no observable difference in the outward behavior of the individuals regarding these two scenarios, Searle suggests that the difference between the two scenarios must be explained by pointing to facts which are *internal* to the individuals. So what makes their action *collective* is thought to be an *internal difference*.

The analysis of this internal difference constitutes "the central focus of the study of collective intentionality" (Jankovic & Ludwig 2017b, 1). The standard attempt to explain what this difference amounts to is to analyze the internal states of the individuals involved. To be a bit more precise: What scholars aim to analyze is the *structure of their intentional states*. And as mentioned above, intentional states can be analyzed in terms of their three components, i.e., their *mode,* their *content*, and them being attributable to a *subject.* Accordingly, the question as to what makes collective intentional states *collective* is answered by pointing to (one of) these three components. The results are the so called mode-, content- or subject-accounts of *collective intentional states*. Here are Schweikard & Schmid (2021) explaining the different approaches by analyzing the example of two individuals *visiting the Taj Mahal together*:

> "Content-accounts claim that for *A* and *B*'s intention to visit the Taj Mahal tomorrow to be collective, each *A* and *B* have to intend to visit the Taj Mahal *together.* Mode-accounts insist that the element of collectivity has to extend to the intending; in their view, *A* and *B* have to intend *collectively* to visit the Taj Mahal (together). Subject-accounts hold that the element of togetherness is really in the subject; in their view, *A* and *B* have to form a plural subject or a unified group that is the subject of—and has—the intention to visit the Taj Mahal" (Schweikard & Schmid 2021).

So depending on their school of thought, scholars argue that the distinctive mark of *collective intentionality* can be either explained by a special *mode* that multiple individuals engage in when performing a collective action. Or they argue that what's *collective* about collective intentional states is that the intentional states of the individuals involved have a shared, or collective *content*. Yet others argue that what's collective about collective intentional states is that they are attributable to a special collective *subject*. Now before continuing with more in-depth analyses of the mode-, subject-, and content-accounts of collective intentionality, I want to advance on our understanding of collective intentionality by further clarifying two of its features:

First, collective intentionality is an umbrella-term, under which different phenomena are analyzed. While some take collective, or joint *intentions* to be the paradigm of collective intentional states (cf. Schweikard &

Schmid 2021), theories of collective intentionality do not exclusively aim to analyze collective *intentions* and joint *actions.* They also try to explicate the nature of collective intentional states such as *joint attention*, collective *memories*, collective *emotions*, or collective *experiences* (for an overview see: Jankovic & Ludwig 2017a: Part II). These latter forms of collective intentional states will not be in the focus of my thesis, and I will not discuss them in depth.

Second, the distinction between mode-, content-, and subject-accounts of collective intentionality is not the only relevant way to divide the debate. The reductive / non-reductive divide also figures prominently. What does this divide amount to? Basically, scholars disagree on whether collective intentionality is a *genuine* phenomenon, or whether it can (or cannot) be reduced to individual intentionality. Saying that *collective intentionality* reduces to *individual intentionality* is meant to imply that collective intentionality is *nothing more than* individual intentionality*,* or that it's *nothing over and above* individual intentionality. Other ways to express this idea in the vocabulary of reduction is that collective intentionality *fully depends upon,* or that it is *constituted by* individual intentionality (see: van Riel & van Gulick 2024). So regarding collective intentions and collective actions, *reductive* views assert that they are to be best understood in terms of the properties and concepts that we already have available when we try to understand individual intentions and action. *Non-reductive* views, in turn, deny this assertion (cf. Alonso 2017, 34).

But to say that there's a conceptual distinction between *collective* and *individual* intentionality, or that the former can (or cannot) be reduced to the latter, invites for an ambiguous understanding: It could be understood that *individual intentionality* belongs to an *individual subject*, and that *collective intentional states* are the states of a *collective subject*. Authors then would argue whether the intentional states of such a *collective subject* can (or cannot) be reduced to the intentional states of the individual *subjects* involved. But this is not the only way to understand this distinction. Another way to understand this distinction is to focus not on the subject-component, but on the mode- and content-components.

The work of John Searle (especially: Searle 1990) figures as a prominent example of such a non-reductive view which focusses on the mode-component. According to Searle, what makes collective intentional states "collective" is that they are held is a special, collective *mode,* or attitude.[9] This collective mode is one in which an individual possesses so called *we-intentions*, which Searle contrasts with an individual possessing so called *I-intentions*. These we-intentions, according to Searle, are a biologically primitive phenomenon of the minds of individuals agents. So in a way, they are *individual intentions*. But critically, Searle claims that these *we-intentions* involve "the notion of cooperation" (Searle 1990, 406). According to Searle, this feature of cooperativeness cannot be reduced to individual *I-intentions* plus mutual beliefs (Searle 1990, 404-405). So we- and I-intentions are both the intentions of individuals, but the are intentions of a different *kind*. For

[9] Crone (2025) offers a critique of Searle's ambivalent use of the term *mode*, which she suggests suffers from an equivocation.

Searle, it is *not* the case that we-intentions are *nothing more than,* or *nothing over and above* individual *I-intentions* (ibid). The former cannot be reduced to the latter.[10]

Others, e.g., Bratman (1992, 1993, 2014), argue for a reductive view of collective intentions based not on a special mode, but based on the *contents* of collective intentional states. According to Bratman (e.g., 1992), a collective intention (or as he puts it: a *shared* intention) is reducible to (or: constituted by) *nothing more* than a complex, interrelated structure of the *content* of the individuals' attitudes.[11] On this view, to say that there's a conceptual distinction between *collective* and *individual* intentionality, or that the former can (or cannot) be reduced to the latter, is not to say that *collective intentional states* belong to a *collective subject*, or that they are held in a special *mode*. Both collective intentional states *and* individual intentional states are held only by an individual subject and in the same, individual *mode* of, e.g., intending. Rather, what makes, e.g., collective intentions *collective* is that their *content* refers to a collective action which individuals, under the assumption of common knowledge, intend to perform in accordance with others and by meshing their individual subplans required for the performance of their collective action.

I will refrain from giving a more detailed analysis of collective intentionality and from defending one of the particular views in question. I eventually explore accounts of institutional group action that utilize the concept of collective intentionality (and collective intentions in particular) to explain the actions of institutional groups.

To this end, I will provide a critique of theories of collective intentions and collective intentional action, sometimes labeled as the "Big Four" (the label originates from: Chant, Hindriks & Preyer 2014). These accounts limit their analysis of collective intentional states to paradigmatic small-scale cases of collective action such as two individuals *taking a walk* (Gilbert 1990), *mixing a sauce hollandaise* (Searle 1990), *carrying a piano upstairs* (Miller & Tuomela 1988) or *painting a house* (Bratman 1992). I will try to argue that they are therefore inapt to be applied to explain the collective agency involved in large and complexly structured institutional groups.

---

[10] A handy way to express this thought is to schematize it in order to mark these different *kinds* of modes. For the individual case, I-intentions can be formalized like this: <I> <I-intend> <to visit the Taj Mahal>. In the collective case, we-intentions could be formalized like this: <I> <we-intend> <to visit the Taj Mahal>. The difference expressed in the <I-intend> and <we-intend> brackets marks the difference between Searle's I-intentions and we-intentions. On this view, to say that there's a conceptual distinction between *collective* and *individual* intentionality, or that the former can (or cannot) be reduced to the latter, is not to say that *individual intentionality* belongs to an *individual subject*, and that *collective intentional states* are the states of a *collective subject*. Both collective intentional states and *individual intentional states* are held only by an individual, schematized as the <I>.

[11] Again, a schematic depiction of this helps to understand what Bratman argues for. Bratman's analysis of shared intentions is this: We intend to *J* if and only if:

(1) (a) I intend that we *J* and (b) you intend that we *J*.

(2) I intend that we *J* in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b; you intend that we *J* in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b.

(3) 1 and 2 are common knowledge between us. (Bratman 1999: 121)

If we focus on the first clause and contrast it to Searle's we-intentions, the difference becomes apparent: What's collective about a *Bratmanian* shared, or collective intentional state is not found within the subject-bracket "<I>", or in the mode-bracket "<intend>", but rather in the *content*-bracket "<that *we J*>". Hence, collective intentional states can be explained by the complex, interrelated structure of the *content* of the individual's attitudes.

Let me highlight some abstract commonalities which all of these accounts adhere to. According to Fiebich, "although the various accounts provided in the literature differ in detail with respect to what intentions in a group come to" (Fiebich 2019, 164), "most philosophers"[12] (ibid) agree certain conditions necessary for there to be collective action. The common ground in the debate then seems to be that:

(i) any member of the group intends to do his or her part for there to be a shared intention (individual intention condition),

(ii) that any member of the group does so, because he or she expects the other members of the group to do the same (interdependence condition), and

(iii) that (i) and (ii) are common knowledge in the group (common knowledge condition) (Fiebich 2019, 164f.).

Again, I will provide a more detailed critique of these approaches and argue that theories of collective or joint action that focus on such small-scale cases of collective action are subject to the so called *Upscaling Problem*. Roughly, the *Upscaling Problem* (see Ch. 2.2.1. for a detailed version; see also Poljanšek 2015) argues that accounts of collective action, which take minimal forms of collective action to be paradigmatic, lead to implausible results when applied (or *scaled up*) to large and complex forms of collective action, i.e. cases of institutional group action. Another, related reason to refrain from a more in-depth analysis is that some of the above-mentioned authors (especially Bratman) have since updated their initial theories to be applied to large and complex forms of collective action. So instead of first setting up, and then arguing against dusty-old straw-men, I will address these modified versions below.
And instead of committing to any particular view on collective action, I will, at least for the time being, stick to a minimal, and preliminary definition of collective action which omits an analysis of the way in which collective actions *come about*, i.e., whether they are achieved by collective intentional states of individuals; and if so, how these collective intentional states should be conceptualized.[13] My preliminary, minimal definition of collective action then is this:

COLLECTIVE ACTION: A collective action *A* occurs *if and only if* two or more agents S_1-S_n are *A-ing together in a non-distributive sense*.

Alternatively, Hammond defines what he calls "joint action" as a "*complex action, parts of which are performed by several people working together*" (Hammond 2016, 2710). If two or more agents "each want to accomplish a goal, the participation of both is required to accomplish it, and each does his or her action which contributes to accomplishing that goal, this would be a case of joint action" (ibid).

[12] Fiebich discusses the theories of Gilbert, List & Pettit, Bratman, and Searle.

[13] In the third chapter, however, I will focus on one approach of shared, or collective intentions that I find particularly illuminating in explaining the actions of large and complexly structured social groups, i.e, Kirk Ludwig's (2017a, 2017b) account of *we-intentions*, which figure in his *Shared Plan Account* of collective action.

Institutional groups

I will now try to give a definition of the term "institutional group" and consider some paradigmatic institutional groups whose agency I want to explain. I will also list some of the features of such institutional groups that I think any account of institutional group agency should be able to capture. So let me start by clarifying some basic notions. First, we need to consider the *terminology* used to describe institutional groups. When it comes to social groups in general, there is no consensus amongst philosophers as to which terminology should be used to describe the phenomena.[14]

Second, the term "institution" needs to be clarified too. Let me explain how the terms "institution" and "institutional group" relate to one another. Depending on one's definition of institutions, these terms either denote the same thing, or they describe different things altogether. Some authors, e.g., Miller (2001; 2019) or Ludwig (2017b) use the term "institution" and "institutional groups" exchangeably. But at other times, they also talk about "organized institutions" as a specific subclass of *social institutions* (cf. Miller 2019). So actually, let me explain two things: First, what the term "institution" is supposed to refer to in connection with, or in contrast to "institutional groups". And second, what the term "institution" itself is supposed to refer to. To this latter end, Seumas Miller remarks that:

> "in ordinary language the terms ‚institutions' and ‚social institutions' are used to refer to a miscellany of social forms, including conventions, rules, rituals, organisations, and systems of organisations. Moreover, there are a variety of theoretical accounts of institutions, including sociological as well as philosophical ones. Indeed, many of these accounts of what are referred to as institutions are not accounts of the same phenomena; they are at best accounts of overlapping fields of social phenomena" (Miller 2019).

On an general level, the term "institution" is often used to describe rather abstract phenomena such as *money*, *marriage, private property*, *rules of traffic, language,* or *capitalism.* To make matters more concrete, a type-token distinction can be introduced: whereas *money* is a type of institution, the *U.S. Dollar* is a token-

[14] Consider this non-exhaustive list of authors writing about social groups, who all use a different terminology to describe them. Katherine Ritchie (2013, 2015, 2020a, 2020b) fundamentally distinguishes what she calls "feature groups" from "structured groups". Kirk Ludwig (2017a, 2017b, etc) differentiates between so-called "∈-membership groups" and "ε-membership groups" but he also talks about "pure" and "hybrid" institutional groups in contrast to "aggregate groups". Raimo Tuomela (2013) draws a distinction between "We-Mode groups" and "I-Mode groups", whereas Margaret Gilbert (1990; 2006) mostly writes about "plural subjects". Ludger Jansen (2016) differentiates "formal" and "informal" groups, and he also distinguishes "person-collectives" [Personenkollektive, M.G.] from "collective persons" [Kollektivpersonen, M.G.]. Jansen further states that there are "first-order" and "second-order" groups. John Searle (1990; 1995) plainly writes about "groups", whereas Seumas Miller (2001, 2010) draws the distinction between "groups", "organizations" and "institutions". Deborah Tollefsen (2015) differentiates between "aggregative", and "corporate" groups, but she also writes about " "plural" and "institutional" groups. Jennifer Lackey (2021, 5f.) posits that groups exist on a "spectrum" ranging from "collections of individuals" to "highly structured groups". List & Pettit (2011) talk about "organized groups", "social Integrates", "goal-oriented groups" and "collectives". Peter French (1995; 2020) writes about "aggregate collectives" (such as "pods" and "clods") in contrast to "conglomerates" (such as "corporations"). Dave Elder-Vass (2010) differentiates between "norm circles", "organizations", "interaction groups" and "associations". Finally, Christopher Kutz (2000) states that there are "ephemeral" and "institutional" groups. Garcia-Godinez (2020) writes about "institutional groups".

instance of said type. Marriage is another type of institution, whereas the *marriage between Paul and Patricia Churchland* instantiates a token of this type.

Two main theories of institutions deal with such phenomena: the theory of *institutions-as-equilibria,*[15] and the theory of *institutions-as-rules,*[16] as well as combinations of such theories.[17] My project is not about institutions in this abstract sense and I do not provide an analysis of institutions such as *money*, *marriage*, or *languages.* Rather, I will endorse one particular view of institutions, which is of interest for my thesis: the view of institutions as *organizations,* or *organized groups*. Let me explain what organized groups are, and how they differ from other kinds of groups.

I hold institutional groups, understood as organized groups, to be *social groups*, i.e., groups of *people* (human people or otherwise social creatures). I do not aim to give a definition of what makes groups "social" other than they are realized by, or composed out of people (or other social creatures). Defining *sociality*, especially in a non-circular fashion, is a difficult, if not impossible task (cf. Ritchie 2020b, 402f.). There are however obvious examples of *non-social* groups, which will not be of interest to me: a group of chairs, mathematical sets, geological formations such as the *Vadito Group* in New Mexico, or the group of islands called the *Indonesian Archipelago*.

Next, we can distinguish social *groups* from social *pluralities,* or social *aggregates*. Social pluralities are the mereological sum of their parts. Two people compose a plurality *iff* these two people are the only members of the plurality. In turn, two pluralities are identical to one another if and only if they are both composed of the exact same people. Social groups, however, are different from social pluralities. Two social groups are not necessarily identical to one another only because they are both composed of the exact same members. Groups, but not pluralities, can be both extensionally coincident and numerically distinct (cf. Ritchie 2015,

---

[15] On the institutions-as-equlibria view, institutions are regarded as *equilibria* that exist in game-theoretical scenarios. In game theory, an equilibrium is "a profile of actions or strategies, one for each individual participating in a strategic interaction" (Guala 2016, xxiv). What's special about these profiles is that they describe states of strategic games in which "each strategy of an individual player within such a game is a best response to the action of the other players" (Hindriks and Guala 2015, 464; also: Hédoin 2021, 75f.). According to this view, e.g., *driving on the right side of the road* is an institution that facilitates the coordination of individuals participating in traffic. For two influential versions of the institutions-as-equlibria approach see Lewis (1969) and Sugden (2005).

[16] An influential account of the *institutions-as-rules* view has been put forward by John Searle (1995, 2005, 2010) who claims that institutions should be understood as: "any collectively accepted system of rules (procedures, practices) that enable us to create institutional facts" (Searle 2005, 21f.) According to Searle, the constitutive rules that enable the creation of institutional facts "typically have the form of *X counts as Y in C*" (ibid), where some object, person, or state of affairs (the X-term) is assigned a special status (the *Y* status) in a given context. This Y-Status "enables the person or object to perform functions that it could not perform solely in virtue of its physical structure, but requires as a necessary condition the assignment of the status" (ibid). *Money*, for Searle, is such a paradigmatic institutional fact which is based on the assignment of a status functions (the Y status) onto physical objects. There is, however, a connection between the theory that views institutions as rules and those institutional groups that I want to target with my analysis. I will explicate this connection in detail below (see especially the sections on *constitutive rules* and *status functions and collective acceptance* in Chapter 3.1). Another influential theory of institutions, which can be considered as a variation of the *institutions-as-rules* view, has been provided by Raimo Tuomela (2003; 2013), who defines institutions as *social practices.*

[17] For a general overview of these two theories: Greif & Kingston 2011; Hindriks 2017; Hédoin 2021; Miller 2019. A recent account of institutions, provided by Francesco Guala (2016) and Frank Hindriks (Hindriks & Guala 2015) aims to unify the two main approaches to institutions into what they call the *institutions-as-rules-in-equilibria* approach.

313). So institutional groups are *social groups* in contrast to both non-social groups *and* social pluralities. To further characterize institutional groups, I will now follow Katherine Ritchie's *structuralist account* of *organized social groups* and argue, that institutional groups are *organized social groups* in the way Ritchie defines them.

Ritchie fundamentally distinguishes *organized* social groups from *feature-based* social groups. Feature-based groups are groups of people that are based on their individual members having a common (socially dependent) feature, or exhibiting a common (socially dependent) property. There's a wide range of social groups which can be described to be feature-based, e.g., gender groups (e.g., women, men, non-binary people etc.), racial groups (e.g., blacks, whites, latinos), economic classes (workers, the 1%), or sexual orientation groups (e.g., heterosexuals, queers, asexuals, etc.) (cf. Ritchie 2020b, 414f). Ritchie also calls these groups whose features are dependent on social factors *social kinds* (ibid).

Contrast these groups to what Ritchie calls *organized groups*. Organized groups, e.g., baseball-teams, committees, book-clubs, or military units, differ in critical respects from feature-based groups. Ritchie explains one key difference as follows. An organized social groups' *identity conditions*

> "rely on the way they are organized or structured. For example, a baseball team's persistence seems to require some things playing functional roles (e.g., pitcher and catcher) that are specified by its organizational structure. The baseball team would not persist if it had no organization; it would be merely some individuals […] Having a particular structure or organization is not relevant to the identity conditions of [feature-based] groups like Latinos, gays, or Whites" (Ritchie 2013, 313).

Organized groups differ from feature-based groups in that organized groups have two components: they not only have *members* but they also have a *structure*. An organized group's structure captures its "functional organization" (Ritchie 2020b, 411): "For instance, a baseball team's structure captures the functional roles of the catcher, pitcher, outfielders, etc. Relations might include *calling the pitch*, *pitching to*, *returns the ball to*, and so on" (ibid).

But organized groups are not *identical* to either their structure or their members. They *have* a structure, but they are not *identical* to their structures. Rather, they are *instantiations* of a structure. They are structures that are instantiated, or have been *realized* by individuals (cf. Ritchie 2013, 268-271; 2015, 316). Because organized groups have a structure, they are not identical to the sum of their members. For individuals to constitute an organized group, these individuals have to stand in certain relations to one another. These relations are interesting, as they allow for two features of organized groups. First, the same set of individuals can constitute numerically distinct organized groups, because the same set of individuals can stand in *different* relations to one another, depending on which groups they constitute. Consider, e.g., a biker gang of which all members are also the only members of a book club dedicated to reading Hegel's *Phenomenology of Spirit*. The book club and the biker gang have the same individuals as members, but they constitute numerically distinct groups. Second, organized groups, in virtue of them having a structure, can survive a change of membership. If one member leaves the biker gang and another individual takes up her position within the group structure, the biker gang persists.

To see how organized groups come to exhibit these features, we need to take a closer look at such an organized group's *structure.* To say that groups have a structure, is to say that the individuals that instantiate such groups stand in particular *functional relations to one another* (cf. Ritchie 2020a, 95). At the most abstract level, a group's structure is composed of what Ritchie calls *nodes* and *edges*. Edges in a structure capture the relations that hold *between* nodes. In turn, nodes are defined in terms of their relations to *other nodes* (cf. Ritchie 2015, 316). Now in the case of organized groups composed out of people (or other social creatures), the nodes of such organized groups can be identified with positions, or *roles*, and the *edges* can be identified as the *relationships* between such positions, or roles. In a less abstract fashion, Ritchie defines the structures of organized groups as follows:

> "the structures of organized groups consist of roles and relations between them. Roles are defined in terms of relations to other roles, tasks that role-players are allowed or required to carry out, and in some cases specific features a role-player must have. Relations between roles might be hierarchical or non-hierarchical. Relations that involve deference and power are hierarchical. For instance, a role might allow a role-player to give orders to individuals playing other roles. Relations of seconding a motion or reporting on a project involve relations between group members that are non-hierarchical. Relations between roles also capture the ways playing a role depends on other roles being played" (Ritchie 2020a, 95).

Roles, in an important sense that I will explain in more detail below, are *agent-ambiguous*. Besides certain conditions of membership that individuals need to fulfill, it does not matter which *specific* individual occupies a given role. And because membership in organized groups is defined in terms of *occupying roles*, the members of a group can change while the organized group in question persists (more below in Ch. 3.1-3.2). With this minimal characterization of *organized* groups in place, I am now in a position to give a preliminary definition of institutional groups:

> **INSTITUTIONAL GROUPS:** Institutional Groups are organized groups that consists of an embodied (or realized) structure of differentiated roles. *Institutional roles* are defined in terms of their interdependency, and in terms of tasks that individual members have to perform, as well as rules regulating how these tasks are to be performed.

In the third chapter, I will examine related accounts of institutional groups more thoroughly and provide a more fine-grained description of both institutional groups and institutional roles. Until then, let us stick with

this preliminary definition of institutional groups, which is fairly uncontroversial and enjoys - while admitting some variations - a wide use throughout the existing literature on institutional group agency.[18]

To further characterize such institutional groups, let me list six features that such groups may exhibit: 1) COMPLEXITY, 2) STABILITY, 3) TRANSFORMATION, 4) EMBEDDEDNESS, 5) EXCLUSIVITY/INCLUSIVITY, and 6) ACTIVITY/PASSIVITY.

COMPLEXITY: Institutional groups can be complexly structured and they can have large numbers of individuals as their members. Consider, e.g., *Walmart*, the multinational retail corporation. With more than two million individuals, more people are employed at Walmart than there are citizens of Slovenia (Khanna 2016). For complexity, consider the European Union. The European Union has, besides its 27 member states, seven central institutions, seven so called *EU bodies*, over 30 decentralized agencies, 20 administrative agencies, four inter-institutional services, as well as *non-standing executive agencies*, *European Union corporate bodies*, and EU *joint undertakings* (see: European Unions Directorate-General for Communication 2024).

EMBEDDEDNESS: Institutional groups can be *embedded in*, or be comprised of *other institutional groups*. Consider the UN as an example of such a *meta*-institution, which has over 190 sovereign states as its members. The *Bochum Police Department*, in turn, is an institutional group that is *embedded* in the North Rhine-Westphalia (NRW) State Police Force.

TRANSFORMATION: Institutional groups can *evolve* over the course of time. When, e.g., *Apple* was founded in 1976, it had three members. Today, it's a multinational corporation with more than 160.000 employees and a market cap bigger than most countries' GDP (cf. Economic Times 2023). The ability of institutional groups to transform does not only imply that they can grow and shrink in size. They can also transform in terms of their structure and functions. Institutional groups can re-organize their structure and adapt to changes in their environment. Warren Buffett's *Berkshire Hathaway,* a conglomerate of holdings and investments-funds once started as a clothing manufacturer (cf. Berkshire Hathaway 2024). *Nokia* was founded as a pulp mill that produced card boxes. Then it switched to producing rubber tires, then it switched to producing cell-phones. Today it is on its way to transform into a provider for large-scale telecommunication infrastructure (cf. The Guardian 2013).

STABILITY: Institutional groups can - at least in principle - exist indefinitely. They, unlike us mortal humans constituting them, do not necessarily have a date of expiry. Some institutional groups have been around for centuries and outlasted numerous generations of their members. Also, they can bridge periods where no individual is a member of the institutional group. Consider e.g., the *German Federal Assembly*, which has as its function to elect the President of Germany. The Federal Assembly functions as so called *non-standing* constitutive organ of the German state. It is convened every five years to elect the President of the German Republic. After this election, the Federal Assembly dissolves, only to re-instantiate itself in the next electoral cycle (see for further notes on this feature: Ludwig 2017b, vii).

---

[18] see: Miller 2001; 2010; 2019; Bratman 2021; 2022; Kutz 2000; Tuomela 2013; List & Pettit 2011. Similarly, Garcia-Godinez defines institutional groups as "a realization of a formal group structure" (Garcia-Godinez 2020, 39). And Ludwig defines institutions as an "organized differentiation of roles directed toward joint action, which may be occupied successively by distinct individuals" (Ludwig 2017b, 2; also: 2020). John Rawls, in his *Theory of Justice*, defines institutions as "a public system of rules which defines offices and positions with their rights and duties, powers and immunities, and the like. These rules specify certain forms of action as permissible, others as forbidden; and they provide for certain penalties and defenses, and so on, when violations occur" (Rawls 1999, 47f.).

EXCLUSIVITY/INCLUSIVITY: Institutional groups consists of an embodied structure of differentiated institutional roles. But not every individual is eligible to occupy any role, and so not every individual can become a member of just any institutional group. Institutional groups can have *conditions of membership*. The most basic condition is often taken to be *voluntary acceptance of membership* on behalf of the individual which joins a group (Ludwig 2017b; Miller 2019, Sec. 3; Searle 2010). But voluntary acceptance of membership is neither necessary nor sufficient to actually become a member of an institutional group. It is not sufficient, as there can be other criteria for membership that pertain to features that an individual has to exhibit, e.g., a certain age, a certain height, academic or scholarly titles, certificates of education, etc. For some institutional groups, it simply does not suffice that one recognizes oneself to be a member of them. One does not, e.g., become a member of the U.S. Supreme Court by simply thinking of oneself as a member of the court. Also, voluntary acceptance sometimes may not be necessary to become an institutional group's member, because membership in institutional groups can be *involuntary*: Citizenship is a prominent example of membership in institutional groups which can be both voluntary and involuntary. It is possible to voluntarily become *a legal citizen* by means of naturalization. But other times, individuals automatically receive citizenship at birth in virtue of their parents' citizenship, where this does not require the voluntary acceptance on behalf of the infant.

ACTIVITY/PASSIVITY: Institutional groups can (although they do not need to) have different types of members, i.e., they can have *active* as well as *passive* members.[19] Consider e.g., political parties, labour unions or football clubs. In such groups, there usually are what we could call both *active* and *passive* members. Passive membership in such institutional groups can consist of nothing more than, e.g., paying your membership-fees. It does, e.g. not compel you to attend any of the groups meetings, open and read the letters they send you, nor to take part in any election they hold. In fact, you can completely forget about your membership without this entailing any consequences for you or the group which you are a member of. *Active* membership in such groups is different. It usually implies that the individual is - in some way or the other - involved in the group's *activities*, e.g., it may be involved in *organizing* group meetings, *writing and sending letters* to members, or *convening elections*. Active members in institutional groups are sometimes *employed* by the institutional group, or receive other forms of compensation for their work. But in other cases, individuals do not get compensated, but *volunteer* to be active members of such institutional groups.

Before moving on, please note that these features are somewhat *contingent* and not all institutional group will exhibit every listed feature. Also, they should not be understood as features which are necessary (or sufficient) for a social group to count as an institutional group. Still, I hold them to helpfully characterize the paradigmatic institutional groups which my thesis will focus on, i.e., large and complexly structured institutional groups like *corporations, administrative agencies*, *political parties*, *trade-unions,* etc. So with

[19] The distinction between *active* and *passive* membership is not to be confused with Raimo Tuomela's basic distinction between *operative* and *non-operative* members (see Ch.2.2.3.). For Tuomela, being an operative member means that one is *authorized* to accept certain propositions to hold *for the group*. In contrast, non-operative, or rank-and-file members "must tacitly accept […] or go along with the operative members' collective acceptance of the proposition" (Tuomela 2013, 162f.). To see why active and operative membership are not the same, notice that *one can be both an active and non-operative* member of a group, e.g., when one's institutional role requires one's active exercise of tasks and functions, but merely requires *compliance* with the operative's decisions and acceptances.

these features at hand, let me turn to some examples of institutional groups that I hold to be paradigmatic for my endeavor. Institutional groups which will be of interest for my thesis are groups like, e.g., corporations, military units, police departments, fire brigades, churches, professional associations, political parties, orchestras, labour unions, or football clubs. Corresponding to these types of institutional groups are token-instantiations like, e.g., *Apple*, *IBM*, *Walt Disney Company, Borussia Dortmund*, the *U.S. 82nd Airborne Division*, the *Bochum Police Department*, the *New York City Fire Department*, *United Auto Workers* (*UAW*), the *International Society Ontology Society* (*ISOS*), the *SPD*, etc. Notice, again, that my thesis will focus on *especially large and complexly structured* institutional groups.

As mentioned above, these types of institutional groups have great importance for understanding modern life. Woven into the fabrics of our societies, they have considerable influence on our lives and well-being. Institutional groups "dominate modern life" and "give us the great power that we have achieved over the natural world" (Ludwig 2017b, vii). In order to understand why, and especially *how* institutions can be said to exert such influence, let us now circle back to the concept of *group agency*.

Group agency

Colloquial speech purveys the idea of institutional group agency and of institutional group agents. Take as examples sentences like "Apple intends to use 100% recycled cobalt in battery production by 2030", "Russian military plans to attack other countries", "UAW wants to unionize Tesla", or "the AfD acts to dismantle democratic institutions". We intuitively seem to grasp the meaning of such sentences. However, there are different philosophical views of what *group agency* actually is and how claims for *group agents* should be understood. And just like in the other philosophical domains discussed so far, the terminology to describe the different views is contested too. To follow a useful taxonomy suggested by List & Pettit (2011) and Lackey (2021), we can first contrast *eliminitavist* with *realist* theories of group agency. Eliminitavism comes in two varieties: First, as a *metaphor-theory,* and second as an *error-theory*. Both views are united in their claim that it is *literally false* to speak of group agents and group agency. Within the *realist* camp, we can distinguish *deflationary* (or *reductive*) views from *inflationary* (or *non-reductive, holist*) views of group agency and group agents. Both realist views admit that sentences about group agents performing actions are literally *true*, but differ in their analysis of what this truth amounts to, or consists of.

Three consecutive questions can clear the field (cf. List & Pettit 2011, 7). The first question is whether talk about group agency is always metaphorical. If one answers this affirmatively, one subscribes to the *metaphor-theory* of eliminitavism, sometimes also called *fictionalism* (see: Moen 2023). According to the metaphor-theory of eliminitavism, it is *literally false* that groups perform actions or that there are group agents. Rather, talk about group action is either not meant to convey literal truth, but rather its just mere elliptical talk about actions of individual agents. Or, if taken as true, it is in error. Anthony Quinton often has been described (by e.g., Lackey 2021; Moen 2023; List & Pettit 2011; Schmitt 2017) as a classical metaphor-type eliminitavist regarding group agents. He states that:

> "We do, of course, speak freely of the mental properties and acts of a group in the way we do of individual people. Groups are said to have beliefs, emotions, and attitudes and to take

> decisions and make promises. But these ways of speaking are plainly metaphorical. To ascribe mental predicates to a group is always an indirect way of ascribing such predicates to its members … To say that the industrial working class is determined to resist anti-trade union laws is to say that all or most industrial workers are so minded" (Quinton 1975, 17).

According to eliminitavist views of group agency there exist only individual agents. Metaphor-theory conceives of group agents as metaphorical shortcuts, and talk about them to be *literally false*. When individuals cooperate in groups, so the eliminitavist, "they do not bring novel agents into existence" (List & Pettit 2011, 3). In turn, to think that institutional groups "are genuine corporate agents that act on the basis of collective attitudes would be to mistake metaphorical shorthand for literal characterization" (ibid). The second question is whether *non*-metaphorical group-agency talk is always misconceived, or in error. If one answers the second question affirmatively, one subscribes to *error-theories* of eliminativism. Here, talk about group agents is intended to be taken for literal truth, but error-theories suggests that it is misconceived, and ultimately false anyway (ibid).

Now if one negates that non-metaphorical talk about group agency is always in error, one begins to charter *realist* territory. Here, the consequent third question to ask is whether talk about group agents, while not misconceived, is taken to be *reducible* to individual-level talk (cf. List & Pettit 2011, 7).

The two main options here are either to endorse a "redundant realism" or to be a "non-redundant realist" about group agents. Other ways to express the difference (see: Lackey 2021; Moen 2023) is to say that one is either endorsing a redundant, i.e. *reductive* or *deflationary theory* about group agents, or whether one commits to non-redundant, i.e., *non-reductive* or *inflationary* theories about group agents.

Critical discussion of reductive as well as non-reductive views of group agency will be at the heart of the second chapter of my thesis. For now, and to move forward, it can be summarized like this: Reductionists claim that group agency, i.e., the capacity of groups to perform actions can be understood entirely in terms of individual members and their agency, including their capacity for *collective action*. Reductive views admit that it is literally true that groups perform actions, but "such claims are made true entirely by individual members of the group" (Lackey 2021, 4) acting in certain ways. So groups are not genuine agents "in their own right" as their agency can be reduced to the agency of individuals that stand in certain relations to each other. A group agent, to paraphrase the reductive theorist Kirk Ludwig, is a *group of agents* (cf. Ludwig 2017a, 296-298). Reductive views then commit to group agents only in "weak, ontologically non-committal sense" (Paul 2020, 142).

Non-reductionist, or inflationary views reject this view and argue that groups can form intentional attitudes which are *not* reducible to those intentional attitudes of their members. Inflationary theories of group agency view groups to be agents with "minds of their own", or "in their own right". Regarding this irreducibility, Moen, notes that the non-redundant realist view on group agency is

> "therefore distinct from the idea of ‚shared agency', associated particularly with Michael Bratman, where a shared intention is understood as a state of affairs where individuals hold the same intention to perform an action together. As Bratman notes, such an intention is not attributed to a ‚superagent'. The mental states occur only in individuals' minds. Group-agent

> realists, on the other hand, do attribute mental states to a superagent, or group agent. Certain groups, such as political parties, multi-member courts, legislatures, expert committees, commercial corporations, and tenure committees, they argue, have their own attitudes, and should therefore be understood as agents" (Moen 2023, 45).

According to such theories, certain - though not all- groups can form and consequently act on their own, irreducible intentional states. Non-redundant views of group agency should then not be equated with the concept of *collective*, or *shared action*.

There are different explanations for how to make sense of such irreducible intentional attitudes, and how groups could be said to realize them in their own right (for a historical overview of such theories see: List & Pettit 2011, Ch. 3.3.; also: Rovane 1997). As Moen's use of the term "superagent" may seem to suggest, the claim that groups have "minds of their own" has often been (mis-)understood to involve some spooky, metaphysically obscure element, like an *vis vitalis*, or a Hegelian *Weltgeist.* But non-reductive views of group agents do not necessarily have to involve such mysterious "transcendent realities" (List & Pettit 2011, vii). As will be shown below (Ch. 2.1.), theories of group agents rather aim to *demystify* the concepts of irreducibility and irreducible intentional states. They work by wit and not by witchcraft.

On, e.g., List & Pettit's view, there need not to be an ontologically independent and free-floating "hive mind" of a super-agent involved in such theories. The non-reductive intentional states of group agents, as List & Pettit argue, still supervene on the contributions of their members. Rather, group-level intentional states and the individual intentional states these group-level states are supposed to supervene on "relate in such a complex way […] that we have little chance of tracking the dispositions of the group agent, and of interacting with it as an agent to contest or interrogate, persuade or coerce, if we conceptualize its doings at the individual level" (List & Pettit 2011, 76). The independence of List & Pettit's group agent is an *epistemological* independence, rather than an *ontological* one.

In order to move forward, let me summarize the main points of these sections: First, I will use the term "action" to refer to events that have appropriate mental items such as beliefs, desires, and intentions as their causes. In turn, agency (minimally) describes the capacity to perform such actions, or to act, i.e., to bring about change in one's environment on the basis of states such as beliefs, intentions, and desires. The term "agent" refers to any entity with such a capacity. Second, a *collective*, or *joint* action occurs iff two or more agents S_1-S_n are *A-ing together in a non-distributive sense*. Third, the term "institutional group" refers to organized social groups that consist of an embodied (or realized) structure of differentiated roles. Fourth, in trying to understand claims regarding the agency of such institutional groups, reductive (or deflationary) theories of group agency argue that the agency of institutional groups can be reduced to the *collective actions* of their individual members. In turn, non-reductive (or *inflationary)* theories argue that the agency of institutional groups cannot be reduced to collective action, but that *groups themselves* are genuine agents, i.e., entities with the capacity to perform actions. While these minimal clarifications leave many questions unanswered, I hope that the reader can at least grasp the basic thoughts behind these concepts. In order to move forward, let me outline the structure of my thesis.

Structure of the thesis

Here's my thesis in a nutshell: In Chapter 2, I will give a systematic overview of the existing literature on institutional group agency and argue that both reductionist and non-reductionistic accounts have certain problems in explaining the agency of large and complexly structured institutional groups. In Chapter 3, I assemble what I call "role-based" theories of institutional agency and argue that they are the best option available to explain the agency of institutional groups. In Chapter 4, I develop a novel argument against the way in which these "role-based" theories conceptualize institutional roles. In Chapter 5, I develop my account of "Role Agency", which aims to provide for a deepened understanding of the way in which individuals relate to their institutional roles.

But let me be a bit more precise. The chapters of this thesis are organized in the following way: The main goal of the second chapter is to provide the reader with an overview, as well as a critique of the current debate about the agency of institutional groups. The chapter is divided into two main sections. In the first section, I will turn to prominent, realist approaches to institutional agency which argue for the possibility of genuine, irreducible group agents. As mentioned, such realist, or non-reductive theories advocate for the existence of group agents "in their own right". During the course of this section, I will provide reasons why such non-reductive theories of group agents can ultimately be forsaken. In the second section, I will then turn to reductive theories that try to explain the agency of institutional groups through the *collective actions* of the individual group members. According to the latter type of explanation, a group's agency can be reduced to the agency of its individual members *acting together*. While I ultimately think that this is the right way to understand institutional agency, I will put forward reasons for thinking that the depicted theories fail to account for the particular features of *institutional* group agency, which my thesis wants to illuminate.

So beginning in Ch. 2.1.1., I will discuss one of the classical accounts of group agency, i.e., Peter French's theory (1979; 1995; 1996) of the *corporation as a moral person.* French argues that certain institutional groups, which he calls *corporations*, are *agents* that have *corporate intentions*. According to French, corporations are agents in virtue of them having *Corporate Internal Decision* (*CID*) *Structures*. Such CID-Structures regulate both *how* members of corporate groups make decisions and *which decisions* they make. Because of these two functions, French argues that CID-Structures allow for the re-description of the intentional actions of the corporation's members as the *corporate intentions* of the group itself. And because a agent is an entity that is guided by intentions, French concludes that corporations are agents in their own right.

A similar approach to institutional agency will be examined in Ch. 2.1.2. Michael Bratman (2022) recently argued for institutional groups to be agents because of their *procedural rule-based infrastructure*. Bratman argues that the outputs of what he calls *social rules of procedure* are functionally equivalent to *intentions*. By arguing that such *institutional intentions* can provide *robust guidance*, he takes institutional groups to have a *Frankfurtian standpoint,* and thus agency.

In Ch. 2.1.3., I will turn to a related, non-reductive theory of group agency, established by Christian List and Philip Pettit. List & Pettit define a *group agent* as an agent that consists of individual persons (cf. List & Pettit 2011, 8f.). What makes these individuals comprise one numerically distinct group agent is that they have

suitable coordinated dispositions to think and act in such a way that properties of the group *holistically supervene* on the dispositions of the individuals. Like French, the authors argue that distinct decision procedures, which they call *aggregation functions,* are key to understand this holistic relation of supervenience. The authors here rely on the so called *Doctrinal Paradox* to show how aggregative functions can generate group-level attitudes which are different from - and do not seem to be reducible to - those of the individual members. Groups, then have a *mind of their own*.

In Ch. 2.1.4., I turn to Deborah Tollefsen's *interpretivism* about group agency (2002a; 2002b; 2015). Her theory can be summarized by the following claim: Because our rich and sophisticated practices of interpreting groups *as if they were agents* are successful (e.g., in explaining their past actions, or in predicting certain actions to be performed in the future), we can infer that they actually *are* agents. Taking the so called *collective stance* towards groups lets us discern patterns of group behavior that we would otherwise miss out on.

The second section of the chapter examines theories that hold the agency of institutional groups to be based on the *collective actions* of their members. To sort out theories of collective action unfit for application to institutional groups, I will first discuss the so called *Upscaling Problem* in Chapter 2.2.1. Roughly, the problem argues that "scaling up" accounts which base their analysis of collective action on small-scale, egalitarian and highly interdependent cases leads to implausible results when we use them to explain the agency of large and complexly structured institutional groups. I then focus on theories which explicitly target such institutional groups, i.e. the theories of Christopher Kutz and Raimo Tuomela.

In order to explain the agency of large and complexly structured institutional groups, Christopher Kutz (Ch. 2.2.2.) develops a minimalist account of collective action that he claims to be "parsimonious in its metaphysics and philosophical psychology, sufficiently undemanding to account for the cooperation of loosely-linked agents, and anti-egalitarian enough to reconcile collective action with hierarchy" (Kutz 2000, 2). According to Kutz, the concept of *overlapping participatory intentions* is key to understand the form of collective action that characterize institutional groups. Such participatory intentions are formed if an individual agent understands her individual action as *contributing to a collective end* and thinks of her actions as *doing her part* in order to contribute to that end.

In Chapter 2.2.3., I turn to the *positional theory of group agency* formulated by Raimo Tuomela (2013). Tuomela argues that the agency of institutional groups can be explained by pointing to the so called *we-mode*. According to Tuomela, when individuals act (reason, deliberate, etc.) in the we-mode, they are acting (reasoning, deliberating, etc.) *as a group member* and *for the group*. Under certain conditions obtaining, individuals then constitute *we-mode group agents*. Such *we-mode group agents*, however, are *fictitious* entities (cf. Tuomela 2013, 48ff.). Besides his we-mode, Tuomela also offers a theoretical explanation for how groups come to have structures and hierarchies. To this end, Tuomela provides a *positional* theory of authority which allows us to distinguish between *operative* and *non-operative* group members.

The third chapter focusses on *role-based theories* of institutional group agency. By this, I refer to theories that, at their core, argue for the claim that an institutional group action consists of (and consequently can be reduced to) the individual contributory actions of its members, who *act in their assigned* roles, or *qua role-occupancy*. To say that members of an institutional group act in their assigned roles, is to say that they perform the functions and tasks definitive of the roles they occupy.

I focus on three authors who endorse such role-based explanations of institutional agency: Katherine Ritchie (2020a), Seumas Miller (2001; 2010; 2019) and, perhaps most prominently, Kirk Ludwig (especially: 2017b). I want to stress that these authors do not explicitly label their own theories as "role-based". I will, however, provide the reader with reasons to think that such a label is nevertheless warranted. All authors agree not only on how to characterize both *institutional groups* and *institutional roles*, but they also agree on the way in which institutional roles figure in the explanation of institutional group agency.

So Chapter 3.1. will start with the concepts of institutional groups and institutional roles employed by such accounts. The upshot of this section is that institutional groups are best viewed as interrelated structures of institutional roles, which in turn are defined through tasks and functions, that an individual role-occupant must perform. Here, I follow Kirk Ludwig's view of institutional roles as *collectively accepted status functions*. As the analysis of collective acceptance often invokes concepts like mutual knowledge, or recursive beliefs, I will discuss whether such a view is feasible in light to the *Upscaling Problem*. Ludwig's *shared plan account* of collective action (2017a; 2017b), I will argue, is apt to explain the existence of institutional roles without invoking the concept of mutual knowledge, or recursive belief-cascades. These requirements, Ludwig argues, are neither necessary nor sufficient for institutional roles to exist (cf. Ludwig 2017a, 221). This, in turn, saves the account from the *Upscaling Problem*.

In Chapter 3.2., I investigate how the concept of institutional roles figures in explaining institutional agency. Clarifying this will require first to keep track of, and ultimately overcome two problems that reductive explanations of institutional group agency encounter: The problem of *Action Integration* and that of *Diachronic Group Constitution*. I will argue that role-based explanations can solve these problems by highlighting critical features of institutional roles in relation to institutional agency: First, institutional roles provide the individual with so called *role-based reasons* for action. Second, institutional roles *specialize* the actions of group members, leading to a differentiation of tasks and a division of labour. Third, the actions of individual role-occupants are functionally integrated into a *layered structure*. Finally, institutional roles allow for *representative*, or *proxy action*.

With this characterization at hand, I turn to the question whether role-based explanations can capture the *anonymity* and *compartmentalization* of institutional action in Chapter 3.3. According to Katherine Ritchie (2020a) members of institutional groups can be *minimally cooperative* in virtue of 1) them playing roles in an organizational structure and 2) them having a common goal (cf. Ritchie 2020a, 93). This claim, however, needs to overcome a *skeptical challenge* to explain how the condition of having a common goal could be harmonized with the feature of anonymity. If having a common goal includes symmetrical mental states, e.g., mutual knowledge or beliefs about each other's mental states, then cooperation most likely cannot occur anonymously. But if having a common goal means that the actions of individual role-occupants are *functionally integrated to achieve an end*, then cooperation may indeed occur anonymously. To show how such functional integration of role-performances could be understood, I draw on Seumas Miller's "Collective End Theory" (CET) of joint (or collective) action (Miller 2001). Here, I will argue that role-performances of individuals can indeed be functionally integrated to achieve an end without common knowledge being involved in this process. What makes such functional integration possible, is that institutional roles are *action-specific* but *agent-ambiguous*. Roles are agent-ambiguous because they are interchangeable, and thereby do not specify *who* performs a task. They are nonetheless *action-specific* by prescribing certain

actions that occupants must perform. On this basis, I conclude that mutual belief among individuals does not seem necessary for collective actions, at least in cases where these collective actions are constituted by functionally integrated role-performances.

Having presented such role-based explanations of institutional agency, the fourth chapter aims to identify certain problems that these accounts face. I will try to show that the way in which role-based theories characterize *the very concept of institutional roles* is under-theorized, especially in light of so called *discretionary powers*. To this end, I will pose a challenge to these accounts, which I christened the *two Problems of Discretion.*

Chapter 4.1. can be summarized this way: If we follow the standard characterization of institutional roles, they are primarily defined through tasks and functions that individuals need to fulfill, as well as rules that regulate *how* such tasks are to be performed. As such, institutional roles are interchangeable and therefore *non-specific* regarding the individuals that occupy them. But this *non-specificity* (understood as impersonality) of institutional roles comes with a price. As institutional roles are non-specific, they can only give any *particular* role-occupant generic instructions or *generalized directives* on what to do, without specifying how to execute the tasks and functions *in situ*, i.e., given a certain context or in a specific situation. But theorizing about role-dependent actions *in vacuo* is something different than actually performing such actions *in situ*. So in order to actually apply their tasks and functions, institutional roles encompass so called *discretionary powers*. But these discretionary powers, which allow for flexibility and adaptability of roles to specific circumstances, changing environments and unprecedented situations, give rise to the *two Problems of Discretion* (Ch. 4.2.). The conclusion to be drawn from these two problems is twofold: First, I conclude that the standard way of characterizing institutional roles is insufficient, and that such an insufficient characterization leads the established theories to be unable to make sense of the *two Problems of Discretion*. Second, I conclude that closer attention must be paid to the relation between individuals and the institutional roles they occupy. This serves as the motivation to develop my account of *Role Agency* in the fifth chapter.

My own, novel account of *Role Agency* aims to describe a form of agency that individuals engage in when acting in institutional roles that includes both the *internalization* and *idealization* of institutional roles. In Chapter 5.1., I demarcate my theory of Role Agency from Michael Schmitz's "role-mode" and argue, that it should primarily be understood as a *modification of individual agency* that stems out of an individual's capacity for *perspective* taking. Such perspective taking, in turn, occurs not by switching to a special *mode*, but rather on the basis of *role-specific reasons for action*. In a next step, I explain how *role-perspective taking* is connected to a role-occupant's ability to both *internalize* and *idealize* her institutional role. The process of *Role-Internalization* (Chapter 5.2.) helps us to answer how role-occupants may come to understand and have control over their institutional roles in the first place. I will analyze the concept of Role-Internalization along two broad dichotomies, i.e., along a *theoretical (*or *knowledge-based)* and *practical (*or *application-based)* dimension; and along a *formal* and *informal* one. From this, I derive four different dimensions of Role-Internalization: a 1) formal-*theoretical*, a 2) formal-*practical,* an 3) *informal*-theoretical, and an 4) *informal*-practical dimension of Role-Internalization.

Whereas Role-Internalization explains how role-occupants may come to understand and have control over their institutional roles, the process of *Role-Idealization* explains the ways in which individuals "step back",

and critically reflect on the institutional roles they occupy. This evaluative capacity, I argue in Chapter 5.3., is necessary for individuals to overcome problems concerning the discretionary powers vested in their roles. *Role-idealization*, I argue, should be understood as a social *heuristic*, where an individual role-occupant first determines an idealized (or prototypical) role-occupant and in a next step tries to imitate the idealized behavior (e.g., action, judgment, choice, decision, preference, or opinion). *Pace* Ludwig, I argue that such role-idealizations, in order to be *regulative* or *action-guiding,* must in some way be fixed to an external, independent standard. I then identify three levels, an interpersonal-level, an environmental level and a group-level, on which idealized role-performances can be fixed by such an external standard. I show that, on each level, we can identify particular mechanisms that provide the individual with an idealized standard against which she can measure her own, non-ideal role-performance. On the interpersonal-level, this may be explained by the process of *role-modeling*. On the environmental level, processes of *self-stereotyping* can provide such a standard. Lastly, I argue that institutional groups themselves may exhibit processes by which they actively try to provide their members with role-idealizations. I call these processes *processes of institutional ideal formation*. I discuss a real life example of such processes by examining a code of conduct of the *Disney World Resort* in Florida. The key findings of my thesis will be summarized in the conclusion (Ch. 6.), which also gives a brief overview on possible future paths of inquiry.

2. Theories of Institutional Group Agency

For by art is created that great Leviathan called a Commonwealth or State, in Latin civitas, which is but an artificial man, though of greater stature and strength than the natural, for whose protection and defence it was intended […] To describe the nature of this artificial man, I will consider: First, the matter thereof, and the artificer, both which is man.

Thomas Hobbes: Leviathan

‚Und Sie sind auch nicht wirklich so riesengroß, wenn Sie weit entfernt sind, sondern es sieht nur so aus?' ‚Sehr richtig', antwortete Herr Tur-Tur. ‚Deshalb sagte ich; ich bin ein Scheinriese.'

Michael Ende: Jim Knopf und Lukas, der Lokomotivführer.

As mentioned in the introductory chapter, there are different approaches to explain institutional group agency. This second chapter will tread along two broad explanatory paths. The first path to explain the agency of institutional groups is to argue for institutional groups to be genuine, non-reductive agents in their own right. I will call theories that commit to such a view *realist, non-reductive (or inflationary) theories of group agents*. On the other hand, *reductive* or *deflationary* theories will be at the center of the second part of this chapter. These theories do not argue for the existence of genuine, non-reductive group agents. Rather, they argue that the agency of institutional groups is *reducible to* the capacity of the institutional groups' members for *collective action*. Hence, I will call these theories *reductive (or deflationary) theories of collective action*.

The main goal of this chapter is to provide the reader with an overview of the existing literature and display those theories that try to answer the question of how we should understand claims about institutional group agency. I also want to motivate the thought that institutional group agency can be best explained by *reductive theories of collective action.* Ultimately, however, I will argue that existing theories of collective action run into certain problems too. This, in turn, will motivate the move to theories that focus on the concept of *institutional roles* to explain the agency of institutional groups. In Chapter 3, I seek to unify such a role-based explanation of institutional agency, which is itself reductive in nature.

2.1. First Explanatory Path: Group Agent Theories

As just mentioned, I will now examine realist theories which argue for institutional groups to be non-reductive, genuine agents in their own right. Recall that *non-reductive*, or *inflationary* views argue that groups can form intentional attitudes which are not reducible to the intentional attitudes of their members. Inflationary theories of group agency view groups to be agents with, as List & Pettit put it, "minds of their

own" (List & Pettit 2011, 77f). Here, two broad ways of arguing for a non-reductive theory of group agency stand out. First, *functionalist theories* and second, *interpretivism* about group agency.

Functionalism in the philosophy of mind can be - roughly - summarized as the claim that "what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the way it functions, or the role it plays, in the system of which it is a part" (Levin 2023). Similarly, Tollefsen defines functionalism as "the view that mental states are to be defined in terms of what they do rather in terms of their physical make-up" (Tollefsen 2015, 69). Applying this doctrine to groups, functionalism about the mental states of groups

> "seems to open up the possibility that groups could have mental states. If functionalism puts no constraints on the ‚stuff' that realizes functional roles, then it seems as if groups might be the sort of ‚stuff' that could realize functional roles, and hence mental states. As long as there is some state within the group that is playing the appropriate functional role, it will count as a mental state, regardless of the fact that it is not a mental state of a brain" (ibid).

I will discuss three takes on group agency which can be depicted as functionalist theories, i.e. the theories of Peter French (Ch. 2.1.1.), Michael Bratman (Ch. 2.1.2.), and Christian List and Phillip Pettit (Ch. 2.1.3.).[20] After examining these functionalist theories, I will separately discuss Deborah Tollefsen's *interpretivist* theory of group agency (Ch. 2.1.4.).

The three functionalist theories of French, Bratman, and List & Pettit all share the assumption that groups have the capacity to form intentional states in virtue of the functional role that certain *decision-making procedures* and mechanisms play in a group's capacity for intentional action. Further, all three authors try to show that these decision-making procedures and mechanisms can lead a group to realize intentional states that are distinct from, and irreducible to the intentional states of the members, which the group is comprised of.

I will clarify below how such claims for irreducibility are to be understood. But let me state the result of my discussion of these theories upfront. Ultimately, I want to argue against the view that groups are non-reductive, genuine agents "in their own right". Rather, I will pursue reductive explanations of institutional agency that rely on the concept of collective action. Regarding the scope and strategy of my arguments against group agent theories, it should be noted that I only hope to show that for some *type* of groups, i.e., structured institutional groups with interrelated and interchangeable roles, it is not *necessary* to view them as agents in their own right. I will argue that, for each theory, the apparently irreducible phenomena in

[20] This is, of course, only a very simplistic description of functionalism. However, I will leave it at this level, as I wish to argue neither for nor against the truth of functionalism per se. A brief but clear overview of functionalism regarding mental states of groups is provided by Tollefsen (2015, Ch. 4.). More sophisticated theories are provided by Huebner (2014) and Theiner (2018). For a critique of functionalism regarding groups see: Ludwig 2015a. Also, there are scholars that would challenge my division between functionalism and interpretivism. David Strohmaier e.g., claims that List & Pettit (2011) develop not a functionalist theory of group agency (See: Strohmaier 2020) but an *interpretivist* theory. However, List & Pettit (2011; also: List 2018) explicitly state that they draw "on a broadly functionalist theory of agency in order to establish that groups of individuals can in principle count as agents" (List & Pettit 2011, 75).

question, i.e., irreducible intentional states of a group, can be explained in a reductive, or deflationary fashion. There is then, ultimately, a methodological concern at play here. Ontological conservatism (or parsimony) sees for us to refrain from postulating entities (in this case non-reductive group agents) beyond necessity. Introducing new metaphysical resources and postulating the existence of entities should come with the burden of proof. I do not think that this burden is met in the case of group agent theories, and so I stick to ontological conservatism. And instead of committing to the existence of group agents into our ontology, I will, in the second part of this chapter, set out to explain the phenomenon of group agency in virtue of the *individual group members' capacity* for collective agency. This, however, should not be understood as an argument against the *a priori* or *logical* impossibility of group agents. Let me also stress that my arguments against these theories do not depend on the truth or adequacy of *functionalism in general.* It surely is an important question whether functionalist theories in the philosophy of mind are correct in explaining the nature of mental states. But I am not in a position, nor do I wish to answer questions of how to understand the general nature of mental states, or of intentional states more specifically.

2.1.1. French's Theory of Corporate Agents

Peter French developed a theory of group agency (originally put forward in 1979 and substantially revisited later in, e.g., 1995; 1996) according to which corporations are genuine agents because of their so called *Corporate Internal Decision* (*CID*) *Structures.* I will first take a closer look at these CID-Structures and then explain how they relate to a corporation's agency. According to French, CID-Structures consist of two elements:

First, a CID-Structure consists of an "organizational flow chart that delineates stations and levels within the corporation" (French 1996, 151). This organizational "flow chart" provides a corporation with its "grammar" (cf. French 1979, 213). By this, French means that the flow chart gives a corporation its organizational *structure*, by assigning individuals to positions and corresponding responsibilities. These structures can be *hierarchical* by establishing a set of power-relations between the members. Second, CID-Structures consists of *corporate decision recognition rules,* which function as the corporations *logic*. Such recognition rules "reveal how to recognize decisions that are corporate ones and not simply personal decisions of the humans who occupy the positions identified on the flow chart" (French 2015, 1). These recognition rules come in two varieties: First, by rules of recognition that settle how corporations are to arrive at certain decisions and *what* counts as a corporate decision in the first place. And second, they settle a corporation's *policies*, i.e., *what is* being decided upon. A useful analogy to understand CID-Structures, taken from Rönnegard (cf. Rönnegard 2015, 22), is to compare them with games, e.g., the game of chess: The organizational flow chart defines the roles within the structure, with their corresponding lines of responsibility. In chess, the roles are pawns, kings, bishops, etc. In corporations, these roles are roles of clerks, managers, executives, etc. Now the rules settling the *corporation's policy* dictate the goal of the corporation (the aim of the game) and corporate *recognition rules* define what counts as a corporate decision (the moves within the game).

The CID-Structures of a corporation are crucial to understand how corporations could be said to be agents. French's argument for corporate agency rests on two premises. First, he subscribes to Davidson's semantic account of agency, according to which for a corporation to count as an agent,

> "it must be the case that some things that happen, some events, are describable in a way that makes certain sentences true, sentences that say that some of the things a corporation does were intended by the corporation itself" (French 1979, 211).

French here draws on the idea that attributions of intentionality to describe an event are *referentially opaque* with respect to other descriptions of that same event (cf. French 1979, 211f). So events can be described in multiple, different ways. In line with a Davidsonian analysis, French subscribes to the view that an event is an action if and only if there is at least *one* description of what an agent did, which makes true a sentence that she did it intentionally (ibid). So, for French, a *corporate* action is an event such that there is at least one description of what the members of a corporation did (in accordance with the CID-Structure's procedures) that makes true a sentence that the *corporation* did it intentionally (ibid).

The second premise, for French, is that an operative CID-Structure allows for intentions and actions of biological persons to be *subordinated under*, and *synthesized into* corporate intentions and actions. A functioning CID-Structure "incorporates acts of biological persons" (French 1979, 211f), so that an individual agent *A*'s *x-ing* can be *redescribed* as a corporation *C*'s *z-ing* (ibid). Importantly, this attribution of intentional action to corporations is *not* accomplished if attributing intentions to them "is only a shorthand way of attributing intentions to the biological persons" who comprise the group (ibid). So on the one hand, corporations have to have the capacity to act intentionally in order to be agents. And on the other hand, this capacity must be explained in a non-reductive fashion.

So what does make true a sentence that a *corporation* did something intentionally? And how can CID-Structures "incorporate" the actions of their individual members? In a revised version[21] of his theory, French subscribes to the idea that the *planning capacity of CID-Structures* lies at the heart of corporate intentional action. French here relies on Bratman's (1987) planning theory of intentions on the one hand, and the Frankfurtian (1978) notion of action being *guided* by intentions on the other hand. According to this view, actions are being guided by intentions, which in turn are *planning states* of an agent:

> "What is important in deciding whether or not someone is acting is to determine what is going on when the movements are in progress and not what proceeded them. [W]hat we are looking for is whether or not the movements are ‚under the person's guidance', regardless of their antecedents. Actions are ‚guided.' Mere movements are not. Intentional actions are planned. They are undertaken deliberately, on purpose. The operative element in intentional action is

[21] Initially, French argued that for a corporation to do something intentionally, is for it to act on its *reasons*, subscribing to Davidson's belief-desire model of reasons. However, he later revised his view that corporations act on reasons as a pair of *corporate beliefs* and *desires* (See French: 1996, 148ff). His initial proposal also argued for corporations to be moral persons, as he took the Davidsonian concept of agency as "a necessary and sufficient condition for moral personhood" (French 1979, 215). As he later revised his position, I will only focus on his core claim of *agency*.

> planning, and to plan in the relevant way is to make commitments to perform certain future actions" (French 1996, 150).

So what allows a description of corporate actions to be *intentional*, is that a corporation's CID-Structure creates

> "a general set of policies that are easily accessible to both its agents and those with whom it interacts. When an action performed by someone in the employ of a corporation is an implementation of its corporate policy, and accords with its procedural rules, then it is proper to describe the act as done for corporate reasons or for corporate purposes, to advance corporate plans, and so as an intentional action of the corporation. Corporate plans might differ from those that motivate the human persons who occupy corporate positions and whose bodily movements are necessary for the corporation to act. Using its CID-Structure, we can, however, describe the concerted behavior of those humans as corporate actions done with a corporate intention, to execute a corporate plan or as part of such a plan" (French 1996, 152).

Corporate CID-Structures settle a corporation's plan for doing something. And plans for doing something generate commitments to follow through on these plans. According to French, it is therefore true that *corporate plans* generate *corporate commitments,* which are not reducible to the individual commitments of the group's members. For French, to say that "Corporation X plans to build a new plant in Mexico" does not entail that "all (or any) of the personnel employed by X plan to build anything in Mexico" (ibid). And so French concludes there to be "corporate commitments that are not merely reducible to the plans and commitments of those who directly or indirectly participate in the corporate planning processes" (ibid). If the members of a corporation act in virtue of these corporate commitments, i.e., if the operative CID-Structure allows for the intentions and actions of biological persons to be *subordinated under*, and *synthesized into* corporate intentions and actions, then it is adequate to re-describe these actions as the actions of the corporation. It is on this basis that French holds corporations to be agents.

Assessment of French's theory of corporate agents

Let me provide the reader with some reasons for why I think French's claim that groups are genuine agents might ultimately fail to be decisive. A first thing to notice about French's theory, for which he has been criticized, is his narrow focus on *corporations* as agents, but not on institutional groups in general. Within his own taxonomy (2020), French distinguishes between two broad types of social groups, which he calls "aggregated" and "conglomerate" collectives. According to French, a group of individual people is an aggregate collective "if it is nothing more than a gathering of folks" (French 2020, 13f.). The identity of such an aggregate "is just the sum of the identities of its parts" (ibid). In contrast, conglomerate collectives are groups "such that the identity of the collective is not exhausted by the sum of the identities of its individual members. A conglomerate collective can be composed of disparate types of people with disparate views who bind together by some sort of cementing factor and endure for some period of time" (French 2020,

16). Now in "many cases" (ibid), the way in which disparate types of people with disparate views can be bound together is by "a collective agreement on a decision procedure by which courses of collective action are chosen and tasks relative to the agreed-upon actions are assigned among the membership" (ibid).[22]

The point for which French has been criticized here, is that he falls silent on what exactly makes *corporate* internal decision structures unique in contrast to other conglomerate social group's decision structures. The internal decision structures of *corporations* allow for the intentions and actions of biological persons to be synthesized into corporate intentions and actions. But what about the internal decision structures of other institutional groups, e.g., research groups (see for this criticism: Tollefsen 2015, 55)? Do they allow for the intentions and actions of biological persons to be synthesized into *corporate intentions* and actions too? And can we expand French's view so that just every group that employs just *any* decision procedure therefore is to be counted as an agent? If not, which exact features of corporate internal decision structures are necessary and/or sufficient?

A second worry about French's theory is his somewhat eclectic application Bratman's theory of intentions (1987) as planning states to account for a corporation's agency (see for this line of criticism: Keeley 1981; Rönnegard 2015, 23). As mentioned, French originally subscribed to Davidson's belief-desire model of reasons to argue that corporations are agents. So initially, French tried to analyze corporate reasons in terms of a complex of the corporations *beliefs* and *desires*. On this view, to describe a corporate action as intentionally done by the *corporation itself*, is to describe it being caused by the corporate reasons - as a complex of beliefs and desires - that causally led to the action being performed. He later renounced this view that corporations act on reasons as a pair of *corporate beliefs* and *desires*, because this overly formalized the notions of desire and belief to fit his Corporate Internal Decision (CID) Structure approach (cf. French 1996, 148ff). So French suggests that "corporations cannot, in any normal sense, desire and believe" (ibid). However, French, by subscribing to Bratman's planning theory of intentions, still thinks that corporations can *intend*. It's somehow puzzling why this should be the case. The original idea of Bratman's planning theory of intentions was that intentions, being genuine mental states, cannot be *reduced* to complexes of beliefs and desires. He did not, however, argue that beliefs and desires were thereby not necessary for an entity's capacity for agency. So even if corporations could be said to have (the functional equivalent to) some mental states (i.e., intentions), but not others (i.e., beliefs and desires), it is unclear why corporations could be said to be *agents* in virtue of them having (the functional equivalent of) intentions only. On the contrary, one might hold beliefs to be basic. Tollefsen, for example, suggests that "in order to intend anything at all I have to be a believer. I can't intend to open a window without also having beliefs about what a window is and how to open it" (Tollefsen 2015, 56).

A third problem with French's theory is that it seems to conflate the possibility for a CID-Structure to *represent* intentions to perform an action, with the intentions to perform the action itself. It is one thing for a CID-Structure to represent the corporate policies and procedural rules that make for a corporate plan to perform an action. And while a CID-Structure may in some form contain a *representation* of rules, goals, and

[22] French gives an example of such a conglomerate performing a collective action. He asks us to imagine visitors of a saloon who, upon hearing rumors about cattle thieves, come to be bound by the spontaneously formed commitment to lynch alleged the thieves (cf. French 2020, 15ff.).

procedures, it is quite another thing for individual members in the corporation to adhere to these rules, goals and procedures (See for a related criticism: Hess 2020, 115f.).

Imagine, in analogy, that my grandmother writes me a very detailed shopping-list, with multiple steps to follow and several places to visit (even at different times of the day, so that I can greet the butcher on her behalf). I may derive from this list a certain plan and certain intentions to perform the action of buying the items on the list. Also, we can imagine this to be the case only because the shopping-list of my grandmother represents the plan to do so in question. But why would I need to assume that the shopping-list itself realizes the intention of buying the items on the list? All that is needed is that I, myself, realize the intentions to buy the items on the list. Neither do I need to *subordinate myself under,* or *synthesize myself into* the intentions of the shopping list. The shopping list may *represent* intentions, but it does not posses them. So all the actual intentions and commitments that result from the application of a CID-Structure might actually be realized only by the individual members of the group, but not by the corporation itself. What we are left with, then, are the individual members' intentions and commitments to act in a certain way, which is specified by some set of rules, goals and procedures. But this does not show that these intentions are in any meaningful sense the metaphysically distinct intentions of the corporation itself.

A member *m* may have a certain plan and accordingly certain intentions to perform some action *A* only because he works at a corporation *C* and *C*'s CID-Structure *represents* corporate policies and procedural rules such that *m* is ought to perform *A*. But why would we need for *C* to have intentions, too, in order to explain *m's A-ing?* The reason why French forgoes to answer such questions might be that ultimately, his theory of corporate agency is based on Davidson's *semantic account* of agency (For a related criticism, see: Ouyang and Shiner 1995). According to this analysis, for a corporation to count as an agent, "it must be the case that some things that happen, some events, are *describable* in a way that makes certain sentences true, sentences that say that some of the things a corporation does were intended by the corporation itself" (French 1979, 211; also: French 1996, 151-152). But such a semantic account (only) tells us "when it is a meaningful and useful *use of language to attribute actions* to beings or things. However, a semantic account does not tell us what metaphysical characteristics agents have" (Rönnegard 2015, 26).

In the next section, we will look at Bratman's recent theory of institutional agency. Bratman's own theory is, in many respects, similar to that of French. For one, he also holds the concept of *intentions* to be central in arguing for groups to be agents. Second, he also subscribes to the idea, that *procedural rule-based decision procedures* are central to understand how institutional groups could be said to have such intentions, and thus, agency. Third, he also applies the Frankfurtian view of agency as being *guided* by intentions to institutional groups. However, he does not rely on a Davidsonian semantic analysis, but explicitly argues against Davidson's ideas on agency.

2.1.2. Bratman's Rule-based Account of Group Agency

Michael Bratman (2021; 2022) recently developed a theory of institutional agency, according to which institutional groups, or as Bratman calls them "organized institutions", can act intentionally and hence be viewed as intentional agents (cf. Bratman 2022, xix-xxii).

The (rather complex) argument of Bratman can be summarized like this: Bratman's theory of institutional agency is fundamentally based on his *planning theory of intentions*, according to which intentions are mental representations of future state of affairs that guide and organize the actions of agents. Now this *planning agency* is a "core capacity" involved not only in individual (Bratman 1987) or shared (Bratman 1993; 2009; 2014), but also in *institutional* agency (cf. Bratman 2022, xi). Bratman's guiding thought of this "core capacity conjecture" is this:

> "[G]iven that our planning capacities are central to the organization of our individual agency over time and of small-scale cases of our thinking and acting together, it is reasonable to explore the conjecture that the deep structure of human organized institutions also involves these plan-infused forms of human practical organization" (Bratman 2022, xvi).

The fundamental concept of Bratman's theory of *institutional* agency, in which this core capacity is involved, is the notion of *rules of procedure*. According to Bratman, organized institutions are based on an infrastructure of *authority-according social rules of procedure*. These rules of procedure, in turn, generate *crystallized, action- or acceptance-focused outputs of organized institutions*. In a next step, Bratman argues that these institutional outputs satisfy the functional specifications characteristic of *intentions*, including the function of *coordination-inducing rational guidance*, as well as supporting *consistency* and *means-ends-coherence* (cf. Bratman 2022, 136ff.). Based on his characterization of institutional intentions as c*rystallized, action- or acceptance-focused outputs* of *authority-according social rules of procedure*, institutional groups can be said to be *intentional agents*. This latter claim is mainly supported by negative arguments against the *Davidsonian* view of intentions as belonging to a *dense holistic subject,* and by positively arguing in favor of a *Frankfurtian* theory of an agential standpoint (Bratman 2022, Ch. 8-10).

Social rules of procedure

Bratman takes the planning-capacity to be central for understanding institutional agency. But how is this capacity connected to the above mentioned *social rules of procedure*? What are social rules of procedure to begin with? It's useful to clarify first what *social rules in general* consist of. The basic idea of Bratman is that social rules are expressions of shared intentions, i.e., they are the result of a shared intention to endorse a public social pattern of action (cf. Bratman 2021, 58; 2022, 85f.). According to Bratman, for two individuals to share an intention, both have to have the intention to *J* together, and their *subplans* regarding their individual contributions to this intention to *J* together have to *mesh*, or *interlock*. This, in turn, requires a form of *sensitivity*, or mutual *responsiveness* in one's way of acting, so to not disturb the others from executing *their* sub-plan and vice versa. (see, e.g., Bratman 1993).

Bratman refers to the planing capacity in order to explain how individuals achieve and maintain such coherent interlocking of their shared intentions. This is what he calls the "plan-theoretic construction of shared intentions" (Bratman 2021, 56ff.):

> "For example, if I intend that we paint the house, wide-scope demands of means–end coherence on my intention induce rational pressure on me in favor of helping you, if you need it, in your role in the house painting; and widescope demands of plan consistency induce rational pressure against forming intentions to act in ways that would baffle our joint house painting" (ibid).

Small-scale cases of shared cooperative activities, such as two individuals painting a house, therefore require coordination, and underlying such capacity for coordination is the mentioned *core capacity* for planning. Rational planning figures in Bratman's explanation of shared intentions because it normatively constraints individuals in acting certain ways, i.e. it encourages or prohibits certain ways of acting.

In a next step, and in order to show the "modest and modular place" (Bratman 2021, 56) that such shared intention play in human organized institutions, Bratman merges his plan-theoretical approach of shared intentions with a concept of rules, which he adapts from H.L.A. Hart's theory of law (Hart 1994). A first way to approximate the idea of Hart-type social rules (e.g., taking off one's hat in church) is to contrast them to (mere) social *regularities* (e.g., habitually having tea for breakfast). Whereas social regularities can be analyzed in terms of individual preferences, social *rules* are forms of regular behavior which are based on, and restricted by *other individuals* endorsing the same pattern of action.[23] Social rules, if followed collectively, are based on *shared policies,* which "involve interlocking and interdependent intentions of each in favor of the group's acting in a certain way in a certain kind of circumstance" (Bratman 2022, 20).

So for a social regularity to be a rule for a group, the (or at least *some*) members of this group have to share an intention that the group should conform to this regularity. This, in turn, pressures the individual group members to conform their own behavior to this regularity and to also monitor, help and criticize *others* to conform to it. Social rules, then, involve shared policies and "thereby a kind of shared intentionality continuous with that of small-scale cases of shared intention" (Bratman 2022, 57). Bratman's "Hart-type" model of social rules as "shared policies" then bottoms out in the analysis of rules as social regularities satisfying four interrelated conditions. Rules involve:

(a) a *public social pattern* of action, the full explanation of which involves
(b) *endorsement* by participants of that pattern as a *common standard* that prescribes for the group as a whole, where those endorsements are embedded in a public social web, and where
(c) these embedded endorsements of the pattern support characteristic interpersonal *criticisms, demands, and guidance* of action; and

[23] Here, Bratman draws a parallel between social regularities and social rules on the one hand, and individual action and shared intentional action on the other: "When we walk together on the basis of our shared intention, we each are set to reason, and to reason together, with an eye on our shared end rather than merely strategically in pursuit of personal ends given what the other is doing. And the thought is that the contrast between mere social behavioral regularities and social rules involves a somewhat analogous idea: social rules of the sort of interest involve a kind of sharing that is like that involved in shared intention" (Bratman 2022, 42f).

(d) there is a *rational dependence* of the endorsements in (b) on the general conformity in (a). (Bratman 2022, 85f).

Social rules are only effective if they are being followed. By whom? Bratman calls those to whom a rule applies *participants* in a rule. Importantly, Bratman further differentiates the group of participants, or the "*population-base"* of a rule (Bratman 2022, 121) into a *kernel* and a *penumbra*. According to this distinction, not all members of an organized institution have to share a policy in favor of the group's conformity to a social rule. Rather, those who *do* are to be considered as the *kernel* of participants in a social rule. They are the ones actively enforcing and monitoring rules within a group. They are also the ones who actually share the intention that the group should conform to a given regularity. The others, who Bratman calls the *penumbra* of participants, are those who (merely) intend "generally to conform to R in ways that mesh with the R-conforming behavior of those in the kernel and for reasons that are induced by the kernel" (Bratman 2022, 68). The penumbra of participants is less engaged, or active in enforcing, altering, applying and monitoring rules.[24] They nevertheless "go along" with them. With this distinction in place, Bratman is able to account for cases where individuals in an organized institution are *alienated* and may act for personal reasons only, such as collecting a pay-check (see: Bratman 2022, 89-92). So individuals are either part of the kernel, i.e., they share the intention in favor of the group's conformity to a given set of social rules, or following these rules is something they do because this sort of behavior is expected of them, given the shared policies in place (cf. Hindriks 2023, 5). Importantly, the divide between kernel and penumbra allows for cases where the underlying social rules of procedure is not shared by everyone, but only by a kernel at a given time.

Let me now turn to the central concept of *procedural rules*, which are wider in scope than social rules, and regulate the behavior of participants in a more general matter. Bratman characterizes *rules of procedure* to be functionally relevant for solving certain problems that emerge in (institutionalized or non-institutionalized) groups of rule-participants which try to coordinate their actions. His example is a group of musicians trying to promote their music in the area where they live. This group may already employ and follow multiple, specific social rules, or regulate their behavior according to a social pattern of action. Examples of such rules of the musicians might be: "regularly meeting on Wednesday", "only playing modal Jazz in months that end with the letter y", "not allowing the Kazoo to be played during songs in the minor-key", etc. However, Bratman proceeds that groups will encounter situations in which they require a social decision procedure in order to solve certain problems. In these situations, a "gap in the social rules"(Bratman 2022, 100f.) has to be filled by the participants.

Now in order for groups to respond to such problems, and fill such gaps within their rule-framework, they may employ a *procedure* to come to a solution for this coordination problem:

"perhaps a majority vote among participants, or a Quaker-inspired consensus procedure, or shared deliberation, or deference to a certain subgroup or to those who have a certain status

[24] In analogy to Raimo Tuomela's positional theory (2013) these kernel-participants could be described as *operative* participants in a social rules. In turn, the penumbra, merely going along with the decisions of the kernel-participants, could be viewed in analogy to group members having a *non-operative* participatory status (see Ch. 2.2.3),

> conferred by that procedure or occupy a certain office created by that procedure. This will normally involve the operation of a different kind of social rule—a *social rule of procedure*. And [Bratman's] Peter French–inspired conjecture is that the introduction of such a social rule of procedure is the basic step from a cluster of social rules to a rule-guided, organized institution" (Bratman 2022, 101) [own emphasis].

As a form of "social technology" (Bratman 2022, 149), Bratman says these social rules of procedure "issue in coordinated guidance of temporally extended, socially embedded thought and action" (ibid). Rules of procedure are *mechanisms*, or more generally *means* by which groups can settle certain issues, resolve conflicts or reach consensus on what to do.[25] On the one hand, social rule of procedure can be individuated by their content, which specifies certain procedures to solve certain problems, e.g., whether to have majority-voting on collecting member-fees. But on the other hand, the content of social rules of procedure does not only settle a certain course of action, but also a relevant *follow-through*, i.e., the relevant means by which the procedural decisions are to be realized.

Bratman characterizes the content of a social rule of procedure as follows: "Given relevant intentions to solve a certain problem, (i) proceed in such-and-such ways for arriving at a resolution of that problem, and then (ii) follow-through in such-and-such ways with the outputs of those procedures" (Bratman 2022, 101f). So rules of procedure are a way of making decisions and settling on the means by which these decisions are to be realized, or "followed through".

If rules of procedure are established within organized institutions, and if they are consequently *followed through*, they lead to decisions which Bratman calls *institutional outputs*. These institutional outputs can either be acceptance- or action-focused outputs (cf. Bratman 2022, 103ff.). So the institutional outputs may specify certain actions (A) which are to be performed in order to arrive at a solution of a problem, or these outputs may specify certain propositions (P) to be accepted. The latter is exemplified by a company, whose "procedural-social-rule-generated acceptance-focused crystallized output" (Bratman 2022, 105) lead the members to accept certain propositions about climate change. If these propositions are - via the procedural rule output - accepted, they can function as premises "in further reasoning—individual, shared, and/or institutional—concerning its business practices" (ibid). If the output of a procedural social rule is action-focused, this might specify a particular type of action to be performed by one or several members of an institutional group. But the action-focused output of a procedural social rule might also be more general, not specifying one particular instance of an action to be performed, but rather rule-like *activity-types*: "They can favor a general way of acting in certain general circumstances. Think of an output that says quite generally to wear masks [during a pandemic], or to follow a majority-rule procedure to settle certain issues" (Bratman 2022, 106). To say that such outputs are *crystallized* is to say that they enjoy temporal stability and come about in virtue of a "thick, temporally extended social-psychological web that favors performing A or accepting P" (Bratman 2022, 103f). An acceptance-based output of an institution, e.g., the acceptance of

[25] Social rules of procedure, as a special kind of general social rules, then resemble the above mentioned Corporate Internal Decision (CID) Structures proposed by Peter French (1996) or (premise- or conclusion-based) voting procedures of List & Pettit (2011).

a proposition P, in virtue of being *crystallized,* allows this proposition to figure as a premise in further decision-procedures (cf. Bratman 2022, 105).

Social rules of procedure may not only specify certain procedures by which groups can solve problems, but by being *authority-according* or *-generating*, they can assign positions, functions or statuses to individual group members. A university department's procedural rule for solving a problem concerning student admissions might, e.g., see the establishment of a sub-committee as a way for arriving at a resolution of this problem. Thus, the procedural rule might involve the designation of a subcommittee with a specific function in solving this problem. Importantly, this *authority-according* or *-generating* feature of procedural rules can give individual participants *authoritative powers*.[26] So, e.g., "certain individuals—those who are at a given time members of that admissions subcommittee—are accorded a distinctive status within the processes that respond to this problem" (Bratman 2022, 102). Pertaining to this position or status, the individuals in question can be assigned the authority, the right, duty or permission to perform certain actions (action-focused authority-according rule of procedure), or to accept certain propositions for the group (acceptance-focused authority-according rule of procedure) which corresponds to the (domain-specific) problem they ought to solve. Because solving certain problems might require particular skills, knowledge or experience, social rule of procedure can also establish membership-conditions, or confer roles or positions according to relevant expertise. This, in turn, means that "certain individuals are accorded a distinctive role or status within processes that aim to solve a relevant problem" (ibid).

Procedural rules may not only generate institutional roles and positions, but in virtue of these roles and positions being interrelated, or *interlocking*, they can specify the individuals' interrelations of power, so that some will acquire the *right* to settle certain matters, while the other individuals "will be set (defeasibly) to see some as having accorded duties of deference to authoritative decisions" (Bratman 2022, 107). So individuals occupying certain positions (or offices) are authorized to make decisions to which other individuals, occupying corresponding roles, are to defer to. Given "the hierarchical complexity that is common within organized human institutions", Bratman concludes that "at least normally such institutions will involve authority-according social rules of procedure" (Bratman 2022, 107). Bratman calls this the "*institutional centrality of authority-according social rules of procedure*" (ibid).

Take again the example of an institutional group settling on procedural rules on how to choose amongst different candidates in a hiring process. Here, the group might establish a hiring-committee (inducing institutional positions, offices, and/or statuses), the decisions of which can be counted as institutional *outputs of authority-according social rules of procedure*. And in virtue of the hiring committee being *structured,* the interrelated positions of the committee can be consequently occupied by different individuals. By assigning institutional positions, offices or statuses, *authority-according social rules of*

[26] The notion of authority can be expressed as individuals possessing the relevant rights, duties or permissions to act on their conferred status. Bratman here appeals to Searle's theory of institutional reality, and especially to his concepts of constitutive rules and of *deontic powers* (Searle 1995, 2010). Bratman explicitly states that he aims to "incorporate a version of Searle's emphasis on 'counts as' social rules and associated ideas of institutional roles and statuses" and he makes "room for the Searle-friendly idea that in some cases social rules accord (relativized, *de facto*) authority to those with a certain accorded status" (Bratman 2022, 131). This results in Bratman's claim that "when we turn to complex human organized institutions, […] we do need to put acceptance of ‚deontic powers,' within a given institution, front and center" (Bratman 2022, 113).

procedure generate structures and hierarchies, which in turn, secure an organized institution's structural *robustness.*[27] Given the ways "in which individual participants come and go" such authority-according positions provide "persistence of institutional structures despite change in individual participants" (Bratman 2022, 174). However, such persistence is not only involved in institutional structures, but also in their outcomes. Take again the example of a hiring committee. The positions of such a committee can be occupied by different individuals throughout time, all while the *institutional outputs* of the committee can remain robust, i.e., invariant of the actual individuals who are part of the committee at a given time.

The argument for institutional intentional agency

Bratman's model of rule-guided organized institutions, sees that "social procedural rules in an institutional web normally issue in crystallized outputs that shape downstream thought and action" (Bratman 2022, 135), where these outputs can be action- or acceptance-focused. According to Bratman, these outputs of procedural rules are functionally equivalent to *intentions*. For Bratman, intentions are *plan states*, and the characteristic function of plan states is to provide "temporally extended, cognitively sensitive framing and guidance of thought and action" (Bratman 2022, 156). And as Bratman holds that the *authority-according social rules of procedure* can issue outputs which shape downstream thought and action, in a way that provide such framing and guidance, he concludes that they are the *institutional intentions* of organized institutions (see: Bratman 2022, 135-138). His next step is to move from the claim that organized institutions can realize *institutional intentions* to the claim that organized institutions are *institutional agents.*

His argument for institutional *agency* consists of a number of consequent steps. A straight-forward argument for institutional agency would be for Bratman to conclude from the claim that (1) there is guidance by an institutional *intention,* to the claim that (2) there is guidance by an institutional intentional *agent* who so intends. Rather than applying this straightforward conclusion, Bratman first takes an intermediary step.

Here, he takes it that "we can reason from (1) to [the claim that] (3) there is guidance by an intentional agent" (Bratman 2022, 172). Now to accept (3) (the claim that there is guidance by an intentional agent) is not to say that the institutional group itself is the intentional agent, i.e., that there is guidance by an *institutional* intentional agent who so intends. However, Bratman's goal is to establish just that.

[27] Bratman models this robustness of social rules of procedure in terms of a "thick web of Lockean ties, both over time and across participants" (Bratman 2022, 176) The basic idea here is that, within an institutional group, there will be a sufficient, cross-temporal overlap of individuals following a given set of social rules of procedure that govern the conduct of the institution. This, in turn, secures the persistence of said rules despite the (constant) change of membership of those following the rules within the institution. The appeal to Lockean cross-temporal ties, which secure the robustness of social rules of procedures "involves an analogy with a Lockean view of sameness of person over time. On such a view, a person can persist over time, despite various changes along the way, if there are the needed Lockean interrelations—where these will involve certain constancies of content (as in memory and intention) and cross-temporal connections. And the present idea is that a social rule can persist over time despite various changes in participants at times along the way. This can happen if, despite changes in individual participants, there are the needed Lockean ties across synchronic networks that involve intentions with relevantly constant contents" (Bratman 2022, 77f.).

So in order to arrive at this conclusion, Bratman (Ch. 9) first rejects the view that institutional intentions require for a Davidsonian *densely holistic subject*. On Bratman's interpretation of his work, Davidson "argued that there is agency, strictly speaking, only when there is intentional agency; and intentional agency is the agency of a minded subject, one who is the locus of a dense holistic, broadly coherent web of attitudes" (Bratman 2022, 142ff.; For criticism of Bratman's interpretation of Davidson see: Garcia-Godinez 2023, 840). According to Bratman, and *pace* Davidson, institutional groups can form and hold opinions on only a limited amount of subjects. He gives the example of a medical organization sending supplies to a specific region in the world:

> "Suppose that Medic Supply arrives at an institutional crystallized output that is an institutional intention to send medical aid to C. The proper functioning of this intention will involve a dispersed but in relevant respects unified web of coordinated temporally extended individual and shared intentions and activities, including those of managers concerning the overall activity of sending the supplies and of employees concerning truck driving. This will normally involve a more or less consistent web of intentions, adequate specifications of means, and minimal agreement about what it is to send such aid. *But it need not involve a holistic, Davidson-friendly institutional subject. It need not be settled within the attitudes of Medic Supply whether the aim is simply relief from suffering, or supporting the local government or the local economy, or respecting promises made to donors, or satisfying certain religious obligations.* There may here be substantive disagreements among participants and good reasons for Medic Supply not to take a stand. There can also remain significant disagreement in beliefs about the context of sending the aid (e.g., about features of the target country) and what is involved in sending the aid. *In many cases the institution itself need not take a stand on these matters of disagreement and may be well advised not to*. *Nevertheless* […] *this crystallized output of the social rules of procedure of Medic Supply can be an institutional intention of Medic Supply, functionally speaking*" (Bratman 2022, 157) [own emphasis].

So for Bratman, institutional intentions can be *limited* and they must not require a dense, holistic web of beliefs, or opinions. Hence, he concludes that holism about the mental is not a necessary assumption in order to explain how institutional groups can form intentions. As long as the output of the procedural rules remain effective, i.e., as long as they result in the relevant *follow-through*, substantive disagreement among participants is possible. Also, institutional agents will not necessarily have to form opinions, or "take a stand on these matters of disagreement" (ibid) in order to realize institutional intentions. As such, organized institutions will not meet the requirements for a dense, holistic subject that is the agent of intentional actions. If agency requires such a dense, holistic subject, then institutional groups will fail to be intentional agents. So Bratman rejects the idea that agency requires such a densely holistic subject. But how does he argue from (1) to (2) without invoking such a Davidsonian holistic subject?

In order to arrive at (2), Bratman first contrasts the two cases of *shared* intentions and *institutional* intentions in regard of their *robustness*:

> "Suppose that you and I share an intention to paint the house together. If I were to drop out and some third person were to take my place, our shared intention to paint would no longer exist. Our shared intention is *fragile* with respect to changes in participants. In contrast, an institutional intention will be *robust* with respect to certain changes in participants. Suppose that such an institutional intention successfully guides. So, there is guidance by an intentional agent. Who? Well, this guiding institutional intention is robust in the face of change in individual participants. And an initially plausible principle is (5) *If the intention that guides is robust in the face of change in individual participants, then so is an intentional agent who so intends and thereby guides*" (Bratman 2022, 172).

From this, Bratman concludes that, if guiding intentions *persist despite changes in membership*, and if this guidance by the persisting intentions induces guidance by an intentional agent, then:

> "it seems plausible that that intentional agent will also persist despite these changes in participating individual agents. So, if the intention that guides is robust then an agent who so intends and thereby intentionally guides is robust. But if, given the guidance by an institutional intention, we are looking for a relevant, robust intentional agent, the obvious candidate is the institution itself. So, we can reason defeasibly from (1), (3), and (5) to (2): a relevant intentional agent who so intends and thereby guides is the institution itself. We thereby arrive at (2) by appeal to robustness, but without appeal to a densely holistic institutional subject" (Bratman 2022, 173).

We then have a *prima facie* argument for the existence of an institutional group agent. But Bratman concedes to a possible objection regarding (5). Here, the antecedent, i.e., the claim that an institutional intention may remain robust, may not guarantee the validity of the subsequent claim that there is *a robust agent* who's intending:

> "Why think this guidance by this robust intention constitutes the guidance *by any agent at all whose intention it is*? Why not say that even though, in such a case, the guiding *intention* is robust, there is, quite simply, no robust intentional *agent* whose intention it is and who thereby guides?" (Bratman 2022, 174).

Now, to defend this view expressed in (5), Bratman turns to Harry Frankfurt (see: 1978; 1988). According to Bratman, Frankfurt offers a theory of an *agential standpoint* (see especially: Bratman 2022, Ch.10.2.), which allows him to demarcate his theory of institutional agency from that of Davidson's *dense holistic subject*. Frankfurt distinguishes between first-order, and second-order (or meta-) desires. Whereas first-order desires are at "simply desires to do or not to do one thing or another" (Frankfurt 1971, 7), second-order

desires are desires that are aimed at, or have as their content, first-order desires.[28] According to Bratman's interpretation of Frankfurt, an agential *standpoint* of an individual then is constituted by higher-order desires which are in favor of (not) being moved by first-order desires. According to Bratman,

> "Frankfurt thought that certain structures of will -in particular, hierarchies of desire- are at the heart of where 'the person himself stands' and of acting of one's own free will. And [Bratman's] somewhat analogous thought will be that certain structures of the 'will' of an institution are at the heart of an institutional standpoint whose functioning supports the status of that institution as intentional agent" (Bratman 2022, 175).

Importantly, these structures do not need to be embedded in a Davidsonian dense holistic subject (ibid). Such a standpoint constitutes "non-homuncular, strong form of individual agency", at least if the guidance provided by such a standpoints meets three conditions: "First, the attitudes that constitute the standpoint support a thick web of interconnections over time […] Second, these attitudes rationally shape and anchor basic forms of practical thinking within the system. Third, these attitudes are appropriately stable" (ibid).
If these three conditions are met for institutional groups, then the guidance provided by such a standpoint constitutes an *institutional intentional agent*. Bratman sees that institutional group can fulfill this trio of conditions and that

> "an institution's robust rule-guided procedures for settling important practical issues, and their institutional crystallized outputs, are a central aspect of that institution's standpoint and thereby, potentially, its intentional guidance" (Bratman 2022, 176).

So for Bratman, the "stance-providing, knitting-together, rationally anchoring, organizing roles of robust social rules of procedure and their institutional crystallized outputs" (ibid) fulfill all three conditions. Since these institutional outputs are (the functional equivalent of) intentions of the institution, "we arrive at the conclusion that guidance by institutional intention can help constitute guidance by an institutional intentional agent who so intends" (ibid).

Assessment of Bratman

Now let me begin to assess the theory of Bratman with some remarks about both the goal and scope of Bratman's theory. As we know by now, the target phenomenon that Bratman wants to capture is the agency of institutional groups, his main goal being here to identify an "abstractly specified infrastructure that is common to and important for a wide range of human organized institutions, despite variability across those institutions" (Bratman 2022, 128). Bratman's theory aims to illuminate such forms of an unified institutional infrastructure, "while allowing for different degrees of commonality and of pluralism" (Bratman 2022, 185).

[28] For Frankfurt, second-order desires are distinctively human, and connected to a *persons* freedom of the will: "No animal other than man, however, appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires" (Frankfurt 1971, 7). This point isn't discussed by Bratman.

So the way in which Bratman models the agency of institutional groups allows for there being multiple, and different ways in which the phenomenon of institutional group agency may be realized. His theory therefore only sets out to give *sufficient*, but not necessary conditions for institutional agency. This approach, which he labels the "strategy of sufficiency" (see Bratman 2022, 19ff.; 201), does not amount to the claim that institutional agency *must* be built on top of the core capacity for planning agency. Rather, Bratman only wants to show that the planning capacity *can*, or *may* provide an explanans for institutional agency, with there being the possibility of alternative explanations.

> "In some cases, however, we can ask whether one such realization is more explanatorily fecund than the others. In some such cases, an affirmative answer will take the form of an inference to the best explanation that supports a conclusion that privileges one such sufficient construction" (Bratman 2022, 19).

One of the merits of this strategy of sufficiency is the potential compatibility of Bratman's account with other theories of institutional agency. But this strategy of sufficiency also drew criticism. Frank Hindriks criticizes Bratman's social rule model of institutional agency for leaving undetermined as to which institutional groups (or as he calls them: organizations) the sufficient condition for agency (i.e. having a Frankfurtian standpoint) is supposed to apply to. Under the strategy of sufficiency, Hindriks notes, Bratman's model will apply to a "wide range of institutional agents" (Hindriks 2023, 15), but not to all of them. And while "Bratman does not say so explicitly, this implies that there can be organizations that lack agency. His theory is restricted to those organizations that are agents. This deserves to be highlighted, because the presumption usually is that, if organizations can be agents, all of them will be" (ibid).

Further, the fact that Bratman's theory only gives sufficient conditions for the existence of institutional group agents has been criticized for *being itself insufficient*. Miguel Garcia-Godinez explained that Bratman's project becomes "explanatorily superfluous" by not providing "proponents of other accounts any reason to reject their views, or to recognize his own as better placed to explain the phenomena" (Garcia-Godinez 2023, 839). He concludes that, "in the end, Bratman's proposal is only helpful to those who already have reason to doubt that such other views can successfully explain institutional reality" (ibid). As such, his theory does not provide us with reasons to prefer his account over the other competing alternatives.

Now one reason to hold alternative accounts to offer a better explanation for institutional agency is the relation between the actions of individual group members and the agency of institutional groups. According to Bratman, what truly matters for an explanation of institutional agency is that the authority-according procedural rules of groups give way to robust *outputs* that shape downstream thought and action. Bratman's description of institutional intentions, however, has a lopsided focus on the *outputs* of procedural-rules, while neglecting their *inputs.*

This is problematic for several reasons. A first line of criticism is put forward by Frank Hindriks (2023, 13-15) who argues that Bratman's rejection of the "two Davidsonian dogmas" leads to inconsistencies in his theory. The first "dogma" pertains to the idea of dense holism, i.e., that "intentional agents possess an extended web of closely interrelated mental states" (Hindriks 2023, 13). The second "dogma" rejected by Bratman is that actions are performed *for reasons* and that intentions are *formed on the basis of reasons* (ibid). Now

according to Hindriks, Bratman's rejection of holism, together with the fact that Bratman holds institutions to be capable to form intentions only on a limited range of topics, has problematic consequences for understanding institutional agency. Recall that in cases like the above described medical organization deciding to send supplies to a specific region, there can be disagreement amongst the group members as to why the company sends aid, and "the institution itself need not take a stand on these matters of disagreement and may be well advised not to" (Bratman 2022, 156). According to Hindriks,

> "[i]n such situations, institutional agents form intentions by means of their procedures without there being a reason for which they act. And they might act on such intentions. Thus, it can be that an institutional agent acts but not for a reason of its own […] This view faces a number of problems. First, actions performed for no reason are unintelligible, both for others and for the agent themselves. Bratman seems to bite this bullet and accept that institutional agents are not always intelligible […] The second problem concerns motivation. Why would an intention move the agent to action if it is baseless? Other things being equal, it is rational for an agent to act on their intentions. However, one of those other things is presumably that it had a reason for forming the intention. At least if it knows that it is not aware of any, it is difficult to see how the intention can still motivate or even persist. Third, how can an intention provide rational guidance if it is not based on a reason? And when would it be appropriate to reconsider it? These considerations suggest that rejecting the second Davidsonian dogma is problematic after all" (Hindriks 2023, 14f.).

According to Hindriks, these problems originate from a "common source" which is the description of institutional intentions "only in terms of their downstream effects, to wit their outputs" (ibid) but not in terms of their input. And there is, I think, another reason for why neglecting the *input-relation* of institutional intentions is problematic.

The basic idea I want to put forward against Bratman's theory is that he fails to do justice to the often extensive influence of individual decisions within institutional groups, especially in cases in a) which such decisions of individual group-members guide the actions of institutional groups, where b) the capacity of an individual to issue such decisions, and thus exert such an influence, can be traced back to authority-according rules of procedure. Cases where individuals exercise extensive influence on institutional actions are problematic, at least insofar as institutional intentions are supposed to be *robust* with respect to changes in participants. To put it somehow bluntly, I think that Bratman's theory runs into the risk of *mis-identifying* the intentions of powerful individuals with the intentions of groups they have power over. Yet, Bratman's output-focussed analysis would render (or rather: misrepresent) such individual decisions as *institutional* intentions in virtue of them being procedural rule-based, authority according, acceptance- or action-based crystallized outputs.[29]

[29] Which in this context would mean that the decision of such a dictatorial individual is embedded in a "temporally extended social-psychological web" (Bratman 2022, 103f.) of rule-participants that favors this individual performing action A or accepting proposition P.

Recall that on Bratman's view, institutional positions and offices are *means* or *vehicles* by which groups solve coordination problems. By being authority-according, procedural rules can assign deontic powers to offices or positions and thereby establish an institutional group's hierarchical *structure*. But if we follow through on this idea, then the assignment of institutional positions and offices in virtue of authority-according rules of procedure ultimately implies that the decisions of just *one* individual group-member, holding a powerful position in a given institution, can settle an institutional group's intention, which guides its "downstream thought and action". Let me elaborate on my criticism by first discussing an "extreme" case of "dictatorial" groups and then, in a next step, by extending this criticism.

The first, and possibly most ostentatious case for such a mis-identification of a powerful individual's intention with the intentions of the group, which such an individual has power over, is the case of what List & Pettit call "dictatorial groups" (List & Pettit 2011, 36). In dictatorial groups, the members

> "authorize a 'dictator' to form the group agent's attitudes, without doing anything more active on the group's behalf […] The dictator may then be the sole member acting for the group, with other group members being merely complicit in the group agent's existence, under the dictator's direction" (ibid).

Here, the authority-according procedural rules of a *dictatorial* institutional group may specify that only *one* individual's dictatorial position of power issues in its institutional output. In cases of such a dictator ruling supreme, an institutional intention, i.e. the output of a procedural rule that shapes downstream thought and action, can be traced back to just *one* individual. Also, such an individual may have the power to decide - in his "*L'État, c'est moi!*" kind of governance - on a relevant "follow through". But in such a case, I think it is apt to identify *that particular individual* (and not the whole group) as the agent who issued the output of a procedural rule that shaped the group's downstream thought and action, *simply because said individual provided the only input to it.* I tend to agree with List & Pettit that a group action constructed around a 'dictator' can be seen as "just *an extension of that individual's agency* rather than as a group agent proper" (List & Pettit 2011, 59) [own emphasis].

We can now easily extend such dictatorial cases by first noticing, that "dictatorial groups" are not merely freakish anomalies of the social realm. In fact, a group centered around one "dictator" might just be one extreme example of something that occurs within a large class of institutional groups with structures involving relations of power. And many institutional groups will, to some degree, in fact exhibit such "dictatorial" features. This becomes apparent if we notice that paradigmatic institutional groups, e.g. companies, military units, political parties, labour-unions, governmental agencies, etc., tend to be structured by both *vertical* and *horizontal relations of power*.[30] By saying that an institutional group's power-structures are *vertical,* I mean that the relations between individual members include authority-based chains of command and relations of *authorization of*, and *deference to* deontic powers (i.e., *being someone's boss,* and giving orders to others; or *having a boss*, whose *orders* one is to defer to). *Horizontal power*-structures, in contrast, do not depict interpersonal relations of authority, but rather functional

[30] I will give a more detailed analysis of such power-structures of institutional groups in Ch. 2.2.3. and in Ch. 3.1. and 3.2.

specifications of domains, departments, or sectors. This means that within *horizontally*-structured institutional groups, specialized sub-groups are assigned different tasks and assignments. Companies, for example, are horizontally structured if they have specialized departments, e.g., for *HR*, *Sales*, *R&D*, *Marketing,* etc. Now within horizontally-structured institutional groups, members may authorize different individuals to form the groups attitudes regarding *different domains and subjects*. A group, e.g. may appoint an individual each to be head of *HR*, *Sales*, *R&D*, *Marketing*, *Accounting, Finance*, *Operations*, or *Administration,* etc. And such individuals may each have the power to ultimately decide on a relevant "follow through" regarding their domain-specific area.

So once we depict the authority-according structures of institutional groups in a more broad and de-centralized way, we notice that institutional groups can have many, independent, *domain-specific* "dictators" which each reign with their own domain-specific authority. But in such cases, again, I think it is apt to identify *those particular* individuals - and not the group itself - as the agents who issued the output of a procedural rule that shaped the group's downstream thought and action. Why? *Because, again, those individuals may have provided the only input to it*.

When explaining the agency of institutional groups, we ought to incorporate notions of hierarchy and power into our theory of how, and why institutional groups could be said to perform actions the way they do. Bratman's characterization of authority-according rules of procedure giving rise to offices, positions and hierarchically related *status roles* surely does justice to this feature of institutional groups. But recognizing the role of hierarchies, power-structures and authority-based decisions in institutional groups ultimately reveal to us how individuals exert their influence in the decisions, goals and actions of institutional groups such individuals have power over. Sentences like "It was *Elon Musk*, who decided to change Twitter's name to ‚*X*'" or "*Trump* ultimately decided for the USA to withdraw from the Iran Agreement" clearly seem to suggest that we can make institutional actions intelligible by understanding them *to be based on the decisions and judgement calls of individuals,* especially of those individuals in power. And they also suggest that often times, such decisions can be made on a whim, and reflect the idiosyncratic and erratic intentions of those individuals in power. This is because institutional roles, such as *being a CEO*, come with varying degrees of individual autonomy to make decisions. This feature of institutional roles is usually captured in terms of *discretionary powers* vested in individuals occupying institutional roles.[31] Such discretionary powers, however, cast doubt on Bratman's argument for institutional intentions to be *robust* in the face of change in individual participants. The decisions, plans, and goals of institutional groups *can* and in fact *do* often change in virtue of particular group members joining, or leaving the higher echelons. Thus, the invariance of output should not be seen as necessarily given, as the assignment of power to a "dictator" can undermine such robustness.

All of this is not to say that the *outcome* of such decisions is something that can never be attributed to the institutional group itself, or that the decisions of powerful individuals are ultimately not based on decision-making processes or rules of procedure. But I hold Bratman's theory to downplay, or diminish the role that individual decisions, especially of those "on top" of institutional groups, play in bringing about such outcomes. Looking only at the downstream effect of such decisions and attributing these outcomes to the

[31] A full discussion of the relation of institutional roles and discretionary powers, as well as the implications for group agency is provided in Ch. 4.

institutional groups themselves, leads us to disregard the actions and decisions of -potentially powerful- individuals in institutional actions. We then run the risk of, as Manuel Velasquez put it, waving "our hands before the corporate veil" (Velasquez 1983, 15), incapable to "travel behind the corporate veil to lodge with those who knowingly and intentionally bring about the corporation's acts" (ibid).

2.1.3. List & Pettit's Theory of Group Agency

Let us now take a closer look at another, influential account of group agency. Just like French's account, Christian List's and Philipp Pettit's account of group agency initially stems out of the debate about corporate moral responsibility.[32] The prominent account has escaped the realm of practical philosophy ever since. Their goal is to establish a *realist*, non-reductive account of group agents. Accordingly, their main claim is that there are group agents "in their own right, as it is often said, groups with minds of their own" (List & Pettit 2011, 77f; see also: Pettit 2003).

The way they argue for the existence of genuine group agents[33] is somewhat intricate but boils down to this: There are certain ways (called *aggregative functions*) in which groups generate attitudes by aggregating the individual attitudes of their members, where the resulting attitudes on the group-level do not seem to be reducible to those of the individual members. Such ways in which these group-level attitudes are generated also meet so called *rationality constraints,* and they uphold a relation of *holistic supervenience*. On this basis, List & Pettit argue for the autonomy of groups regarding their capacity to form intentional states. And this autonomy, in turn, is crucial for their claim that groups are agents.

Although List & Pettit somewhat provocatively claim that groups have are agents with a "mind of their own", it should be made explicit from the get-go what this really amounts to. *Group agents* in the sense of List & Pettit do not have consciousness, qualitative states or visual, auditive, tactical experiences, etc. A group agent in the sense of List & Pettit is an agent that consists of individual persons (which may have all of these features) (cf. List & Pettit 2011, 8ff.). What allows those individual persons to comprise one numerically distinct agent is that the individuals in question have suitable coordinated dispositions to think and act such that the properties of the group *holistically supervene* on the dispositions of the members. Group agents then are not *ontologically* autonomous entities, yet, List & Pettit claim that they are autonomous in a different sense:

> "The agency of the group relates in such a complex way to the agency of individuals that we have little chance of tracking the dispositions of the group agent, and of interacting with it as an agent to contest or interrogate, persuade or coerce, if we conceptualize its doings at the individual level […] The autonomy we ascribe to group agents under our approach is epistemological rather than ontological" (List & Pettit 2011, 76).

[32] Unless otherwise indicated, I will follow the exposition in List & Pettit 2011. See also: Pettit 2003; 2007; Pettit & Schweikard 2006; List 2018. For critical discussion see: Tollefsen 2015; Paul 2020; Townsend 2013; 2020.

[33] The authors ultimately argue for an even stronger claim, namely that groups are organizational or institutional *persons* (List & Pettit 2011, Chapter 8). I do not asses their claims here as I am not concerned with group personhood but group agency.

So what the authors really want to establish is *epistemological autonomy,* which Pettit also calls "psychological autonomy" (Pettit 2003, 167). To better understand List & Pettit's claim that groups can be genuine agents, we may first look at their underlying concept of an agent. Their concept includes three key features (F1-F3) and corresponding "standards of rationality" (SoR). In order for some entity to count as an agent, it has to have:

F1: Representational states
F2: Motivational states
F3: The capacity to process its representational and motivational states
SoR: attitude-to-fact; attitude-to-attitude; and attitude-to-action standard of rationality.

The first two features link the capacity for agency to intentional states.[34] The third feature describes a potential agents capacity to "intervene suitably in the environment whenever that environment fails to match a motivating specification" (List & Pettit 2011, 20) The standards of rationality, List & Pettit argue, account for the agent to be a *rational* agent, i.e. have the capacity for minimal rationality. According to this minimalistic framework of agency, any entity is an agent if it exhibits these three features of agency and meets the minimal rationality constraints that come with them (cf. List & Pettit 2011, 32f).[35]
We can, on this basis, assemble a group agent. A group agent

> "is a group that exhibits the three features of agency, as introduced above. However this is achieved, the group has representational states, motivational states, and a capacity to process

[34] Representational states, depicting "how things are in the environment" are those intentional states, that have a mind-to-world direction of fit, such as, e.g. beliefs or assertions. Motivational states, which are informative of how an agent wishes or requires things to be can be seen as intentional states with a world-to-mind direction of fit. Regarding F1 and F2, List & Pettit conceive these intentional states to be "configurations in an agent's physical make-up that play a particular role or function in engaging with other such states and in producing action" (List & Pettit 2011, 21). With such a functionalist conception of intentional states, List & Pettit do not make assumptions about the precise physical nature of intentional states, which "may be electronic or neural configurations of the agent, for example, depending on its robotic or animal nature. They may be localized in the agent's brain or central processing system or dispersed throughout its body" (ibid). The only requirement for these states (or configurations) to count as intentional states is that they "play the appropriate functional role" (ibid).

[35] For this to be the case, the agent must meet i) *attitude-to-fact* standards regarding the capacity to accurately form and maintain truth-tracking or veridic representations of how things actually are, but also the capacity to establish, follow and monitor rules for cases in which the forms of representation do not accurately track reality, such as states of delusion, paranoia etc. (cf. List & Pettit 2011, 24f). To count as minimally rational, an agent further has to meet the ii) *attitude-to-attitude* standards of rationality, which concern the capacity for creation and maintenance of coherence in one's attitudes (e.g., not simultaneously believing p and not-p) or to reach deductive closure (e.g., believing "p" and believing "p → q" also entailing believing "q"). Rationality also entails iii) *attitude-to-action* standards which roughly concern the coherence of connection between one's reasons and one's actions. If, e.g., Alf has reasons to eat the Turner family's cat Lucky (which we may conceptualize here as a mix of representational and motivational states of Alf concerning Lucky) but his reasons cause Alf to eat Spaghetti instead, his capacity of acting rationally in light of the attitudes he has can be said to be corrupted or disturbed. Repeated occurrences of such disruption can ultimately lead to the breakdown of an agents capacity to act rationally (not to act though) (ibid).

> them and to act on that basis in the manner of an agent. Thus the group is organized so as to seek the realization of certain motivations in the world and to do so on the basis of certain representations about what that world is like" (ibid).

A group will count as an agent if the group will have the capacity to form rational and coherent intentional states by aggregating them in a specific way from the individual-level attitudes of its members. The group, List & Pettit argue, must

> "ensure that whatever beliefs and desires it comes to hold, say on the basis of its members' beliefs and desires, form a coherent whole. This is certainly the case with attitudes close to action: those desires on which it acts and those beliefs on which it relies for directing its action. Otherwise the group won't be able to enact its intentional attitudes in the manner of a well-functioning agent" (List & Pettit 2011, 32).

A group can now fulfill the necessary requirement for rationality by employing aggregative functions that breach "certain initially plausible conditions on how its attitudes relate to those of its members" (ibid). By choosing those aggregative functions, the group is organized

> "so as to seek the realization of certain motivations in the world and to do so on the basis of certain representations about what that world is like. When action is taken in the group's name – say, by its members or deputies – this is done for the satisfaction of the group's desires, and according to the group's beliefs" (ibid).

So how are groups agents in this sense, i.e. how do they realize the above mentioned features of agency? List & Pettit argue that groups can "collectivize reason" which leads a group "to perform in the manner of a reason-driven agent by deriving its attitudes on conclusions […] from its attitudes on relevant premises […]" where this is established in a way "guaranteeing that the attitudes of the group are consistent as well as derived in a rational process" (List & Pettit 2011, 58f.). For a group to "collectivize reason" is for this group "to aggregate the intentional attitudes of its members into a single system of such attitudes held by the group as a whole" (ibid). Such a single system of attitudes is the group's locus of agency. Groups, in turn, are agents insofar as they establish and maintain such ways of collectivizing the reasoning of their individual members. So let us next look at how this may come about.

Aggregation functions

The key here is that groups can establish certain decision procedures, which they call *aggregation functions*. Those aggregative functions concern the connection between the individual members attitudes regarding some propositions and the way those attitudes are "collectivized" into a single system on the group-level.

Roughly, aggregative functions concern the relation of individual attitudes as input and group-level attitudes or judgments as outputs (cf. List & Pettit 2011, 48f.)

Now there are many ways in which a group can arrive at a judgement and not every way a group does so makes it eligible to talk about collectivizing reason. How can we separate the grain from the chaff here, i.e. how can we tell which aggregative functions are appropriate? Not all aggregation functions are able to meet four conditions which List & Pettit identify as necessary in order for a group to meet the rationality constraints described above. These four conditions are the 1) Universal Domain, 2) Collective Rationality, 3) Anonymity and 4) Systematicity.[36] Now List & Pettit argue that any group can only realize three out of the four conditions, and they further argue that for *robust group rationality* to be fulfilled, the fourth condition of systematicity has to be abandoned, or relaxed (cf. List & Pettit 2011, 50ff; 67ff.)[37].

So how does such an aggregation function look like? List & Pettit (see also: Pettit 2003) discuss the so called *discursive dilemma* to showcase such a function. The dilemma concerns the relation between a group's judgments on the one hand and the individual judgments of the members of the group on the other hand regarding a certain set of propositions. Regarding this relation, a so called *doctrinal paradox* (List & Pettit 2011, Ch. 2) can arise. The doctrinal paradox aims to show that one and the same group of individual people can - as a group - arrive at different conclusions regarding the set of premises, depending on how the individual judgements of the members are to be aggregated. So the judgments of a group can vary while at the same time, the individual members do not vary in the judgments they give.

This variation of output stems out of the way in which those individual judgments (the input) are aggregated into a group judgement (the output). There may be two different aggregation functions employed: For one, conclusion-based functions and second, the (slightly more sophisticated) premise-based aggregation functions. Both functions have in common that the group accepts that majority votes settle

---

[36] Those four conditions are characterized this way: 1) Universal domain: The aggregation function admits as input any possible profile of individual attitudes towards the propositions on the agenda, assuming that individual attitudes are consistent and complete. 2) Collective rationality: The aggregation function produces as output consistent and complete group attitudes towards the propositions on the agenda. 3) Anonymity: All individuals' attitudes are given equal weight in determining the group attitudes. Formally, the aggregation function is invariant under permutations of any given profile of individual attitudes. 4) Systematicity: The group attitude on each proposition depends only on the individuals' attitudes towards it, not on their attitudes towards other propositions, and the pattern of dependence between individual and collective attitudes is the same for all propositions. (cf. List & Pettit 2011, 49.).

[37] The weakening of the systematicity-condition concerns the supervenience relation of group-level and individual-level attitudes. What does weakening the systematicity condition amount to? According to List & Pettit, the systematicity of group attitudes consists of two interrelated parts: Independence (the group attitude on each proposition depending only on the individuals' attitudes towards the this exact proposition, not on their attitudes towards other propositions) and neutrality (the pattern of dependence between individual and group attitudes being the same for all propositions). By weakening both parts of the systematicity-condition, List & Pettit allow for a group to rationally aggregate the intentional attitudes of its members into a single system of such attitudes where this process of aggregation can prioritize some propositions over others (weakening neutrality) and letting the group attitudes on a first set of propositions determine its attitudes on the latter (weakening independence). Aggregative functions, in order to give rise to rational group attitudes, then must admit as their input only those individual attitudes which are consistent and complete and produces as their output only those group attitudes that are consistent and complete as well. Further, this way of arriving at group attitudes is democratic: All of the individuals' attitudes are given equal weight in determining the group attitudes and no one member is more important in setting the group's attitude than any other.

official positions of a group. The crucial question is *where* to apply the majority votes: on the conclusion or on the premises? Their standard example is a hypothetical panel of judges (see also Pettit 2003) deciding on whether a defendant is liable for a harm she caused (see Table 1 below). The logical form of the decision is: P1 $\wedge$ P2 $\rightarrow$ C. In this case, the defendant is liable for the harm done (C) only if she both P1) was the cause of harm and if she P2) also had a duty to care about the harm done.

	P1 Cause of harm?	**P2** Duty to care?	**Conclusion:** Liable for harm?
A	Yes	No	No
B	No	Yes	No
C	Yes	Yes	Yes
premise-based procedure:	Yes	Yes	Yes

Table 1: Matrix 1 (from Pettit 2003, 168)

The conclusion-driven aggregation of individual judgements works like this: Let the judges reason individually whether P1 and P2 and then let the majority vote of their individual judgments regarding the conclusion rule the verdict. In Table 1, individual *A*'s judgment would not see for the defendant to be liable, as she denies that the defendant had a duty to care (P2). Same goes for individual B, who thinks that, although the defendant had a duty to care, she was not the cause of harm (P1). Only individual C sees the defendant guilty, as she both judges her to be the cause of harm *and* her to have had a duty to care. Read this way, then, A and B hold the defendant to be innocent while C sees her to be guilty. Per majority-vote on the *conclusion* (2 Judges: not guilty/ 1 Judge: guilty) the group would settle for the defendant to be acquitted.

Contrast this with the premise-driven aggregation function of individual judgments. Premise-driven aggregations work like this: Let the judges reason individually whether P1 and P2 and let the group's attitude *on each premise* be settled by a majority vote on it. In this case, the majority of judges think that the defendant is both the cause of harm (A and C: yes vs B: no) *and* that she had a duty to care (A: no vs. B and C: yes). In a second step, the function sees it that the accepted premises lead to the conclusion by means of deductive closure. The defendant is found liable for the harm caused.

According to List & Pettit, conclusion-based functions do not fulfill the conditions necessary in order for them to give rise to rational group attitudes.[38] Does the premise-based aggregative function fulfill the conditions necessary in order for it to give rise to rational group attitudes? First of all, it admits as input any possible profile of only individual attitudes towards the propositions on the agenda, which here translates to admitting as input the belief that P1, P2 and the conclusion C are either true or false. Second, all individuals' attitudes are given equal weight in determining the group attitudes. For consistency, it also passes the test: Adopting this procedure, the judges *as a group* will endorse a consistent conclusion because the majority of the judges sees P1 and P2 as satisfied and follow through on the conclusion, inferring C from P1 and P2. Yet, if the judges adopt this procedure, they - as a group - apparently endorse a conclusion which a majority of them individually reject.

This is quite astonishing, as List & Pettit seem to have shown that individual and group attitudes can come apart this way. This premise-based procedure, List & Pettit claim, "lead a group to perform in the manner of a reason-driven agent by deriving its attitudes on conclusions (or 'posterior' propositions) from its attitudes on relevant premises (or 'prior' ones)" (List & Pettit 2011, 58ff.). This is what Pettit called the "imposition of discipline of reason at the collective level" (see Pettit 2003, 175). The above mentioned *collectivization of reason* therefore describes the process of aggregating intentional attitudes of group members into a "single system of such attitudes held by the group as a whole" (List & Pettit 2011, 58ff.) where this procedure guarantees "that the attitudes of the group are consistent as well as derived in a rational process" (ibid).

While procedures like the conclusion-based approach maximize the influence of, and responsiveness to *individual* judgments, List & Pettit claim that a group will have to settle for aggregation functions like the

[38] First of all, they admit as input any possible profile of individual attitudes towards the propositions on the agenda, which here translates to admitting as input the belief that P1, P2 and the conclusion are either true or false. Second, all individuals' attitudes are given equal weight in determining the group attitudes. Notice, however, that this function does not produce consistent group attitudes towards the propositions: Adopting this procedure, the judges together will endorse an *inconsistent conclusion*. Why? Because the majority of the judges sees P1 and P2 satisfied. Yet -*as a group*- they do not follow through on the conclusion that C from P1 and P2. As a group they do not reach the above mentioned deductive closure, necessary for attitude-to-attitude rationality: If the *group* judges that P1 and P2, and knows that this entails C, then the group *also ought to judge that C*. Yet, they don't. While the majority of the group sees that the defendant was the cause of harm and a majority of the group sees it that she had a duty to care, the majority of the group does not judge the defendant to be guilty, although her guilt follows from both of the accepted propositions. Importantly, the judges fail to rationally reason *as a group,* but not individually. The inconsistency resides exclusively on the group-level as every individual judge is consistent in *her* judgment. Rationality on the individual level remains intact, but it's somehow lost when we look at the group-level. Individual rationality plus this sort of aggregation function then does not, on its own, secure *collective* rationality.

premise-based approach in order to promote its causes, for groups will not be effective in promoting their goals if they tolerate (too much) inconsistency in their judgements (see, e.g., List & Pettit 2011, 40f).[39]

Holistic supervenience

The last thing that needs clarification is which explanatory role the notion of "holistic" or "set-wise" supervenience plays in getting us from the observation that groups establish ways to rationally and consistently settle on judgments, to the claim that groups are agents in their own right. Recall that List & Pettit define a group agent as an agent that consists of individual persons (cf. List & Pettit 2011, 8f.). What makes these individuals comprise one numerically distinct agent is that they have suitable coordinated dispositions to think and act in such a way that the properties of the group *holistically supervene* on those dispositions. We've just seen how adopting a premise-based decision procedure describes such a way in which individuals coordinate their dispositions to think and act. But how do the group-level properties *supervene holistically* on those dispositions?

List & Pettit want this relation of supervenience to be modeled as an aggregative function, which meets the standards of rationality, i.e., consistency and deductive closure (ibid). They call this the *requirement for robust group rationality*:

> **Robust group rationality**: The supervenience relation determines consistent and complete group attitudes on the relevant propositions for any possible profile of consistent and complete member attitudes on these propositions (List & Pettit 2011, 67).

For the supervenience relation between individual- and group-level attitudes to secure robust group rationality, two conditions must be fulfilled:

> "First, the supervenience relation must determine the group attitudes on the relevant propositions for any possible profile of members' attitudes. Unless this condition is met, the relation is not one of supervenience – that is, of necessary determination – but only one of contingent determination. Second, the group attitudes determined by the supervenience relation must robustly meet the appropriate attitude-to-attitude standards of rationality, such as consistency and completeness. Unless this second condition is met, the supervenience relation does not secure the robust presence of group-level agency" (ibid).

[39] A group can also faced with pressures of rationality *diachronically* when it is faced with decisions it made in the past which now have to figure as the premises on how to decide on another issue. Deciding in consistency with past judgements can therefore constrain the judgments of a group for new cases (resembling somewhat the role of commitments in the individual case). Pettit here identifies cases, where, if at t_1 group g per majority-vote settles on an issue A, g must take A to be a premise for further reasoning at t_2. In March, e.g., a governmental party may decide that it needs to tackle the unbalanced budget by *cutting* spendings. In November, the party is asked again whether to endorse some tax reform which would *increase* spendings. In November, the party cannot - at least ceteris paribus - endorse the *increase* of spendings based on the decision it made in March, for this would be detrimental to making consistent and coherent decisions. This is not to say that the group cannot vote in favor of the tax reform in November, but only that it would be inconsistent with the group's *previous* decision to cut them (cf. Pettit 2003, 172ff).

List & Pettit claim that so called *proposition-wise supervenience* won't suffice to fulfill the condition of robust group rationality.

> **Proposition-wise supervenience:** The group attitude on each proposition is determined by the individual attitudes on that proposition, where the mode of determination may differ from proposition to proposition (List & Pettit 2011, 67-68).

To secure robust group rationality, the "group's attitude on a particular proposition cannot generally be a function of the members' attitudes on that proposition" (List & Pettit 2011, 69). Rather, they must be modeled as "more complex" (ibid). In an attempt to account for such rationality in group attitudes, List & Pettit now introduce their notion of *holistic* supervenience:

> **Holistic supervenience**: The set of group attitudes across propositions is determined by the individual sets of attitudes across these propositions (ibid).

Holistic supervenience is consistent with robust group rationality, at least if certain aggregation functions to generate group-level attitudes are employed. The premise-based procedure, while leading to a breach of *proposition-wide* supervenience, is such an aggregation function which secures robust group rationality. And this breach of proposition-wide supervenience "shows that individual and group attitudes can come apart in surprising ways, thereby establishing a certain autonomy for the group agent" (ibid). List & Pettit consider two cases of a premise-based procedure regarding the question whether $P \wedge Q$ (see Table 2 and 3).

	P	**Q**	**P ∧ Q**
Individual 1	True	True	True
Individual 2	True	False	False
Individual 3	False	True	False
Premise-based procedure:	True	True	**True**

Table 2: A profile of individual judgments (List & Pettit 2011, 70)

	P	**Q**	**P ∧ Q**
Individual 1	True	True	True
Individual 2	False	False	False
Individual 3	False	False	False
Premise-based procedure:	False	False	**False**

Table 3: A modified profile of individual judgments (List & Pettit 2011, 70)

Notice first that in both scenarios of Table 2 and 3, the premise-based procedure gives logically consistent results: In both cases, deductive closure is reached on the individual- and group-level. No individual is behaving irrational in either of these two scenarios. And on a group-level, Table 2 shows that the group concludes the *truth* of whether P ∧ Q from the truth of the group-level attitudes towards the two premisses being true. But in Table 3, the group concludes that P ∧ Q is *false* from the group-level attitudes towards the two premisses being false. So under the premise-based procedure, the group may form opposing judgments towards the proposition whether P ∧ Q *while the individual judgments whether P ∧ Q do not change in both scenarios*. In both scenarios, each individual 1, 2 and 3 arrives at the same individual conclusion whether P ∧ Q is true or false: In both Table 2 and Table 3, Individual 1 holds P ∧ Q to be true; and Individual 2 and Individual 3 hold P ∧ Q to be false. So the proposition-wise relation of supervenience is breached, because the group's judgment whether P ∧ Q is not determined by the individual judgments regarding whether P ∧ Q alone. Hence, to make sense of the group giving logically consistent, and rational results, List & Pettit argue that group-level attitudes must be modeled to supervene *holistically* on the individual attitudes, i.e., the group's attitude towards a proposition (whether P ∧ Q is true or false) is determined not only by members' attitudes towards *that* proposition, but by their attitudes to other, logically related propositions (whether P is true or false; and whether Q is true or false) (see for further elaboration: List & Pettit 2011, 69-72).

Assessment of List & Pettit

Having presented the account, we are now in a position to assess List & Pettit's claims. Because the account has been subject to strict and extensive criticism elsewhere (see, e.g., Smith 2012; Tuomela 2013, Ch. 5; Miller 2001; Miller and Mäkelä 2005; Townsend 2013; 2020; Sylvian 2012), I will focus on a more specific aspect, i.e., List & Pettit's distinction between group- and individual-level attitudes.

A first way, however, to criticize List & Pettit is to assess both the scope and breadth of the account. The doctrinal paradox was exemplified by the case of a panel of judges deciding in a court case. The underlying problem, according to List & Pettit, is not confined to the legal realm but arises in every group that has to make judgements where majorities in a group each support one premise but different majorities support different premises, while those majorities do not overlap (cf. Pettit 2003, 169). The claim is that any structured and organized group will - at some point in their existence - encounter the problem of how to derive at a judgement based on given premises. The question now arises as how to position this theory within the debate of *institutional* group agency.

Within their own conceptual apparatus, List & Pettit merely distinguish what they call "collectives" from "groups", where the latter are distinguishable from the former because they "have an identity that can survive changes of membership" (List & Pettit 2011, 39). Accordingly, List & Pettit identify a great variety of groups, which may count as institutional groups under our preliminary definition. The examples of List & Pettit include trade-unions; political parties, commercial entities like corporations, or civic entities like chess clubs, universities, churches or cycling societies (cf. List & Pettit 2011, 40). All such groups, on the basis of the mechanisms described above, can potentially be group agents (insofar as they collectivize the reasons

of their members to install group-level rationality etc.). Collectives, on the other hand, are taken not to be agents in their own right (ibid).

Now, List & Pettit focus their theoretical explanation of group agency on what they call *democratic* groups, which engage in specific forms of deliberation and decision-making, thereby excluding (the above mentioned) "dictatorial" or authoritarian groups where one member (or just a few in the higher echelons) gets to make judgements on the group's behalf (see: Pettit 2003, 170). These "dictatorial", or non-democratic ways of decision making then are then simply identified as *degenerate* cases of group agency (cf. List & Pettit 2001, 59). As mentioned above, List & Pettit go so far as to regard dictatorial groups as a mere "extension of that individual's agency rather than as a group agent proper" (ibid).

Now as to criticize the exemplificatory breadth of the account, it is not obvious why such democratic groups should function as the paradigmatic entities on which to build a theory of institutional group agency. It is also unclear why hierarchically organized types of groups should be regarded as "degenerate" cases. One could also make the case that they, in fact, are the norm. Corporations, for example, are by any means one of the most discussed type of groups in the context of institutional group agency. However, corporations are usually thought to be groups that base their decisions on authority structures, power relations and membership-positions, and not purely on democratic deliberation. If the phenomena of hierarchical organization and authority based decision structures are far more common for groups than List & Pettit are willing to admit, it is unclear why democratically organized, deliberative groups should function as the paradigmatic case for establishing a theory of group agency (see for a related criticism about List & Pettit's lack of incorporating *multiple* ways of arriving at decisions: Hess 2020, 118f.).

But even within their own explanatory scope, objections can be made. One main line of criticism can be summarized like this: The only processes of reasoning involved in the cases presented by List & Pettit are processes of individuals. Therefore, their account falls short of what it wants to establish. Recall that List & Pettit motivate their analysis of autonomous group attitudes by claiming to have shown cases of group agents holding attitudes, where we have "little chance of tracking the dispositions of the group agent, and of interacting with it as an agent to contest or interrogate, persuade or coerce, if we conceptualize its doings at the individual level […]" (List & Pettit 2011, 76).

But do we really need to invoke the claim that there are autonomous group agents with a mind of its own for even a simple case of, e.g., a group of three friends adopting the premise-based decision procedure for settling where to get lunch? The examples of List & Pettit aim to show how groups can arrive at rational conclusions, where yet none of the group members came to these conclusions on the basis of their individual reasoning. So far so good. But in a next step, they infer from this the claim that there are processes of collective reasoning which are irreducible to the members of the group. This, I think, is where the argument goes off the rails, because such a distinction between group- and individual-level attitudes can be shown to be only *apparent*.

It has been argued (by e.g., Miller 2001; Miller & Mäkelä 2005; Mäkelä 2007; Ludwig 2017a, Ch. 12.2.) that these examples only show all the members of a group *to reason individually under an established mechanism* in order to aggregate their results. Either an individual derives from any given set of premises an individual conclusion, which is then subjected to a voting procedure, the outcome of which determines the group's view (the conclusion-based approach). Or the members can individually settle on the premises

and then apply a voting procedure which establishes fixed group premises, while the logical consistent *conclusion* is derived from the voted-on premises (premise-based approach). Notice here, that the premises from which the conclusion is inferred are determined by voting. Still, each member of the group can *individually infer this conclusion* from the agreed-upon premises. So while either the premises or the conclusion can be determined by voting, all this really illustrates is that groups can adopt different ways of arriving at decisions. This does *not* amount, however, to the collectivization of reason compelling us to accept the existence of epistemologically or psychological autonomous entities, as there is no extra process needed other than individual reasoning.

Rönnegard handily explained the apparent flaw in Lust & Pettit's reasoning by giving an analogy. He compares the premise based voting procedure to a computational algorithm:

> "It is like programing a computer to run an algorithm (with votes as inputs) and then saying that the output is the autonomous choice of the computer. The fact that the procedural choice is discontinuous with member attitudes does not shift any agency ability or moral responsibility onto the procedure" (Rönnegard 2013, 93).

All that is really shown is that members of a group can choose a premise-based procedure because of *their individual desire to be collectively rational* rather than to have a decision procedure that maximizes their individual responsiveness. But this amounts to nothing more than *individuals choosing a procedure.* All the relevant intentional states here are held by individuals who devise a certain procedure, where those individuals accept to abide by the result of it. A choice of procedure made by individuals does not amount to a decision made autonomously by the group consisting out of those individuals, even if the procedure is the same for each of them, and even if some individuals do not personally desire the outcome of the procedure.

In fact, there is a parsimonious alternative to the claim that a group's attitudes (e.g., believing that p) are epistemically autonomous. An alternative argument, which does not invoke epistemically autonomous group-level attitudes, can be given by re-examining the relation between *group- and individual beliefs*.

Margaret Gilbert (see Gilbert 1987, 2002) established a useful way to think of the disparities and differences between individual and group beliefs, or as she puts it, *collective* (or *joint*) *belief*. In her paper *Modeling Collective Belief* (1987), Gilbert draws a useful distinction between the general notion of Individuals *believing that p* and an individual *personally believing that p*. Every incident of an individual *personally* believing that p is an incident of an individual believing that p. But once we see that the *opposite* does not hold, we see how apparently autonomous group attitudes can be analyzed in terms of individual attitudes. On this distinction, individual group members expressing the belief of a group therefore can

> "view themselves as speaking ‚in their capacity as group members,' ‚as a member of this body,' and so on. ‚I'm afraid that you did not meet our needs', says the department chairman to one of the unlucky candidates. *Conscious of his role as a representative of the department,* he speaks as such. He may personally think that this candidate was the best. Or he may have no personal opinion on the matter" (Gilbert 1987, 196) [own emphasis].

The individuals' capacity to adopt a perspective; to assert or accept certain propositions "as a group member" or to settle on a certain matter "as a member of this body" is key. It directly leads us to understand why an autonomous group agent having distinct intentional attitudes is not necessary in order to understand why individuals can have views that are different from that of the group they belong to. According to Gilbert, an individual believing (that p) *as a group member* does not require this individual to *personally* believe (that p) outside of the group context. This *contextual* or *positional* dimension of beliefs allow for a parsimonious alternative to describe the alteration between individuals believing something and the way we ascribe the resulting beliefs to groups. They also allow us to see why, in the case of discursive dilemmas, it is implausible to suggest that a group is *minded* in any meaningful sense. The adoption of certain decision procedures merely results in an agreed-upon *official view*, *or position*. Such an official view may (or may not) be optimal from the point of view of each individual member's *personal* point of view, but it can still be endorsed by such an individual because of the agreed-upon decision structure. We could call this *official view* of a group its "group attitude", but this does not compel us to accept the conclusion that there is anything over and above individuals accepting this "group attitude" as what was agreed upon. If we look at it from this perspective, those cases of the judges reaching their verdict via premise-based aggregations merely describe individual people - *qua occupying their role of judges* - implementing a certain voting mechanism that allows them to reach a compromise on certain positions. Yet, individually accepted and endorsed compromises are no decisive proof for a supra-individual process of reasoning by a group agent.[40]

2.1.4. Interpretivism about Group Agency

Let's consider another, straightforward way in which groups could be said to be agents: by *interpreting* them as being so. The core idea behind the interpretivist approach to institutional agency is the following claim: Because our rich and sophisticated practices of interpreting groups as agents are successful (for examples at predicting certain things to happen in the future) we are justified in claiming that they actually *are* agents. *General* interpretivism amounts to the claim that for *any* entity to be an intentional agent, it's sufficient that we are able to "make sense" of it as such, i.e. to "understand and interpret its behavior by using our folk psychology" (Tollefsen 2015, 97). *Interpretivism about institutional group agents* argues that groups can be successfully interpreted this way, too, granting them the status of agents in their own right. In ascribing groups certain beliefs, desires and intentions, and thereby successfully predicting certain outcomes as the results of those beliefs, desires and intentions, we have everything we need to infer that institutional groups really *are* agents. It's that simple.

Prominently, interpretivism of groups as agents has been put forward by Deborah Tollefsen in her book from 2015 (as well as articles from 2002a and 2002b). The underlying, general theory of interpretivism can

[40] To postulate this alternative view, of course, begs the question of how to understand such claims of an individual's capacity to act, or believe *qua being a group member*, or *qua role-occupancy*. In Chapters 3. and 5., I will explain in detail how we could make of sense of such a notion.

be traced back to Daniel Dennett (1987). I will unpack Dennett's theory below, but for now, let me give several reasons why interpretivism might seem attractive.

A first motivating reason to find interpretivism promising is constituted by the fact that it borrows from functionalist theories in the philosophy of mind the idea that we must not focus on the material make-up of mental states in order to explain them. Rather than speculating about the metaphysical underpinnings and presuppositions for agency (like, e.g., List & Pettit do), interpretivism starts with the practice of *attributing* agency to institutional groups. No nitty-gritty metaphysical theory of actions and agents is needed where this practice has primacy. A second, corresponding reason one might find interpretivism attractive is that our practices of attributing agency to institutional groups pick out the ubiquitous, everyday and ordinary ways we actually talk about groups and their associated actions, e.g., in newspaper reports. Wouldn't it be odd, the interpretivist might ask, to think that we are all fundamentally delusional or misguided when we read sentences like "Ukraine wants to take back Crimea" or utter sentences about our favorite soccer-team "acting far too passively after the change of sides"? After all, we tend to grasp the meaning behind these sentences and they seem to reveal something actually occurring in the world.

But why is interpretivism in a position to infer from this form of talk about group agency that groups really could *be* agents? What are the arguments here? Let us take one step back and look at Dennett's original project of interpretivism. I will now provide a rather detailed explanation of Dennett's theory because my argument against Tollefsen's interpretivism is (partially) motivated by a small, yet crucial mis-analogy Tollefsen uses herself in reference to Dennett's theory.

Dennett's interpretivism

Dennett's interpretivist account of intentionality (see: 1971; 1981; 1987; 2009) argues that in order for us to understand intentionality, we shouldn't conduct a search for it in the brains, neural configurations or bio-chemical processes etc. of individuals, or think about it as a feature of mental states altogether.

Instead, we ought to think of intentionality a concept functioning in *explanatory strategies* that we employ to make sense of the world. Intentionality, according to Dennett, is first and foremost a *practice* of making sense of an entities behavior (e.g., a concrete physical object). Now, when we make sense of an entities behavior, intentionality only comes into play under two assumptions that we have when interpreting it. We take it to be 1) a rational agent with 2) states such as beliefs and desires.[41] These interpreted entities are then called *intentional systems*. Note here, that these entities are not interpretable as intentional systems *because they have intentional states* as, e.g., an inherent mental property. Dennett's proposal works the

[41] For now, our pre-theoretical understanding of these terms must suffice, for on the one hand, Dennett explicitly refers to a "folk psychological" understanding of these terms and, on the other hand, we will turn to the question of what a "true believer" amounts to below anyway. In fact it would be a whole different story to explain what exactly Dennett's theory of rationality and intentional states amounts to (and whether this still really is "folk psychology"). *In short*, rationality, in Dennett's sense, means that one assumes an entity to "believe all the implications of their beliefs and believe no contradictory pairs of beliefs" (Dennett 1987, 21). Beliefs, in turn, are characterized by Dennett to be informational states that are made predictable by our behavior and they can be contrasted with more sophisticated cognitive states, which Dennett calls "opinions". Opinions (or Beliefs-with-a-capital-B) are more complex informational states, being "linguistically infested cognitive states - roughly states of betting on the truth of a particular, formulated sentence" (Dennett 1987, 19).

other way around: By his definition, any entity whose behavior can be successfully interpreted by attributing rationality and intentional states to it is an intentional system.

One way to make sense of this (maybe confusing) practice of *taking the intentional stance* is to contrast it with two other practices of interpretation, i.e. the *physical* and the *design stance (*or *strategy*).[42]

The physical strategy is introduced by Dennett in the following way:

> "if you want to predict the behavior of a system, determine its physical constitution (perhaps all the way down to the microphysical level) and the physical nature of the impingements upon it, and use your knowledge of the laws of physics to predict the outcome for any input" (Dennett 1987, 16).

Contrast this with the strategy of taking the *design stance*, where "one ignores the actual (possibly messy) details of the physical constitution of an object, and, on the assumption that it has a certain design, predicts that it will behave *as it is designed to behave* under various circumstances" (ibid). Taking the design stance to predict the behavior of an entity is *riskier* than predicting it from physical-stance because it relies on extra, and *normative* assumptions: First, that an entity is actually designed as one supposes it to be, and second, that it will operate according to that design, and not malfunction (cf. Dennett 2009, 2). While this stance yields a higher risk, i.e. it's more prone to lead to a false interpretation, the payoff in terms of achieving predictability might also be higher when one takes this stance. Alternatively, the payoff is higher because one achieves the same quality of prediction more easily, i.e. with less effort and time.

Take this example: Understanding why an entity emits certain sounds in certain intervals of time is both achievable on the physical and design strategy. The design strategy might, however, be more efficient: If the interpreter assumes that the entity in question is an *old grandfathers clock* which is designed to ring its bell at the full hour; and if one also assumes that the grandfather clock will operate according to that design, one can predict its sound-emitting behavior. Further, one can - ceteris paribus - predict that it will behave *as it is designed to behave*, even when one never looked inside of the clock and saw all its mechanic parts and their relations to one another (i.e., if one never took the physical stance towards it). Now, the more complex an entity tends to be, the more the design stance might be pragmatically warranted at generating simplified, accurate predictions of the entities behavior.

We now take the *intentional stance* towards an entity, when we treat it *as if* it possesses intentional states in order to make its behavioral patterns intelligible to us. According to Dennett, it works like this:

> "First you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and

[42] All three strategies operate at different levels of abstraction, where the physical stance is the most concrete, i.e. non-abstract way of understanding an entity; the design stance having an intermediate position and the intentional stance being an abstract way of understanding an entities behavior. The overall aim of all these strategies is to understand (i.e. to maximize the predictability of) the observable behavior of an entity.

> finally you predict that this rational agent will act to further its goals in the light of its beliefs" (Dennett 1987, 17).

Initially, this not a controversial claim when operating on the level of *interpretation*. Interpretations may be accurate or inaccurate, simplified or confusing, vulgar or sophisticated, tainted, biased, leaning or balanced, etc.[43] Similarly, interpreting entities to be intentional agents can be a handy way to understand them. Let's take the following example: Instead of studying molecular biology for years to adequately take the physical stance towards it, one can explain the mutation of a virus this way: the virus mutates because *it wants to replicate as much as possible* and it *believes that mutating to a less deadly variant* will allow it to do so. That is of course not what is *really* going on, but to think about the virus in this way is a handy and *pedagogically useful metaphor* for understanding.

When discussing the intentional stance as a handy way of making sense of entities, Dennett similarly admits that "[a]ll that has been claimed is that on occasion a purely physical system can be so complex, and yet so organized, that we find it *convenient*, *explanatory*, *pragmatically necessary* for prediction, to treat it *as if* it had beliefs and desires and was rational" (Dennett 1971, 91f.) [own emphasis].

After laying out this groundwork of the intentional stance, one might naturally think that Dennett's next task would be to distinguish those intentional systems "that *really have* beliefs and desires from those we may find it handy to treat *as if* they had beliefs and desires" (Dennett 1987, 22) [own emphasis].

However, Dennett proposes another, stronger reading of the analysis: Abandon the distinguishing-task altogether and get rid of the metaphor-theory, too! In short, his claim is that by applying the intentional stance and successfully interpreting an entity as an intentional system, we can actually discover truths about the interpreted entities, which lie in so called *objective patterns.* By discovering these objective (or *real*) patterns, we already moved from *as-if-intentionality* to genuine cases of an entity possessing intentional states. The apparent distinction then collapses. This amounts to Dennett's central claim for what it means to be a *true believer*, i.e. to really have beliefs and not just metaphorically being ascribed beliefs:

> "All there is to being a true believer is being a system whose behavior is reliably predicable via the intentional strategy [i.e. by discerning patterns of human behavior indescribable from the physical or design stance], and hence, *all there is* to really and truly believing that *p* (for any propositions *p*) is being an intentional system for which *p* occurs as a belief in the best (most predictive) interpretation" (Dennett 1987, 29).

The way Dennett tries to argue for this claim is to imagine a prediction contest, (i.e. a way to compare the success of the different stances through the measure of predictiveness) between a Martian, who is a *Laplacian super-physicists* (which is his way to say that this interpreter is maximally competent in taking the physical and design stance, so that they never had the need to use the intentional stance) and an interpreter who uses the intentional stance.

---

[43] My interpretation, e.g., of the movie *Shrek* is that it's about the proletarian uprising of the fairytale creatures against the bourgeoise Lord Farquaard, aiming to establish classless society and communist utopia. This - of course - is just one way to watch *Shrek*. For a quite different, and way more nuanced interpretation see, e.g., Caputi (2007).

> "The Earthling and the Martian observe (and observe each other observing) a particular bit of local physical transaction. From the Earthling's point of view, this is what is observed. The telephone rings in Mrs. Gardner's kitchen. She answers, and this is what she says: ‚Oh, hello dear. You're coming home early? Within the hour? And bringing the boss to dinner? Pick up a bottle of wine on the way home, then, and drive carefully.' On the basis of this observation, our Earthling predicts that a large metallic vehicle with rubber tires will come to a stop in the drive within one hour, disgorging two human beings, one of whom will be holding a paper bag containing a bottle containing an alcoholic fluid. The prediction is a bit risky, perhaps, but a good bet on all counts. The Martian makes the same prediction, but has to avail himself of much more information about an extraordinary number of interactions of which, so far as he can tell, the Earthling is entirely ignorant […] The Earthling's performance would look like magic! How did the Earthling know that the human being who got out of the car and got the bottle in the shop would get back in? The coming true of the Earthling's prediction, after all the vagaries, intersections, and branches in the paths charted by the Martian, would seem to anyone bereft of the intentional strategy as marvelous and inexplicable" (Dennett 1987, 26f).

So why is the intentional stance on par (or even better) at understanding the situation than some super-physicist? Dennett concludes:

> "Our imagined Martians might be able to predict the future of the human race by Laplacian methods, but if they did not also see us as intentional systems, they would be missing something perfectly objective: *the patterns in human behavior that are describable from the intentional stance, and only from that stance*, and that support generalizations and predictions" (Dennett 1987, 25) [own emphasis].

So to sum up: the intentional stance provides a vantage point for discerning "real patterns" of behavior which are otherwise missed or indescribable from the physical or design stance. And utilizing this vantage point in understanding and predicting the behavior of intentional systems is what intentional states boil down to, i.e. what intentional states really are. These patterns are epistemically objective in the sense that "they are there to be detected" but they exist not completely ontologically independent from the subjects. They are patterns composed partially of our own "subjective" reactions to what is out there (cf. Dennett 1987, 39). With this rather long exposition in place, we can now turn to the interpretivist approach of group agency.

Tollefsen's interpretivism of group agency

Deborah Tollefsen's theory claims (2002a; 2002b; 2015)[44] that we are justified in interpreting groups as agents because our doing so seems to successfully explain their behavior. She demarcates her position against what she calls "methodological individualism", which is the view that an explanation of group agency must be wholly reducible to statements about the mental states of individual members and their interactions. Against this individualism, she positions her methodological collectivism.

The attributions of agency, viz. intentional states form the basis for explaining and predicting the actions of groups, where this "practice of interpreting the actions of groups is just an extension of our practice of making sense of individuals, and it is governed by the same constitutive rules" (Tollefsen 2015, 104). By this, she means that by treating groups *as if they were* intentional agents, i.e. via taking Dennett's *intentional stance* towards them, we simply assume that they have a unified perspective and that they meet certain norms of rationality. Once we assume that a group possesses rationality, "and a rational point of view is in place, we attribute beliefs, intentions, and desires to groups in the same way we do to individuals" (Tollefsen 2015, 104).

Tollefsen's claim for group interpretivism is guided by two challenges for those who are hesitant to grant groups the status of possessing mental states and the capacity to act: The first challenge Tollefsen's theory poses to reductionists is in explaining not only why particular token-actions of groups took place but also why we - especially in the realm of the social sciences - seem to be able to explain *types* of institutional group actions. The problem with institutional action *types* seems to be that they are *multiple realizable (*see, e.g., Tollefsen 2015, 87f). One and the same type of action, e.g., a nation *declaring war* on another nation, can be multiply (and sadly frequently) realized, although for each case there might be heterogeneous states of the group members leading a nation to do so. Different nations (consisting out of different members within different structures) each have their own way of bringing about this event. If one now was to explain such a social phenomenon by giving a reductive analysis in terms of individual attitudes of the group members and the interaction between those members, one would not be able to find any regularities (or patterns) in the social world. So comparing the different groups to each other would not lead us to give us any particular explanation why these *types* of actions seem to occur. Yet, social scientists, Tollefsen's theory suggests, are interested in answering general and abstract questions like "why do firms or companies, in general, merge under certain economic situations and why do nations adopt certain political policies as a result of economic instability?" (Tollefsen 2015, 88). So the appeal of interpretationist approaches to group agency actually seems to reside in the shortcomings of other, reductive or individualistic accounts of group agency. Tollefsen states that in answering these questions:

> "we cannot appeal to individual intentional states because, in many cases, there may be nothing in coming at the [token] micro-level that would explain why these companies or nations acted in the same way. Given the explanatory aims of social sciences, I think there are

[44] Unless otherwise indicated, I will follow the exposition in Tollefsen 2015.

> strong methodological reasons to accept the adequacy of [type] macro-level explanations that appeal to the mental states and processes of groups" (ibid).

The second challenge Tollefsen's view poses to methodological individualism/reductionism is in parallel to the challenge which Dennett poses to the Laplacian martian. Recall that the Martian could observe everything from the design and physical stance, but that she did not take the intentional stance towards a human beings actions. The key here is that, according to Tollefsen's view, methodological individualists are in exactly the same position as the martian when it comes to the intentional states of social groups. They may observe all the detectable behavior that is given from their vantage point, yet, as they do not take the intentional stance towards groups, *they miss out on something crucial.*

So let us turn to her example. Tollefsen's theory asks us to imagine that two philosophers, a methodological individualist on the one hand and a methodological *collectivist* on the other hand, are engaging in a prediction contest. The individualist tries to explain and predict social phenomena strictly by "appealing to individual intentional states" whereas the collectivist "believes that predicting the behavior of an organization involves viewing the organization as a rational agent" (Tollefsen 2015, 106). The contest is about predicting what the *Ford Motor Company* will do in response to an enormous increase in gas prices. Tollefsen states:

> "In order to predict what Ford will do, the individualist will have to find out who the operative members of the organization are, how each member voted and why they voted that way, and whether they are telling the truth about their intentional states. The collectivist, on the other hand, knowing that individuals are likely to stop buying large vehicles during a time when gas prices are high, and knowing that Ford sells a great deal of these vehicles and wants to continue to maximize profits, will predict that Ford will discount these vehicles in the near future. The prediction is a bit risky, perhaps, but a pretty good one nonetheless. The individualist may make the same prediction, but he will have expended a considerable amount of time and energy. Compared to that of the individualist, the collectivist's performance will look like magic. How did she know Ford would lower its prices on large vehicles without even talking to the president of the company? There are real patterns that emerge from interactions of individuals *that require* for their interpretation the intentional stance" (Tollefsen 2015, 106f.) [own emphasis].

These patterns, discernible only by the interpretative stance towards groups, are "real patterns of social behavior, *patterns that are missed* if one attempt to explain the social world by appealing only to individual mental states" (ibid) [own emphasis]. She concludes, again in analogy to Dennett, that "groups and individuals really have mental states, but their ontological status is more akin to centers of gravity than tables and chairs" (ibid).

Discussion of Tollefsen

I will now assess Tollefsen's argument. To put my cards on the table: I think that Tollefsen's theory ultimately fails to establish a sound proof of group agency. I also think that the methodological individualist is in a superior position to explain group agency. In order to make my case, I will - after turning to some general objections to Tollefsen's account - dismantle the two challenges that Tollefsen poses for the individualist. There, I will first turn to the second challenge. I will discuss two readings of Tollefsen's analysis, a weak and a strong one. According to the weak reading, Tollefsen's view merely claims that taking the intentional stance towards groups is a handy *shortcut* and *pedagogically useful metaphor* for understanding the collective actions of individuals comprising certain social groups. On a stronger reading, Tollefsen's theory claims that taking the intentional stance isn't just useful, but actually *necessary* to discern "real patterns" of group behavior. On this reading, methodological individualists cannot -in principle- explain group phenomena that emerge from interactions of individuals. Methodological individualism would then fail to accurately explain the agency of groups. After discussing these two readings, I will turn to Tollefsen's first challenge for individualistic explanations to account for not only token actions but also for explaining action *types*.

A first and general thing to assess is the limitation of her account. For Tollefsen's theory, interpretivism about group agency is just *one way* to explain the agency of groups and "we don't always take the intentional stance toward groups […] and not all group will be the appropriate target of the intentional stance" (Tollefsen 2015, 104f.). As with Dennett's interpretivism, the ultimate measure of success for taking the intentional stance towards groups is how great the explanatory power of the intentional stance is. Tollefsen accordingly restricts her view to what she calls "organizations", which are "the paradigm case of a group" that the intentional stance can be aptly applied to. According to Tollefsen, organizations have a structure which "allows for complex behavioral patterns to arise and the synthesizing of individual perspectives into a unified perspective" (ibid). As examples she states the U.S. Navy and the Ford Motor Company (ibid). So can assume here that she is talking about *institutional groups* in the sense I defined them above, because this also resonates with her initial taxonomy of groups, where she divided groups into either *aggregate groups* or *corporate groups*, the latter of which are defined through them 1) having a structure, and 2) a decision-making process (cf. Tollefsen 2015, 3). It is left unclear where exactly the line between groups that the intentional stance should be applied to, and those groups where the design stance is more appropriate, should be drawn. Tollefsen states that there are possible cases, where we might not be dealing with a group "whose agency is extended over time in a way that makes their behavior complex *enough* to warrant the intentional stance" (Tollefsen 2015, 105) [own emphasis]. In those cases, "the design stance might be more appropriate" (ibid). She concludes that "the explanatory power of the intentional stance constraints which groups counts as intentional agents" (ibid). But let us simply assume that Tollefsen is correct here: having a structure and a unified group-perspective enables the rise of complex behavioral patterns. And because we are justified in taking the intentional stance towards such groups, we are justified in treating them as intentional agents.

Another critique of her argument (which is pointed out by Katherine Ritchie 2016, 175) rests on the fact that her argument for group agency relies on the truth of interpretivism. Tollefsen herself acknowledges the

fact, that she does not offer any particular arguments for why interpretivism should be the correct theory of mental states. However, this critique might be unwarranted. Tollefsen does not have to argue for the truth of interpretivism in *general*, but only that *interpretivism about group agency* is the correct way to explain the agency of paradigmatic groups, which she calls "organizations". For the intentional states of individual human beings, the story might be different - but this would be a *different* story than Tollefsen's. My critique of Tollefsen will try to match this scope. It should not be understood as an attempt to criticize *interpretivism in general*. Rather, it should be understood as an attempt to show where Tollefsen's argument for *interpretivism about group agency* fails.

A more interesting objection, also put forward by Ritchie, is that Tollefsen's theory does not offer a specific description of the *interpreter*, who, by taking the intentional stance, can discern the real patterns that constitute the social groups' intentional states. Ritchie criticizes that

> "[Tollefsen's] focus on our actual ascription practices suggests that [she] takes interpreters to be ordinary agents like us. We, however, make errors. You and I might interpret the same person or group in different ways. According to the interpretivist *what it is* to, for example, believe that it will rain later today is to be interpreted as having the belief that it will rain later today and for the interpretation to be explanatorily powerful. In asking, ‚Does Bert believe that it will rain later today?' or ‚Does Ford believe it should lower prices on SUVs?' we want a determinate answer that does not vary by interpreter and protects against interpreter error" (Ritchie 2016, 175).

This is a rather minor objection, as it could be easily discharged (Ritchie herself admits this). Tollefsen without making it explicit, seems to appeal to an idealized interpreter, whose features could be further specified. But if we stick with the question of *whose interpretation is to do the trick here*, we might find a better reason to cast doubt on Tollefsen's argument.

Tollefsen's second challenge

A first way to criticize Tollefsen's analysis is that her second challenge to methodological individualism invites for an ambiguous reading. On a weak reading of her example of the prediction contest, she seems to suggest that, thinking about groups *as if they were agents* may merely be a handy *shortcut* and *pedagogically useful metaphor* for understanding the collective actions of individuals comprising certain social groups. Recall that, in the prediction contest, the individualist "may make the same prediction" having "expended a considerable amount of time and energy" (Tollefsen 2015, 106). On this reading, all that has been claimed then - to paraphrase Dennett - is that on occasion, a social group can be so complex and organized, that we find it *convenient*, *explanatory and pragmatically necessary,* or *pedagogically useful* for prediction, to treat it *as if* it was an agent, i.e. *as if* it had beliefs and desires, and *as if* it was rational. Here, the appeal to "spending considerable amounts of time and energy" does not amount to a proof that there is a fundamental *impossibility* of explaining group behavior by exclusively appealing to individual mental states. So it might not be *necessary*, but only *convenient* to take the collective intentional stance to interpret

groups. On this weak reading, taking the collective intentional stance is not necessary, because *nothing is shown to be missed out on* if groups are interpreted in individualistic terms.

On the stronger reading of her analysis (which has also been thoroughly criticized by Backes 2021), her theory seems to argue for more than the pragmatic utility of such *as-if-attributions.* She claims that there indeed are certain group-level phenomena, i.e., "real patterns" which are *in principle* undetectable, indiscernible or indescribable from looking at them from a purely individualistic point of view. What warrants this stronger reading of the analysis? Recall that Tollefsen claimed that "the interpretative stance we take towards groups is able to discern real patterns of social behavior, *patterns that are missed if one attempts to explain the world by appealing only to individual mental states*" (Tollefsen 2015, 106). Her theory further sees, that "[t]here are real patterns that emerge from interactions of individuals that *require for their interpretation the intentional stance*" (Tollefsen 2015, 107). We then really *do need* the collective intentional stance to discern these "real patterns".

If that was to be the case, then interpretivism would allow us to be "mild and intermediate realists about the intentional states of groups" (ibid). Groups, then would "really have mental states" but the ontological status of these mental states would be "more akin to centers of gravity than tables and chairs" (ibid).

Besides this - perhaps confusing - characterization of the ontological status of mental states, Tollefsen's analysis does not offer any significant examples for a form of group behavior that is not reducible to the individual members intentional states and their interactions. As mentioned above, her own paradigmatic case of the prediction contest does not seem to suffice as evidence for her claim that the methodological individualist misses out on such "real patterns". After all, *nothing is shown to be missed out on* by the individualist. Tollefsen's theory fails to show that there are such patterns which *require for their interpretation the intentional stance.* Therefore, the absence of evidence could be taken as evidence of absence.

In a footnote of her article (2002b), Tollefsen does however give a hint what an individualistic interpreter might miss out on. There, also within the example of the prediction contest, she additionally remarks:

> "The individualist will, perhaps, make the same prediction but will have expended a considerable amount of time and energy. I say ‚perhaps' because the individualist may miss the fact that when individuals act in their organizational roles they often act differently than they would outside of the organizational context. Because the individualist is working at the level of individual psychology he may not have room in his theory for notions like organizational context or role. These concepts are concepts of the social scientist, not the individual psychologist" (Tollefsen 2002b, 409).

But this is hardly convincing: While it is certainly true that individuals (can) act differently in their organizational roles than they would outside of the organizational context, this does not mean that a methodological individualist would not try to explain the actions of the group in terms of the actions of the individuals *acting qua role-occupancy*. An individual acting on the tasks and functions of her organizational role is, after all, an instance of an individual acting. It would therefore be the *exact* place where a methodological individualist would start to look for an explanation of group phenomena. According to the

author, methodological individualism is the doctrine that social phenomena can be explained and predicted by appealing to individual intentional states. But by no means does this imply that acting *qua role occupancy* requires the individuals to form something else than individual intentional states.[45] Tollefsen herself seems to suggest that the methodological individualist indeed would proceed by looking at the individual intentional states of individuals *acting qua role-occupancy*, as she herself notes that "the individualist will have to find out who the *operative members* of the organization are, how each member voted and why they voted that way" (Tollefsen 2015, 106). Being an operative member of an organization is, after all, a status role that an individual occupies.[46]

Second, we could turn her claim upside-down and argue, that the methodological *collectivist*, by interpreting the group as a whole to be a rational agent, would "miss the fact that when individuals act in their organizational roles they often act differently than they would outside of the organizational context" (ibid). Why, after all, would the collectivist concern herself with looking at the actions of individual group members? And how would a collectivist even try to explain the difference between individuals acting *privately* and them acting *qua role-occupancy* if she was only focussing on the behavior of a group *as a whole*? If the difference between individuals acting *privately* and them acting *qua role-occupancy* was in any way important to explain the agency of groups, it's hard to see why it would be discernible from the standpoint of the methodological collectivist taking the intentional stance towards *groups as a whole*.

As this is ultimately not convincing, we might look for another way to defend the stronger reading of her analysis. One way this might be done is by looking at her claim that these real patterns "*emerge* from interactions of individuals" (Tollefsen 2015, 107) [own emphasis]. So there might exist group phenomena that are not reducible to the individual members intentional states because they are *emergent in the strong sense.*[47] And although such strongly emergent phenomena seem not to be involved in the example of the prediction contest, they might be involved in other examples of group behavior.

So how could these patterns emerge from the interactions of individuals? According to Tollefsen's view, group intentional states are to be thought of as dispositional states *of the whole system*, i.e. group:

---

[45] In Chapter 3, I will further clarify the notion of *acting qua role-occupancy*. For now, it may suffice to say that an individual acts *as a role-occupant,* or *qua occupying a role* (or *qua role-occupancy*) if she performs the tasks definitive of her role and that institutional roles involve the performance of assigned tasks and functions. This can be fully explained in individualist terms and *acting qua role-occupancy* does not need to make reference to irreducible we-intentions or other forms of collective intentional states.

[46] Also, equating methodological individualism with "individual psychology" void of notions of "organizational context or role" could be regarded as a plain misrepresentation of the approach of methodological individualism, which figures as one of the most dominant approaches of the social sciences. Consider that, if this *was* a good description of methodological individualism, it would be hard to see how it could be an approach in the social sciences at all. To say that "individual psychology" "may not have room […] for notions like organizational context or role" (Tollefsen 2002b, 409) further misrepresents the study of individual psychology, particularly in the field of *social* psychology which tries to explain how individuals are influenced by the behavior and presence of other individuals.

[47] By this, I suggest that Tollefsen understands emergence in a strong sense. I reside with the characterization of *strong emergence* provided by Chalmers (2008): "We can say that a high-level phenomenon is *strongly emergent* with respect to a low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are not *deducible* even in principle from truths in the low-level domain" (Chalmers 2008, 244).

> "According to interpretivism, mental states are not states of the head or brain but states of the *whole agent or system*. They are dispositional states in that they are defined in terms of what an agent will do, say, and think under certain circumstances […] My proposal is that we think of group attitudes as dispositional states of groups. *We are positing not internal states to the group but global and relational states*" (Tollefsen 2015, 103) [own emphasis].

So it might be these global and relational states, emerging from the interactions of individuals, which are overlooked when the individual intentional stance is taken. These global states, per definition, are states attributed to the whole group and not the individual. But also notice that Tollefsen's theory concedes to the possibility that group behavior can in principle be brought about by mechanisms that are completely in line with individualist or reductive accounts:

> "I think we should distinguish between what makes a group an agent and what mechanisms generate the behaviors that justify the attribution of intentionally. Given the various types of groups in the world, their behavior will be generated in various ways. List & Pettit provide one such way - through the aggregation of votes - but we might imagine it being generated through consensus-building as well. It seems plausible that joint commitments or we-intentionality might also be part of the mechanisms that give rise to the complex behavior of groups such as corporations and organizations. *Even the complex and detailed theories of methodological individualists such as Seumas Miller* […] *might be viewed as providing an account of the mechanisms that generate group behavior*" (Tollefsen 2015, 109f.) [own emphasis].

Two things should be noticed here: First, Tollefsen's theory does not speak in favor of any of these accounts explaining how the mechanisms generating group behavior are to be conceptualized. Rather, she wants to suggest that "such theories ought to be empirically informed, as the question of which mechanisms produce the complex behavior that is interpretable from the intentional stance cannot be discerned from the armchair" (ibid). Now to state that this question cannot be discerned "from the armchair" and that the theories of group behavior "ought to be empirically informed" does not amount to proof that there really are such strongly emergent group phenomena in the first place. Ultimately, Tollefsen's theory then holds the burden of proof here, because theories of group behavior that postulate the existence of such global states emerging from the interactions of individuals rely on an assumption, which reductionistic theories of group behavior do not make. *If* a group agent's "overall behavioral profile" is completely explainable in individualistic terms (i.e. if group behavior is to be generated in the way Miller describes it) then there simply might not be such things as emergent states of a group. If the mechanisms that generate group behavior come about in the way in which Miller describes them, the option of global and relational states seems to fly out of the window. As for Miller, social actions are reducible to forms of individual, interpersonal actions (cf. Miller 2001, 6), without the need to invoke any notion of supra-individual entities on the one hand or any irreducible forms of collectivity on the other. This idea also travels to his theory of the actions of institutional groups. Although Miller acknowledges the *usefulness* of speaking about the actions of institutional groups, properly speaking, he states, "there are no such things as corporate actions"

(Miller 2001, 53) but only individual actions which are related in a certain way. And if interactions of individuals is all there is to the actions of groups, i.e., if the agency of groups bottoms out in the interactions of individuals, then again - the individualist intentional stance-taker does not miss out on anything. She is not in a position of ignorance concerning emergent states of a group, because group behavior is generated in a way which can be captured in strictly individualistic terms.

Connected to this, is another pressing problem for Tollefsen's theory. To me, this problem resides in the fact that we do not find any informative (implicit or explicit) criteria of how to distinguish the *individualist interpreter* from the *collectivistic one*. Yet there ought to be a difference that gives the collectivist an advantage over the individualist when they're both in a contest to interpret social phenomena from their specific vantage points. This seems to be crucial for the argument. Tollefsen argues that *taking the interpretive stance towards groups* and that of *appealing only to individual mental states* are mutually exclusive strategies that demarcate the individualistic from the collectivistic interpreter.

Yet, I simply do not see this as a given: For one thing, both do have the intentional vocabulary and idioms that are needed to take the intentional stance. From what Tollefsen states, we cannot assume that the individualistic interpreter renounces the *intentional* stance altogether and looks at the world from a purely design- or physical vantage point (This is *the* key disanalogy to Dennett's Martian here). Rather, as her own example seems to suggest, the *individualistic interpreter does indeed take the intentional stance*. The only difference, then, seems to lie in the fact that the interpreter does so not towards the *group as a whole* but towards the individual members of that group. So let's call the interpretative strategy of the methodological individualist the *individual intentional stance*, and the interpretive strategy of the methodological collectivist the *collective intentional stance*. But where exactly lies the difference between these two views, so that the interpreter employing the *individual* intentional stance is missing out on something crucial?

Tollefsen states that these "real patterns of social behavior […] are missed if one attempts to explain the social world *by appealing only to individual mental states*" (Tollefsen 2015, 106). Yet, these real pattern "emerge from interactions of individuals" (ibid). Why, one might ask, should we assume that the individualist interpreter - in her exclusive appeal to individual mental states - would miss out on something so crucial as the interactions of individuals? In fact, reductive as well as non-reductive theories of collective intentionality see that *this is exactly the place* to look when we want to find out how intentional states can - in any meaningful sense of the term - be collective or shared.

A closer look at her example does not reveal anything substantial to the claim that the individualist stance taker misses out on anything. The individualist, according to Tollefsen's theory "will have to find out who the operative members of the organization are, how each member voted and why they voted that way, and whether they are telling the truth about their intentional states" (Tollefsen 2015, 106f). So far so good. Now, the collectivist, on the other hand "knowing that individuals are likely to stop buying large vehicles during a time when gas prices are high, and knowing that Ford sells a great deal of these vehicles and wants to continue to maximize profits, will predict that Ford will discount these vehicles in the near future" (ibid).

But why is the knowledge of individuals economic behavior (spending less on things when these things get more expensive) inaccessible to the individualist interpreter? Why is the fact that individuals work together to produce and sell cars for profit, trying to achieve this profit under varying circumstances, something that an individualistic interpreter can - in principle! - not understand? Knowing that individuals are likely to

behave in one way when they are incentivized to behave this way from cues in their environment (e.g., buying certain vehicles during a time when gas prices are high) is something that can be very well understood by appeal to the mental states of these individuals. And this is true even if these cues in the environment are, in some way or another, socially constituted. Likewise, knowing that individuals working together in order to achieve a shared goal or outcome will adopt their plans in light of changing circumstances is something that can be very well understood by appeal to the mental states of these individuals, too; even when these changing circumstances are in some way or another socially situated or constituted. Note that regarding the apparent differences in the individualistic and collectivistic capacity to interpret group behavior, the characteristic features of the *collective* intentional stance would have to be spelled out in much more detail too. The collective intentional stance revolves around the assumption that an interpreter predicts the behavior of an organization by viewing the organization *as a whole* as a rational agent (cf. Tollefsen 2015, 106). My conjecture is that the collective intentional stance relies on some concept of collective intentionality different from that of the individualistic intentional interpreter, who interprets collective intentional states only on the basis of mental states of the individual members and their relations.

But underlying this distinction is the assumption that the concept of collective intentional states used by the collectivist stance taker is itself non-reductionistic, i.e., the assumption that collective intentional states are not reducible to individual intentional states plus e.g., mutual knowledge. This, however, might not be the only way to characterize collective intentional states, as there are reductionistic options available. Tollefsen does not make explicit, what exact analysis of collective intentional states the collectivist intentional stance-taker has to base her interpretation on. We must simply assume that she does not think of the collective interpreter as having a reductionistic concept in mind, because explaining collective intentional interpretation in individualistic terms would simply lead the distinction to collapse altogether: both the individual intentional stance taker and the collective intentional stance taker would ultimately interpret groups by appealing only to the intentional states of only individuals and their relation to one another. Nonetheless, the question remains open which exact analysis of collective intentional states Tollefsen has in mind here.

Tollefsen's first challenge

Let’s consider the first challenge of Tollefsen’s account. Recall that Tollefsen here criticizes the reductionists’ apparent inability to explain not only why particular *token*-actions of groups take place, but also why *types* of institutional group actions occur. Tollefsen states that the reductionistic approach fails because action types of groups are *multiple realizable (*see, e.g., Tollefsen 2015, 87f). There might be heterogeneous states of the group’s members for each action-type so that giving a reductive analysis in terms of individual attitudes of the group members (and their interactions) would not be able to find any regularities in, or patterns of the social world. So when taking the individual intentional stance, one might not be able to answer questions like: "Why do firms or companies, *in general*, merge under certain economic situations and why do nations adopt certain political policies as a result of economic instability?" (Tollefsen 2015, 88) [own emphasis].

Discharging this objection can take two argumentative paths: First, Tollefsen does not claim that individualistic explanation of group action cannot *ever* or *in principle* explain action types. Rather, it might be a problem in "*many* [but not all!] cases, there may be nothing in common at the micro-level that would explain why these [groups] acted in the same way" (Tollefsen 2015, 88) [own additions and emphasis]. There are then (just as many? Most?) cases of group action that may in fact do have something in common at the token-level which explains why these groups acted in the same way. In those cases, the individualistic interpreter does not seem to be anywhere near hot waters.

Second, for Tollefsen the actual problem is that there in fact *are* many cases of group action types that cannot be explained in individualistic terms, because there are (many) token-cases of group action that -although explainable in individualistic terms - have nothing in common. But crucially, Tollefsen does not tell for *which of the many cases in the class of action types* this is supposed to be a problem. So she might suggest that there is *some* class of action types that *can* be explained in individualistic terms and there might be *some other* class of action types that *cannot* be explained this way (stating the action-type of declaring war as a possible example). But why is it, that each time the action-type *A* occurs, the token-mechanisms m_A that produce the token-action *a* have nothing in common, while each time *some other* action-type *B* is realized, the token-mechanisms m_B producing the token-action *b* indeed *do* have something in common?

The question, which Tollefsen leaves unanswered, is how, without an *a priori* principle guiding us, we can tell the explainable types of action types *A* apart from the unexplainable types of action types *B other than by looking at the action tokens a and b and their token-mechanisms* m_A *and* m_B *on the micro-level.* Alternatively, Tollefsen might want to suggest that the question, which types of action types *are* and which types of action types are *not* explainable in individualistic terms, is empirical in nature. But if this is an empirical question, then neither the individualist nor Tollefsen's collectivist are in a superior position to answer it from the philosophical armchair.

Ultimately then, this casts doubt on the methodological superiority of c*ollectivistic stance taking*. Also, the *type-token-relation* can be conceptualized in another way, namely *bottom-up* instead of *top-down*. It might be true that one and the same (or at least a somewhat comparable) action can be generated by different mechanisms (e.g., by two different groups employing two different decision-making processes). However, each and every case, a methodological individualist might be very able to discern the particular mechanisms by which a particular token-action is generated from her individualistic vantage point. One possible route for the individualist, then, would be to find *structural commonalities* in each of those different *token*-mechanisms in order to abstract from this an *idealized* mechanism-*type* (or what social scientists call simply an ideal-type). This type, resembling those core features that are shared by the *token*-mechanisms, is forming - so to speak - the token-mechanisms' undercurrent, while allowing for variation of token-actions. And just because the mechanisms of generating token-actions differ from each other, this doesn't mean that the individualist has no means to analyze, catalog or compare these mechanisms. In fact, in comparative social sciences so called *data-driven,* or *grounded theories* tend to operate this way, as they, while admitting that there might be several and severe differences for each case, formulate inductively-

derived (*Weberian) ideal type models* of social explanations, acknowledging differences while abstracting underlying commonalities.[48]

Summing up interpretivism

I started out in this section stating that interpretivism about group agency is an attractive way to think about the phenomenon in questions because we don't need to get in involved in nitty-gritty metaphysics of collective action; and because of the fact that talking about the actions of groups come with such ease to us, that it would be ludicrous to think that we're delusional when doing so.

Interpretivist theories, trying to exploit this, may therefore be an apt way to capture the phenomena. Dennett's initial proposal was to employ the intentional stance towards entities to discern real patterns in behavior, otherwise unnoticed and unnoticeable. For Dennett, these patterns reveal what the talk about intentional states is actually all about. This might - or might not - be the right way to think about intentional states. However, Tollefsen's approach to extend this towards social groups is ultimately not convincing. She fails to show what might be left out of the picture when applying the intentional stance to groups themselves and not only to the individual members of these groups. Ultimately, positing the necessity for this approach should come with the burden of proof that we otherwise could not capture something essential. This proof, in my opinion, isn't in the pudding of Tollefsen's theory.

If the individualistic side of the camp can explain everything that the collectivist sets out to explain, yet does not rely on positing an additional entity (i.e. a genuine group agent), we should reside with methodological parsimony. Claiming the existence of additional entities is a bit like kidnapping your boss in order to get a pay raise: a desperate measure when nothing else seems to help. It should be the last resort and we are far from being so desperate in the search of an explanation for institutional group agency.

We might, however, not abandon this view altogether, as we can still retreat to the weaker, and for me much more plausible claim: That we should, can and will talk elliptically about groups being agents because doing so is an immensely useful, time- and energy-saving shortcut for talking about the actions of individuals.

---

[48] When she explicitly writes about the social sciences (2002a), Tollefsen claims that social sciences suffer from the paradigm of methodological individualism. To this end, she admits to rely on the Durkheimian school of thought to prove that we need to take an collectivistic stance towards groups in order to explain their actions. I hold this to be a lopsided, somewhat unfairly balanced discussion of the methods of social sciences. Durkheim being the *spiritus rector* of methodological holism, can (and is) often juxtaposed to Max Weber's equally influential approach of methodological individualism. Tollefsen does neither claim nor proof that Durkheim's holistic approach is superior to Weber's individualistic one and we therefore can be skeptical to accept her diagnosis of there being a problem for the social sciences in the first place. See Zahle (2016) for the ongoing discussion between these two schools of thought.

2.2. Second Explanatory Path: Theories of Collective Action

The first part of this chapter examined non-reductive, or inflationary theories of group agency that argued for groups themselves to be agents. I tried to show that none of the approaches above seem satisfying because they are vulnerable to reductionistic charges. In this section I want to present reductive accounts of group agency that try to explain the agency of institutional groups on the basis of the capacity for *collective action* of their individual members. I will call these reductive approaches *theories of collective action* in contrast to the above mentioned *theories of group agents*.

As mentioned above, reductive theories of group agency claim that the capacity of groups to perform actions can be understood entirely in terms of individual members and their agency. Reductive views admit that it is literally true that groups perform actions, but "such claims are made true entirely by individual members of the group" (Lackey 2021, 4) acting in certain ways. If we connect this idea to the above established definition of collective action, we arrive at the claim that an *institutional group action* can be understood entirely in terms of individual members and their capacity for *collective action*.

Let me state the results of this chapter upfront: I hold an institutional group's capacity to cause actions is indeed explained best in terms of collective action. An institutional group's capacity to act can be reduced to the individual group member's capacity to bring about this action *together*. So on the one hand, I think it is correct to reduce a group's agentive capacity to the individual group members' capacities to bring about an action together in terms of collective action. On the other hand, I also think that the accounts in the field of collective action, which I will now present, cannot fully explain the agency of explicitly *institutional* groups. One reason for this is that some of these accounts built up their theories from small-scale cases, which figure as their paradigm for collective action in general. But as I will try to show with the so called *Upscaling Problem*, basing one's paradigmatic theory of collective action on small-scale interactions leads to implausible results when applied to large and complexly structured institutional groups. As a result, we have to look for modifications of these accounts of collective action which explicitly set out to explain the agency of large and complexly structured institutional groups.

I will discuss two such influential modifications, put forward by Christopher Kutz and Raimo Tuomela. In my opinion, these accounts fare well regarding their explanatory power of institutional group agency. They, however, face serious problems too.[49] The main reason for why they eventually fail to explain the agency of large and complexly structured institutional groups, is that they cannot plausibly make room for a crucial feature of institutional groups, i.e. the *compartmentalization* of institutional actions, including the resulting *anonymity* (or *impersonality*) *of interaction* between group members. This feature, however, ought to be accounted for when giving an explanation for *institutional* agency in terms of collective action. In the third

[49] In her book, Jennifer Lackey's (2021) provides arguments which are somewhat similar to my own criticism of the presented theories. I here choose not to discuss her theory at large because ultimately, Lackey does have a different project than I have. This is primarily because she focusses her discussion not on the agency of institutional groups but on notions in collective epistemology—such as *group belief*, *justified group belief*, *group knowledge*, *collective reasoning*, *group assertion* or *group lies*. However, I am more than sympathetic to the negative part of Lackey's book and her criticism regarding what she calls *joint acceptance accounts* (see especially: Lackey 2021: Ch. 2). I want to thank Katja Crone for a helpful discussion regarding the relationship between Lackey's theory and my thesis.

chapter, I will show why role-based accounts of institutional action are particularly apt to deal with cases of such compartmentalized cooperation occurring anonymously.

How large exactly does an institutional group have to be, for it to exhibit such forms of anonymous cooperation? Recently, Collins (2023) pointed to the number of 150 group members. Collins argues that 150 is the so called "Dunbar's number" which describes "the maximum number of humans with which any human can have meaningful contact" (Collins 2023, 203f.). Collins concludes that, once a group surpasses this number, "simple human cognitive limitations" will lead role-occupants to

> "interact with one another in ways shaped by each other's role. Many role-occupants in the organization will not appear to one another as unique individuals worthy of particularized engagement, but rather as faceless and nameless occupants of roles or role-types" (ibid).

I do not wish to give a definite answer to the above stated question, nor do I wish to affirm Collin's thesis that our human cognitive limitations are reached at 150 people. It may be higher, but it also may be lower. Ultimately, this is an empirical question that I am not in a position to answer. However, I will simply assume that, within some institutional groups like corporations, government agencies or military units, members regularly have to cooperate with other individuals that they do not know, and of whose existence they might be fully ignorant.

We can speak of anonymity in a *strong* sense in cases, where an individual's exercise of her tasks and functions depends on the exercise of the tasks and functions of some other agents, whose existence the individual is completely unaware of. An example of such strong anonymous cooperation would be two individuals, call them *James B*. and *Austin P.* working as secret agents for a government agency. Let's suppose that James' task is to stalk a potential suspect and to write a report about the suspect's whereabouts, her habits and where to usually find her on a sunny Saturday. Suppose that James does not know what this report will be used for, and who will read it, but writes it nevertheless. Now suppose that Austin's task is to assassinate the suspect. In order to do so, Austin retrieves the report that James wrote without knowing who wrote it. On the basis of the report, Austin stalks the victim and assassinates her. As such, Austin could be said to be completely unaware of the existence of James: He does not know James personally, nor that he works for the agency. He also does not know that it was James who wrote the report. All Austin might infer is that *someone* once wrote this report. Nevertheless, Austins assassinating the victim depends on the exercise of James' tasks and functions, i.e., James writing the report.

In a weaker sense, anonymous cooperation may occur if individuals solely cooperate on the basis of them each fulfilling the tasks and functions of their institutional roles, but where the relation between these individuals does not entail that they each know which *particular* individual is occupying the given role. Suppose, as an example of such weak anonymous cooperation, that you work in a department of a large corporation and that, one day, your computer breaks down. In order to get it started again, you call the *IT Department*. Some individual, whose name you don't know and whom you have never personally met, answers the phone and guides you through the procedure to fix your computer. With the help of "*the IT Guy*", you get rid of the problem. This is a weaker sense of anonymity than in the case above: You at least know that this *unspecified someone* works at the IT Department and that it is her job to help people out

with their computer problems. If we take anonymity be a social relation between an anonymous person and others "where the former is known only through a trait or traits which are not coordinatable with other traits such as to enable identification of the person as a whole" (Wallace 1999, 23), this, I think, is a case of anonymous, or impersonal cooperation occurring on the basis of each of you *acting qua your institutional roles*.

Ultimately, the theories of collective action of Tuomela and Kutz seem to be unable to account for such forms of impersonal or anonymous cooperation. In contrast, I hold *role-based* explanations of institutional agency, to overcome these shortcomings, and therefore side-step the critical problems that Kutz and Tuomela face. I will lay out such role-based accounts in the third chapter. But let's not get ahead of ourselves.

2.2.1. The Upscaling Problem

As mentioned above, most of the prominent accounts of collective action focus on analyzing small and intimate interactions between spatially and temporally connected individuals as their paradigmatic cases for analyzing collective action. Examples of such analyses include *talking a walk together* (Gilbert), two individuals *preparing a Sauce Hollandaise* (Searle), *painting a house together* (Bratman) or *carrying a piano upstairs* (Tuomela).[50]

Now if one, in order to explain collective action of institutional groups, starts with egalitarian, dyadic interaction between a few, and highly committed individuals as the paradigm case, one inevitably faces what is here called the *Upscaling Problem:*

> **UPSCALING PROBLEM:** Accounts of collective action that take minimal forms of collective action to be paradigmatic lead to implausible results when applied (or *scaled up*) to large and complex forms of collective action, i.e. institutional group action.

Now to say that there's the *Upscaling Problem* is not to say that such accounts are false or implausible within their own explanatory realm.[51] Rather, I hold the flip-side of this view to be problematic: To say that cases of large and complex collective action of institutional groups are just fringe, or extravagant cases of collective action, and that the minimal cases are the *paradigm* of collective agency and interaction is misleading too.[52] To this end, Christopher Kutz, arguing for pluralism about collective action, notes that any "attempts to generalize an analysis of collective action from analyses of specific types" (Kutz 2000, 3) will

[50] See, e.g., Poljanšek (2015) for a general criticism of such paradigmatic examples in the field of collective action.

[51] Also, looking at more minimal forms of collective action may seem to be a promising endeavor to achieve clarity on the basic concepts at play in large-scale cases. See: Crone & Gab (Eds.) (forthcoming).

[52] For a similar critique see: Shapiro 2014. Shapiro's account of "massively shared agency" builds upon his critique of Michael Bratman's initial theory of joint action. Bratman's above discussed theory of institutional agency (2022) both diverges from his initial theory of joint action *and* already incorporated the criticism of Shapiro's paper, a fact that should be credited here. I will therefore not lay out Shapiros views in full detail, in the hope to avoid redundancy. However, Shapiro still offers useful insights into the nature of large-scale, hierarchically-organized action of institutional groups, some of which I will discuss in the course of this project.

lead to frustration regarding the plausibility of the outcomes. Note, e.g., that the above mentioned examples suggest a spatiotemporal immediacy of the individuals involved. For the actions of institutional groups, however, this should not be seen as a given. Scaling up small-scale paradigmatic cases to institutional groups, I will argue, leads to frustrating results because one thereby fails to analyze the distinctive features of the targeted phenomenon. It's not comparing apples with bigger apples, but with oranges.

The *Upscaling Problem* might seem as theft over honest toil: Of course, one might object, such theories will fail to explain these phenomena, because these phenomena lie outside of their explanatory realm in the first place! Michael Bratman's initial theory of shared intentional action, for example, *explicitly* restricted the analysis of collective action to small-scale cases of "modest sociality" (Bratman 1993). Yet, as I will try to show with Gilbert, there do exist accounts that take small-scale social interactions as paradigmatic examples of collective action and explicitly indicate that the theoretical building blocks stemming from these paradigms can be applied to large-scale cases.[53] It's these accounts in particular that are subject to the *Upscaling Problem*. So let me make my case by discussing Margaret Gilbert's theory of the *plural subject.*

Gilbert's plural subject theory

Margaret Gilbert's *plural subject theory* starts with analyzing small-scale cases of collective action such as *taking a walk together* (Gilbert 1990), viewing these cases as paradigmatic instances of collective action. Regarding the scope of her analysis, Gilbert claims that it is in investigating exactly those small-scale cases that we find the core mechanisms of group action *in general* (see: Gilbert 1990, 2). Gilbert writes:

> "I shall propose, more precisely, that analysis of our concepts of ‚shared action' discovers a structure that *is constitutive of social groups as such*. To this extent, then, going for a walk together may be considered a paradigm of social phenomena *in general*" (Gilbert 1990, 2) [own emphasis].

[53] John Searle's analysis of collective intentionality can be read in a similar way. His initial analysis of collective action starts with small-scale cases, e.g., of two individuals preparing a sauce hollandaise together. Searle states that the individuals involved would need to have irreducible we-intentions, which are held in a special we-mode (Searle 1990). While Searle does not explicitly discuss *institutional* agency - his project is rather focussed on the modest goal of explaining the "fundamental structure of human civilization" (2010) - there are implications to be drawn from his general theory of institutional reality. In Searle 2010, he states that "in order for cooperation to take place within an institutional structure, there has to be a general collective recognition or acceptance of the institution, and that does not necessarily involve active cooperation" (Searle 2010, 56ff.). While active cooperation, for Searle, requires unanalyzable and irreducible we-mode intentionality, *collective recognition* can be reduced to I-intentionality plus mutual belief (ibid). Now the reason Searle can be said to face the *Upscaling Problem* is that the establishment of institutional group structures, in which individuals could be said to cooperate, would require that the involved members would have to have mutual belief about each others mental states, or at least know whom to include into the irreducible "we-intention" of active cooperation. I discuss this in depth in Ch. 3.1.

Gilbert claims that in order for people to do something together, they have to form what she calls a *plural subject*[54] on the basis of a *joint commitment.* She holds human social groups in general to be plural subjects and states that, "in order to form a social group, i*t is both logically necessary and logically sufficient* that a set of human beings constitute a plural subject" (Gilbert 1990, 9) [own emphasis]. Examples of such a *plural subject* therefore include non-institutional groups like two people walking together, families, or circles of friends but also large-scale institutional groups like nation states and labour-unions (cf. Gilbert 1990, 9). So Gilbert explicitly states that her account not only captures small-scale interactions. I therefore think that it is legitimate to investigate whether her account is subject to the *Upscaling Problem*.

Let's examine her theory first. The concept of *joint commitment* is central to Gilbert's analysis of collective action. In a certain sense, joint commitments are just like individual commitments. They both give the individuals involved sufficient reasons to act in a certain way: If you commit to going to church on Sunday, this gives you a reason to go, although on Sunday you might not have any desire to do so. Now if two people *jointly* commit to do something together, Gilbert's theory sees that they both come to have a claim *on each other* to conform their behavior to the collective activity they committed themselves to, e.g., by having the right to rebuke the other party of the joint commitment when not behaving appropriately. Imagine you and your friend jointly committing to go to church on Sunday. But when Sunday arrives, your blasphemous friend is trying to talk herself out of it by giving flimsy excuses. According to Gilbert's theory of joint commitments, you have, in virtue of your prior joint commitment to go to church together, gained the right to rebuke, or scold her for abandoning the commonly established plan. It's these *normative forces* of joint commitments, which are central to the analysis of Gilbert's account of joint action.

If two or more people form a joint commitment, according to Gilbert, they commit to something *as a body.* To be jointly committed to X *as a body* has two related aspects to it: "First, the parties are jointly committed to bring it about, as far as is possible, that the parties emulate a single body that *X*s. Second, it is understood that their doing so is to be a function of the joint commitment in question" (Gilbert 2014, 175; see also 2006, 136ff.). *Those and only those parties* that are involved in the creation of a joint commitment are bound by it. The parties to a joint commitment, for Gilbert, are those who comprise both its creator and its subject (cf. Gilbert 2006, 135). When an individual, being jointly committed to espouse X *as a body,* is trying to emulate this *single body* that *Xs*, she can be seen as acting, in effect, as the *mouthpiece* of the whole their commitment tells her to emulate, at least as far as possible (cf. Gilbert 2006, 137). Now Gilbert derives from her analyses a general account of collective action. According to Gilbert, two or more people are acting together if and *only if*:

(1) they are jointly committed to espousing as a body the appropriate goal;

---

[54] Gilbert theory of a *plural subject* can easily be interpreted (or misunderstood) to be arguing for such plural subjects to be genuine, non-reductive agents (see, e.g., Ludwig 2007; Schweikard & Schmid 2021). But I do not think that this is what she wants to argue for: First, plural subjects are not to be understood as having a "single centre of consciousness" or a "distinctive form of ‚subjectivity'" (Gilbert 2006, 135). Second, fact that two parties of a joint commitment "try to emulate as far as possible a single body that *Xs"* (Gilbert 2014, 175) does not imply that, as the result of their attempts, the parties to a joint commitment really *do constitute* such a single body. On Gilbert's view, the individual agents try to act *as if they were* a single agent, but they thereby do not come to *really* act as an agent as such (see also: Ludwig 2017a, 261-264).

(2) they are fulfilling the behavioral conditions associated with the achievement of that goal;
(3) their satisfaction of these conditions is motivated in each case by the existence of the joint commitment (Gilbert 2006, 146).

So for a group to act, this would mean that 1) all the members of such a group are jointly committed to espouse as a body the appropriate goal of their group (which means that they form a plural subject, and that the plural subject in question is the group), 2) they each fulfill the behavioral conditions associated with the achievement of that goal, and 3) that the group members' satisfaction of these two conditions is motivated in each case by the existence of the joint commitment.

With this analysis in place, let us look whether the *Upscaling Problem* applies to her account. The first reason to think so, is that a joint commitment only binds those parties that actually brought it into existence (cf. Gilbert 2006, 135). This seems to have implications for conceptualizing actions of institutional groups. As already mentioned above, institutional groups have a structure in virtue of which they can undergo steady changes in their membership. So one way to explain this feature in virtue of her account would see that institutional groups, with every new member joining the group, have to renew the joint commitment that constitutes them. But this seems to be an implausible result, especially concerning the possibility of the temporally extended agency of such institutional groups. An institutional group performing an action would, upon the entrance of a new member, cease to be jointly committed to espouse as a body the envisioned activity, then reform its joint commitment by including the new member, and only then proceed to continue to espouse as a body this activity. Or institutional groups, after losing one of their members, would cease to be Gilbertian plural subjects, i.e. they would *not* be the appropriate target of investigating joint action on the basis of joint commitment anymore. Now this problem may be overcome, e.g., by specifying the conditions of a joint commitment in a way so that new members of an institutional group can *tacitly* commit to the plural subject they become part of. This is probably the way Gilbert imagines her theory to be modified, so that individuals may join pre-existing joint commitments.

A more pressing reason, however, to be skeptical about her analysis is that it seems to be unable to account for characteristic features of institutional groups, such as the differentiation and compartmentalization of tasks based on the group's structure. Let me explain this by discussing the example of FIRE DEPARTMENT.

> **FIRE DEPARTMENT**: Imagine an individual member of a large fire department working *as a truck driver*. In cases of fire, the individual's task is to drive an equipment-truck to a given supply-site, load of the equipment and drive back to the department to retrieve additional gear. Other members of the department work *as on-site firefighters*. In order to do their part in extinguishing fires, the on-site firefighters rely on the driver to bring the equipment. Now imagine that a forest fire breaks out. Responding to the fire, the driver brings the equipment-truck to a nearby supply site, loads off the equipment and drives back to the department. Throughout the day, the individual drives to the site several times, each time loading off the equipment and thereby supplying the on-site firefighters with the material necessary to extinguish the fire. While driving the gear to the supply site, the driver actually has no idea

which of all her colleagues are, and which are not on active duty at the moment. At the end of the day, the fire is extinguished and everybody goes home.

Does the extinguishing of the fire constitute a collective action performed by the fire department, i.e., is it a collective action of an institutional group? For Gilbert, extinguishing the fire is a collective action if and only if it is brought about in virtue of the *joint commitments* of the individual members of the group. And for Gilbert, for individuals to be jointly committed to X (in this case: to put out the fire) as a body, is to be jointly committed to bring it about as far as possible to emulate a *single X-er*. So, according to Gilbert, we must assume that, as the driver encounters maybe some but maybe not all of the active on-site firefighters (maybe they are all scattered around in the nearby bushes), she is bound by a joint commitment to put out the fire. We must also assume that she is emulating as far as possible a single X-er doing so. But is this really a plausible description of what is going on?

If the above mentioned case really does constitute a collective action of the fire department, this seems to be at odds with Gilbert's condition of everybody being jointly committed to each other. Recall that the truck driver might actually have no idea which of all her colleagues are, and which are not on active duty at the moment. But how are individuals supposed to be jointly committed to the other group members if they do not even know about the other individual's involvement in the group action? Additionally, how should we understand the normative claims that parties of the joint commitment have on each other, if the individuals involved in the group action cannot actually monitor each other's behavior? How should individual members implement their rights to rebuke one another, if they do not know about each other's actions in the first place? If the truck driver has no idea which of all her colleagues are, and which are not on active duty at the moment, then whom does she include in her *joint commitment to espouse as a body the appropriate goal* of extinguishing the fire? And also: Whom, by jointly committing to putting out the fire, is she gaining a normative standing over, e.g., a right to rebuke?

It seems as if the coordinated efforts of the fire department's members to put out the fire may occur without the individuals involved *knowing about each other's involvement*. For institutional groups to perform an action, this often means that the action in question is further divided into sub-actions which smaller, specialized segments of the institutional group are supposed to perform in order to bring about the overarching group action. It is characteristic of institutional groups, like e.g. corporations or military units, to functionally divide their tasks, and create departments, or institutional sub-units which are designated to perform certain, specialized sub-actions. Think of the division of a military corps into brigades, the division of brigades into battalions, the division of battalions into companies; into platoons; into squads, etc. A military company. e.g., might consist out of several platoons, one of which might be a *medical* platoon; an *airborne infantry* platoon or a infantry platoon which includes a weapons squad; which is further divided into two machine gun teams and two missile teams, etc. And within each of these sub-units, action may be further specified or compartmentalized. Or think of the way a corporation might have a sales-, an accounting-, a HR-, an advertisement-, and a legal department where each department works on certain specialized tasks.

If individuals in institutional group contexts are acting together if *and only if* they are jointly committed to espousing *as a body* the appropriate goal, then this puts a heavy toll on every single member of such

institutional groups. Imagine, for example what this would mean for a single clerk in one of the several departments of a multi-national corporation. Each and every time she goes on about her work, she would have to try to *emulate as far as possible a single group agent* (or plural subject) that she is a part of, acting in effect, as the mouthpiece of the whole institution that her commitment tells her to emulate. If Gilbert's analysis is correct in that two or more people are acting together iff they are jointly committed to espousing as a body the appropriate goal, then this also implies that the clerk actually has to have an idea about which goal the corporation is pursuing. It further implies that she understands how her action contributes to the group's overall goal.

I hold this characterization of individual action within institutional contexts to be misguided. While individual action according to a task-specific differentiation might make reference to such a joint commitment of the whole group, it certainly does not need to. Rather, individuals might simply be *individually* committed to fulfilling *their* assigned tasks, where this commitment must neither be derived, nor tacitly be formed by a joint commitment towards espousing as a body the group's goal. It may, e.g., stems out of pure personal or strategic goals of the individual, for example being *individually* committed to fulfill one's assigned tasks in order to get paid money to do so.

Gilbert's analysis of collective action might very well be plausible for minimal forms of collective action, such as taking a walk together.[55] But as I tried to show, her analysis does not seem to handle well more complex cases of collective action, such as FIRE DEPARTMENT. Her account of collective action then leads to implausible results when being applied (or *scaled up*) to larger and more complex forms of collective action, i.e. institutional group action. It is therefore subject to the *Upscaling Problem*.

Choosing small-scale egalitarian groups with high interdependency among the members as paradigmatic examples of collective action comes with postulating conditions for collective action that presuppose some form of symmetry in the psychology of the individuals involved, e.g., in the form of common knowledge, mutual belief cascades, or reciprocal expectations of the participants (see: Fiebich 2019). This symmetry, however, is implausible to assume in accounting for the actions of large and complex structured institutional groups. The more fundamental problem is to think of institutional groups (simply) as an extension of small-scale egalitarian groups with high interdependency among the members, i.e. to think of them as *large-scale* egalitarian groups involving such high interdependency and psychological symmetries. Institutional groups, however, differ in critical aspects from such exemplificatory cases of small-scale cooperation. They are groups that have a *structure*, and on this basis survive the change of membership. They can establish hierarchies allowing for task-differentiation, i.e. the division of (cognitive and/or manual) labour and structural sub-division of groups into specialized units, all of which facilitates anonymous and impersonal ways of cooperation between group members (more on this in Ch. 3).

[55] That is not to say that Gilbert's account hasn't been criticized even within her own paradigmatic realm: See, e.g., for charges of circularity: Tollefsen 2002a; Crone 2021.

Let us now look at accounts that explicitly analyze such large-scale, non-egalitarian forms of collective action: Christopher Kutz's minimalist account of collective action and Raimo Tuomela's *positional* analysis of we-mode groups. Here is the upshot of the following sections:

Christopher Kutz's theory of participatory intentions marks an important step towards a theory of collective action that considers the fact that the human capacity to cooperate does not exhaust itself in taking walks in the park. Instead, he sets out to explain large-scale, "pedestrian but nonetheless genuine forms of collective action" (Kutz 2000, 2) by drawing the conceptual distinction between individuals having *group intentions* via occupying an *executive perspective* on the one hand and individuals having *participatory intention*s via occupying a *subsidiary perspective* on the other hand. This is a first and fundamental step towards a realistic explanation of institutional group agency, as his theory can account for the particular features of institutional groups, such as hierarchical structures or distributed power-relations among the members. Ultimately though, Kutz seems *presuppose* the existence of hierarchy, in order to explain how acting on behalf of a group can take on different forms. I therefore argue to go beyond Kutz's approach and look at Raimo Tuomela's *positional* theory, which does make such hierarchies intelligible in the first place.

Tuomela develops his we-mode account in order to contrast cases where individuals act *as group members* (i.e. acting in the we-mode) to cases where individuals *act as private persons* (i.e. acting in the I-mode). This is an important distinction. But it also gives rise to a fundamental problem in Tuomela's we-mode approach to group agency. When the only alternative to acting as a private person is to act in the we-mode, one eventually comes to have a lopsided understanding of acting as a group member to always involve some sense in which a "we" of a group, i.e. the representation of a group in its totality, seems to be a prerequisite of collective action. If we want to understand and eventually explain the agency of especially large and complexly structured institutional groups, this becomes a problematic assumption. Assuming that "we-thoughts", "we-intentions", forms of "we-reasoning" etc. are necessary elements in the analysis of institutional group agency leads to implausible consequences, or so I will argue.

2.2.2. Christopher Kutz's Participatory Intentions

Christopher Kutz, too, developed his *minimalist account* of collective action in discussion of the *Upscaling Problem*. In his paper "Acting Together" (2000), he starts out with the observation that paradigmatic approaches to collective action focus on "explaining the special case of intimate, tightly reciprocal cooperative activity, such as conversing, walking together, or singing a duet" and that they do so with "great sophistication" (Kutz 2000, 2). Nonetheless, Kutz wants us to acknowledge that the "pedestrian but nonetheless genuine forms of collective action" happening in "broader or more attenuated social contexts, such as voting, working in large organizations, supplying capital for risky ventures-collective" (ibid) play an important part in human sociality. The problem, he diagnoses, is that the predominant ways of analyzing these broader forms of collective action "make implausible attribution of the high degrees of interdependence and mutual consciousness that are at the heart of extant analyses of collective action" (Kutz 2000, 2). In order to fill in this blindspot of collective action, Kutz sets out to develop a *minimalist*

account of collective action that he claims to be "parsimonious in its metaphysics and philosophical psychology, sufficiently undemanding to account for the cooperation of loosely-linked agents, and anti-egalitarian enough to reconcile collective action with hierarchy" (ibid).

His minimalist account of collective action relies on the key concept of *overlapping participatory intentions*.[56] According to Kutz, participatory intentions are *the* common element of collective action, a form of intention that individuals form in light of a collective enterprise. All forms of collective action, according to Kutz, share a "common element in the form of overlapping, individual participatory intentions" (Kutz 2000, 4). For an individual to form such a participatory intention, this is a rather simple process: If an individual agent understands her individual action as *contributing to a collective end* and thinks of her actions as *doing her part* in order to contribute to that end, she has formed a participatory intention. Participatory intentions, then, have two conditions of satisfaction: 1) an individual part (or function) and 2) a collective end (or goal). Based on his theory of participatory intentions, Kutz develops a minimalist analysis of collective action. According to Kutz, "all collective action, hierarchical and non-hierarchical, pre-programmed and dynamic, planned and spontaneous" (Kutz 2000, 27) admits of the following analysis:

> "[A] set of individuals *jointly G* when the members of that set intentionally contribute to G's occurrence by doing their particular parts, and their conceptions of G sufficiently and actually overlap" (Kutz 2000, 27).

I will unpack the key features of this account below. For now, notice that his analysis is reductive at its core. It does not claim for there to be non-reductive group agents, and his analysis bottoms out at the claim that a group acts because its individual members act together. When we try to attribute collective actions to groups, Kutz derives a "general principle" for this to be the case: A group performs G intentionally when *its members do their parts* of intentionally promoting G and overlap in their conceptions of G (cf. Kutz 2000, 28).

So let us take a closer look at the relations of group actions, individual actions and participatory intentions. I will start with contrasting Kutz's account with other theories of collective action in order to draw out its specific features. The merits of Kutz's analysis in explaining large-scale cases of cooperation (and thereby also his immunity against problems of upscaling) can first be highlighted by contrasting two features of participatory intentions with features of *collective intentions* portrayed by above mentioned accounts: 1) Kutz weakens the condition of common knowledge to what he calls *mutual openness* and 2) he replaces the condition of interdependence with the concept of *strategic responsiveness*. Mutual openness of intentions

[56] Kutz characterizes "extensional overlap" to be "a pragmatic concept and always a matter of degree, given inevitable differences in each agent's expectations and conceptions of the group act" (Kutz 2000, 21). Summing up, it simply means that the individuals involved in a collective action must think of their actions as contributions to the same token collective action. This allows for a certain leeway, so to speak, regarding an individual's interpretation of the character of this action: "Agents will have more or less determinate conceptions of the group act, they may be more or less willing to compromise after bargaining on the character of that act, and they may have very different ideas about the scope of the group act, its duration and membership" (Kutz 2000, 21). So individuals involved in collective action don't have to be perfectly informed about what they are doing together, but they have to get it somewhat right in order for their participatory intentions to overlap.

consists in individuals - instead of having common knowledge about each others intentions - being disposed to rely on a *shared cognitive background* which in turn needs not to involve "the object of explicit beliefs in order to serve its role in assigning determinate content to each other's potentially ambiguous utterances" (Kutz 2000, 6).[57] In short, this means that for individuals' intentions to be mutually open, they do not have to explicitly form beliefs about what other individuals believe in order to get a collective action off the ground. They only need to be disposed to believe that it is at least *possible* that "the other knows of or will try to predict our choice, and be favorably disposed to the other's knowledge or anticipation of that choice, at least in the sense that no one would modify his or her plans in virtue of disclosure" (ibid). Second, an agent's intentions are *strategically responsive* if they are minimally sensitive to the beliefs or predictions about what others intend to do. A simple case is two people coordinating their walking down a hallway: If I believe or predict that you intend to go right, I go left. My intention to go left is strategically responsive to what you do and *vice versa*, but it does not imply that our intentions are interdependent: I do not have to believe, that you believe that I go left on the condition obtaining that you believe that I believe that you go right etc. I simply chose left and hope you to be responsive to my choice.

The third and core characteristic feature of participatory intentions can be drawn out by contrasting them with above mentioned analyses of collective action too. Standard cases of collective action seem to describe individuals intending that their group performs an action *in its entirety*. Kutz calls the kind of intention, that has as its scope the entire or total collective end, *group-intention* (to *J*; to *X* as a body; to *X* by individual means of Y). Now *participatory intentions* contrast with *group-intentions* insofar as individuals can have participatory intentions without the intention to realize the *entire* collective end (or goal). Whereas the scope of *group-intentions* seems to require that each participant aims at everyone else's achievement of the collective end *in its entirety*, an individual has a participatory intention if she intends to do *her part* (to fulfill her function, or to play her role) in a collective enterprise all while being mutually open and strategically responsive to the contribution of the other members. While it might be plausible that individuals who group-intend *to X* also intend *to do their parts* of the X-ing (e.g., Searle's (1990) "I <we-intend> to X by individual means of Y), Kutz claims that the converse does not hold. One can have the intention to contribute to (or to participate in) a collective goal without intending the goal to be brought about in its entirety. Kutz pumps our intuition this way:

> "It would ring false to attribute to an individual cellist in an orchestra the intention that ‚we play the Eroica,' or to a single running back the intention that ‚we win the football game.' (A cellist or running back who said this might be thought to take too grandiose a view of his or her role.) Rather, it is far more natural to attribute to the cellist an intention to perform his or her part in the symphony, and likewise to the running back. In contrast, we might say of a conductor, orchestra manager, or coach, that each intends that his or her group perform or win, given the ability of each to influence these total outcomes" (Kutz 2000, 23).

[57] Kutz relies here on Sperber & Wilson's (1986, 38ff) concept of *mutual manifestness.*

Call the perspective, an individual has if she has a group-intention (i.e., if she contemplates her actions from a perspective that suggests that she can settle a group's action in its entirety) an *executive perspective* on group action (cf. Kutz 2000, 21). Call the perspective an individual has if she has a participatory intention, i.e., if she contemplates her actions from a perspective that does not suggests that she can settle a group's action (but nevertheless intends to do her part in it) a *compliance* (or *subsidiary*) perspective on group action (ibid).

Why is this - rather subtle - difference between group intentions and participatory intentions viz. their corresponding perspectives so important? Because, following Kutz, we are now in a position to capture cases where the relation of participating individuals to the overall group action seems weak, i.e. where individual members can only make small or even marginal contributions to bring about a collective end. This includes cases where "cognitively vague, alienated, or dyspeptic agents" (Kutz 2000, 26) are participating in a collective action that is subject to "circumstances of routinized cooperation, hierarchical authority, and compartmentalized information" (ibid). The distinction between group- and participatory intention then allows to capture something central regarding institutional group action: It does away with the picture that individuals participating in collective action always understand their own actions in a way suggesting that *what they do is entirely up to them*, i.e. that it's necessary for each individual in a collective action to occupy an *executive perspective* on the action in question. This, in turn, shifts the focus away from the paradigmatic case of egalitarian, i.e. non-hierarchical groups, where actions are conceptualized as equally distributed among the members. Instead, we can move towards an account of institutional group action that highlights the unequal *distributions of power* regarding the way in which members of a group relate to a group's action (what I call the member-to-action dimension of power) while at the same time recognizing this dimension to play out in the psychology of the individuals involved. Once this divide between executive and participatory perspective is in place, we can account for the contributions of participants who might have no particular views concerning what the group as a whole should do, but who still contribute to the collective action by being individually committed to do their part.

This, I think, paints a more realistic picture of the relation of individual contribution and institutional group action. Above, I mentioned the example of a clerk going on about her day to day work in a sub-department of a large corporation. Let us further suppose that the group in question is a car manufacturer. Recall then my critique of a Gilbert's plural subject theory: Gilbert's theory has it that for the clerk to go on about her day to day job, she would have to try to emulate as far as possible a single group agent (or plural subject) that she is a part of, acting in effect, as the mouthpiece of the whole institution. Now contrast this with an explanation of the clerk's behavior in terms of occupying a participatory perspective. By, e.g., maintaining the infrastructure of the offices, she contributes (vitally) to the actions of the group she's a member of (for the maintenance of offices is necessary for the *other* members for *their* contribution to a collective action etc.). But in order to maintain these offices, she does not need to intend to *manufacture cars and sell them for a profit*, although that is the overarching goal of the institutional group she is a part of and contributing her actions to. Regarding the case of large groups, Kutz writes:

> "individuals whose contributions are marginal will typically not have an executive intention with respect to producing the total outcome or activity. Instead they will have a subsidiary,

> participatory intention, an intention to do their part of achieving the executively-determined goal. *They may have an intention regarding the whole, but they don't need such an intention to identify with and act for the sake of the main goal.* Their individual participatory intentions will in turn serve as executive with respect to further intentions and actions. The cellist, for example, has a subsidiary intention to perform the cello part of the Eroica, which generates further intentions to play in tune and tempo, and to show up for rehearsal on time. The cellist's participatory intention may be subsidiary not only to the music director's intention that the orchestra perform the symphony, but to the cellist's own self-regarding intention to make a career out of music, to play as much Beethoven as possible, and so on" (Kutz 2000, 23) [own emphasis].

Kutz, in this paragraph, rightly points out that while it might be implausible to assume that every member of a large institutional group always occupies an executive perspective regarding the groups action, it does not amount to this being impossible. But his arguments certainly speak against the necessity of all members having "group-intentions“ with the aim at bringing about the collective goal (to J; to X as a body; to <we-intend> to X) in its entirety each time an institutional group acts. So with the conceptual distinction between individuals having *group intentions* via occupying an *executive perspective* on the one hand and individuals having *participatory intention*s via occupying a *subsidiary perspective* on the other hand, we can reconcile institutional group agency with the feature of distributed power-relations among the members.

By power-relations, I do in the first place mean that in a collective action performed by a group, some will have more influence in bringing about this action than others. You can call this a *member-to-action* (MtA) relation of power. Imagine A and B working on a building site for a construction company. Their task is to dig a hole in order to build the basement of an apartment complex. Here, one dimension of these power-relations concerns A's and B's capacity to dig such a hole. If A has greater capacities for digging holes than B, e.g., by being physically stronger, or more experienced in such tasks, or if she has certain licenses for operating machines that help her dig holes, then A has more influence in bringing about the action than B. A's *member-to-action relation of power* then is bigger than B's concerning their collective action. But we can draw from Kutz's division a further way power-relations can be hierarchically distributed: The divide between participatory and executive intentions can also account for the unequal distribution of power in regard of *settling* a group's action, or goal. Some can have more influence on collective action through their having group-intentions (and consequently occupying executive perspectives) and thereby determining a group's goals, while other members' contributions to a collective goal can consist in them *doing their part* of achieving the executively-determined goal. Not everybody has to be on equal footing regarding a group's action. Some give orders and some comply with orders. Those who executively determine the goals of a group typically can - in a way to be further specified below - be thought to be the members with greater member-to-action relations of power. They are usually "higher up" in the echelons, while those members who simply do their part in achieving this goal, the "rank-and-file" members, can be seen as having lesser member-to-action relations of power and are, so to speak, "lower down" the hierarchy. The problem, as I will explain, is that Kutz does not convincingly explain how such hierarchies can come about in the first place. But let us look further clarify some aspects of the minimalist account.

First, notice that simply *having* a participatory intention to do one's part does not seem sufficient for there to be a group action. This is because we need to exclude individuals who have this sort of participatory intention but are not actually part of the group *they are having these intentions for*. If an individual intends to do her part in order for VfL Bochum *to win the Champions League*, but isn't actually part of VfL Bochum, it would seem odd to say that she brought about the event of Bochum winning the Champions League (cf. Kutz 2000, 28ff). Kutz's explanation of this is that groups must establish certain membership-criteria, so that "only the actions of bona fide members of the group can be attributed to the group" and that the group's actions "do not include those of posers, even posers who act for the sake of the group" (Kutz 2000, 30). How exactly should we understand these conditions of membership? Kutz here claims that for *some* institutional groups, internal recognition of membership by other members may be sufficient (cf. Kutz 2000, 28f.), yet some institutional groups might have additional necessary conditions for membership. The case of alienated group members further seems to suggest that Kutz is pointing towards external, rather than internal conditions of membership because one can be a member of a group while neither identifying with the group, nor its goals.

Second, Kutz notes that a group can act in virtue of *only one* member acting: "When IBM's executive negotiates a sale of its microcomputer division, IBM sells its division" (Kutz 2000, 30). Kutz explains this by stating that the action of bona fide members must also be consistent with "the particular powers and limitations on the member's role" (ibid).

But here it is crucial to note that Kutz neither explains what the "particular powers and limitations" of a member's role consist of, nor how these powers might come about. Rather, he seems to implicitly presuppose ways in which member-to-action relations of power seem to be distributed according to a member's particular position within an institutional group: certain group members are able to do certain things, while others are not.

Notice, however, that up to this point, Kutz merely offered an explanation of how the member-to-action relations of power can be distributed among members according to whether they either occupy an executive or a subsidiary perspective. Yet, occupying an executive perspective with corresponding group intentions and *actually being in a position*, or *actually having the power in virtue of one's role* to settle a group's intention are two different things. Crucially, they do not necessarily have to go along with each other. Think of the above mentioned clerk again, who, one day, decides to exclusively act on *group intentions* via an *executive perspective* and sees every contribution of her to bring about the group-level goal of *manufacturing cars and selling them for a profit.* Even if this was a realistic depiction of her day-to-day psychology, it would strike us odd that the clerk thereby gains some sort of power over *actually* determining a group's goals (e.g., the power to sell the electric car division of the company) *just because she took up such a perspective*. Clearly, there are members "higher up the echelon" who get to decide what the groups goals are, i.e. certain group members which are actually able to do these things while others, including the clerk, are not.

But then again, we now might ask why *these* members are in a position that allows them to do so. It remains uninformative to say that they are "higher up the echelon" without specifying what exactly being "higher up the echelon" of a hierarchy actually amounts to, or consists of. What seems to be missing is how individuals are able to determine *whether to occupy a compliance* or *executive* perspective, so that they do

not simply chose to do so on a random basis. And we also miss an explanation for how certain conditions of membership authorize some to perform certain actions, while these same conditions may also restrict (or limit) their power to perform other actions.

Kutz seems to rely on some sort of presupposed hierarchy of members concerning their authority to act on behalf of a group. A hidden decision-structure, so to speak. But simply presupposing the existence of such a hierarchy and equating it with the divide between executive and subsidiary perspectives won't do the trick, for we have to have some criteria which allows us to say whether a group member is actually justified in occupying either of these perspectives. Otherwise, everybody could in principle always be justified in choosing either of the encompassing perspectives on a group's actions. I tried to show that this seems to lead to implausible consequences.

Now in order to fully account for hierarchy within groups, we also need to capture the mechanisms by which *some group members can influence bringing about the action of other group members*. Call this the *member-to-member* (MtM) relation of power: If *A* can order (command, instruct etc.) *B* to Φ, e.g., dig a hole in a certain way, or to operate a machine in order to do so, and B's Φ-ing is a result of her complying (obeying, following) with *A*'s order (all while the opposite direction of command (*B* ordering *A* to *Φ*) does *not* hold) then *A* has a bigger member-to-member relation of power than *B*. Exploring these ways individuals can both assert power in collective action and over the actions of other individuals in such collective actions will be a vital part of shifting away from the paradigmatic case of egalitarian, i.e. non-hierarchical groups, where actions are conceptualized as equally distributed member-to-action relations. It, however, will also require us to go beyond Kutz's minimalistic account, for he remains silent about both the way in which members *qua having certain powers* can act for the group in ways other members cannot, and how such powers play out on relations between members themselves.

The next section thereby focusses on Raimo Tuomela's *positional* theory of group agency, where the distinction between operative and non-operative members can account for this. I will here try to show that Tuomela's *operative members* correspond to the individuals in Kutz's account occupying an executive perspective but that *operative members* come to occupy such a perspective in a certain way. This way - by being collectively authorized - gives us criteria for determining whether they rightfully do so or not. With the concept of collective authorization, we can also discover how power relations between group members (MtM) make it possible for one *operative member* to perform a whole action *of the group* while acting *for the group*, how the unequal distribution of member-to-action relations of power is linked to distribution of the member-to-member relations of power.

2.2.3. Raimo Tuomela's Positional Theory of Group Agency

According to Raimo Tuomela (2013) the agency of groups can be explained by giving an analysis of the so called "we-mode".[58] When an individual acts (reasons, deliberates etc.) in the we-mode, she is acting

[58] I follow his exposition of group agency in 2013. For further analyses of the we-mode see: 1992; 2002; 2003; 2007: 2011 and Tuomela & Miller 1988.

(reasoning, deliberating etc.) *as a group member* and *for the group*. Under certain conditions obtaining, individuals on this basis then can constitute *we-mode group agents*.

So here's a short sketch of Tuomela's theory of group agency: Tuomela's overarching claim is that individual members of a group *acting in the we-mode* constitute a *functional* group agent, i.e. a *we-mode group agent*. When acting in the we-mode, individuals gain a qualitative self-understanding of acting in group contexts that is "irreducible relative to the individualistic, I-mode properties of our common-sense framework of agency and persons" (Tuomela 2013, 91ff.). This *psychological* irreducibility is based on the irreducibility of so called "we-mode reasons" (more below). Those we-mode reasons, Tuomela claims, *cannot* be reduced to individual "I-mode reasons" for they are needed as an explanation of how we-mode groups can achieve (otherwise unexplainable) results in game-theoretical action equilibria, such as the Hi-Lo game (see Tuomela 2013, 11ff.). While Tuomela holds the we-mode to be a *psychologically irreducible* state, the *agency* of group agents always "ontologically bottom[s] out in the behavior of its members" (Tuomela 2013, 21; see also 91ff.). Tuomela's theory of group agency, then, is fundamentally a theory of *collective action* occurring on the basis of the "we-mode". The nature of such group agency can - at least to some extent - be fleshed out by comparing it to individual agency:

> "Analogously to intentional action (or at least a central kind of singular intentional action) by an individual agent, intentional action by a group agent (and its parts, the members) is normally based on reasons for action. Analogously to an individual having to coordinate the movements of her body parts when performing singular action (e.g., a bodily one), the members of a (we-mode) group coordinate their action (indeed all activities including mental ones) both synchronically and diachronically in order to achieve group goals. Analogously to an individual agent who is committed to her intended actions, the group members are committed as a group, that is, collectively committed, to the group's actions. Let us also assume along with common sense that at least some groups, viz., group agents such as we-mode groups and corporate agents more generally, indeed can intentionally perform actions (e.g., a business company buys another one). This intuitive analogy together with the common-sense premise tells us that if a we-mode group acts as a group, its members in general (perhaps not all of them) must act in the we-mode, for a group only acts through its members; and if the group acts in the we-mode, this means that a substantial amount of we-mode acting by the members occurs" (Tuomela 2013, 34).

However close a group agent might come to resemble an individual intentional agent, the Tuomelian we-mode group agent is not an intrinsically intentional agent like, e.g. French or List & Pettit argue for. Rather, Tuomela says that under certain conditions, we can *extrinsically attribute* such groups *as-if mental states* like wants, intentions, and beliefs and, in turn, on this basis attribute them agency (cf. Tuomela 2013, 50ff.). Group agents, then, are *fictitious entities* (cf. Tuomela 2013, 48ff.). A Tuomelian group agent fundamentally is "a collectively constructed functional social action system" without original, or underived intentional states, a phenomenology, or consciousness (ibid). So we-mode group agents are only agents insofar as they are attributed this agency from an external point of view. They do not realize non-reductive, intrinsic

intentional states. In contrast, Tuomela assumes human individuals are *intrinsically* intentional, to have raw feelings and qualia (ibid).

According to Tuomela, such social action systems can exhibit a strong form of unity, as they create interdependencies and normatively binding commitments to act for the individual members. Tuomela claims that collectively committed group members come to identify with the group "in the sense of adopting its goals, views, and norms as their own and, so to speak, giving up to the group, when functioning as group members, part of their natural de facto authority to act" (Tuomela 2013, 21).

The we-mode is central to Tuomela's analysis. How can we further analyze it? Fundamentally, the we-mode describes a psychological *mode*, or *attitude*; a way in which an individual comes to understand herself. The we-mode can be first approximated by contrasting it to the *I-mode* of engaging in group action. Tuomela states that the I-mode and we-mode are different, mutually exclusive "psychologies" (Tuomela 2013, 21). To be in the I-mode is to think of oneself in individual, private terms, whereas the we-mode is describes a way to think of oneself *as a group member*. And as mentioned, the we-mode, according to Tuomela, is irreducible to the I-mode. When individuals identify with a group and on this basis engage in "we-thinking", they are each realizing a collective intentional state, which requires that they

> "intend to act together *as a group* and thus, according to [Tuomela's] approach, for the same *authoritative group reason*, and also satisfy the criteria or markers of *collective commitment* and the *collectivity condition*" (Tuomela 2013, 6; also: 23-24; 55).

Those three concepts, i.e., authoritative group reasons, the collectivity condition, and collective commitment are key for understanding the we-mode.

What is a group reason? In short, it is a reason individuals have for promoting the group's interests and goals (cf. Tuomela 2013, 38f.), where this is not a result of aggregating the individual member's interests and goals, but by means of collective acceptance or recognition. A group reason is a reason that no individual group member *as a private person* has (or needs to have), but a reason that is *collectively accepted* by every individual *as a group member*. Group reasons are normatively binding for the members of a group and they either can be collectively established by the group members themselves (e.g., through a deliberative process) or be imposed by an external authority.

Second, the *collectivity condition* concerns the condition of satisfaction regarding the individual members contributions to a collective action. This condition sees that individuals engaged in collective action are necessarily "sitting in the same boat" regarding the actions of their group (cf. Tuomela 2013, 24; 81). On this condition, a group's goal (e.g., to run up a hill) is only satisfied simultaneously and interdependently for all of the members of the group (if all the members reach the top of the hill).

Third, the collectivity condition ensures that "the group acts tightly as a unit, and so countervailing private reasons are completely set aside and there accordingly is no incentive to free-ride, and each of the members acts in a solidary way towards the others" (Tuomela 2013, 8). The idea of collective commitment entails that members of a group have obligations toward one another to perform actions for the group. Collective commitment is the "central ingredient that accounts for the stability and robustness of group life" (Tuomela 2013, 44) and "the glue that binds the members of a we-mode group together" (Tuomela 2013,

83).[59] By this, it is directly linked to the *group reason*: the group members *qua we-mode* are not primarily individually (or I-mode) committed to promote the group reason, but such an individual commitment can be derived from being collectively committed to the groups reason.

So to sum up, for individuals to be engaged in the we-mode, and consequently constitute a we-mode group agent, three conditions need to be fulfilled: there needs to be a *group reason for action*, the *collectivity condition* needs to be fulfilled and there needs to be *collective commitment* among the group members. If these three conditions are satisfied, then the group members can come to constitute a we-mode group agent in the sense above.

If individuals engage in the we-mode, they act *as group members* and *for the group.* But what does this amount to? First, it is important to emphasize that this aspect of acting *as a group member* and *for the group* in the we-mode travels all the way down to the individuals' capacities for action. Also, it involves a special *mode of reasoning*, which Tuomela calls "*we-reasoning from the group's point of view*" (Tuomela 2013, 15; original emphasis). The idea that individuals who engage in "we-reasoning from the group's point of view" is summarized the following way:

> "we-mode thinking, reasoning, and acting are concerned with thinking and acting as a group member. In the we-mode case, an agent is supposed to identify with the group (or rather with being a group member) and to act as a full-fledged and well-informed member of the group, guided by its goals and norms. To think (believe, want, intend, or feel) and act in the we-mode is to see one's activities essentially as part of what the group is doing" (Tuomela 2013, 153).

In the fundamental case of believing something in the we-mode and *as a group member,* Tuomela argues that the resulting beliefs that an individual comes to hold are not beliefs *proper* but basically "acceptance-beliefs" about a proposition:

> "As to what we-mode acceptance belief amounts to, it […] centrally involves the idea of functioning as a group member. Thus, when g [a we-mode group] believes that p, the members of g, collectively considered, will be assumed to believe (accept) that p when functioning as group members and thus be collectively committed to p. Their private beliefs related to P (here covering p and –p) can be different from those they adopt as members of g" (Tuomela 2011, 86).

[59] Tuomela highlights three characteristic features of *we-mode collective commitment*:

1. We-mode collective commitment is based on a group agent's commitment
2. Qua being collectively committed, the members of a we-mode group are group-normatively committed to one another to perform their parts of the required collective activity, since the group agent's intention cannot typically be satisfied by a single member.
3. Qua being based on the group's commitment, a member is accordingly committed to the group and to its members to further the group's interests (cf. Tuomela 2013, 44).

If one, engaging in *we-reasoning from the group's point of view,* e.g., accepts such an "acceptance belief", e.g., that p, and also accepts that p→q, then an individual *qua group member* can infer (and consequently act on) q without having to *personally* take the truth of neither p nor q at face value.[60] Tuomela formally expresses this idea in his *Collective Acceptance Thesis for Group Sociality* (CAT):

> **(CAT):** A proposition, s, is *group-social* and correctly expresses a *group-social fact* in a primary sense in a group g if and only if (a) the members of g collectively accept s as true or correctly assertable for g, and (b) necessarily, they collectively accept s as true or correctly assertable for g if and only if s is true or correctly assertable for the members of g functioning as group members (2013, 220).

Tuomela coined the term *groupjective* truths (ibid) for the collective acceptance of such group-social propositions, i.e. propositions taken to be true *for the group* and hence for individuals being collectively committed to hold these "beliefs" *as a group member.* On the basis of accepting certain propositions as expressing *group-social facts*, we-mode groups can establish what Tuomela calls an e*thos,* i.e., the groups "constitutive goals, values, and purposes to which group life is dedicated" (Tuomela 2013, 15; see also 26-27; 30-33). By this, Tuomela means the collective acceptance of constitutive and central principles, including the basic goals, beliefs, norms, standards, practices, values, and customs etc. of the group. To act *as a group member* of an institutional group necessarily includes an individual's understanding of what the group's ethos is about and the ways in which to individually further it. He summarizes this the following way:

> "Collective acceptance of the ethos as true or right for the members makes for the correctness of statements of the following kind: ‚We intend to achieve X', ‚We believe X', and so forth, which apply to the group members as a unit, viz., as the group that they conceptually and functionally reify and entify by the collective stand they take toward it" (Tuomela 2013, 153).

So within Tuomela's conceptual apparatus, we can usefully differentiate between individuals acting for reasons they may (or may not) have as private persons, and individuals who *as group members* intend, believe or act *for group reasons;* who come to "*we-reason from the groups point of view*". Conversely, saying that a we-mode group agent believes something, is another way of saying that individuals believe something *as group members and for the group.* Saying that a we-mode group agent intends something is another way of saying that individuals intend something *as group members* and *for the group*. Saying that a group agent did something, i.e. that it performed a certain action, is another way of saying that individuals did something, i.e. performed a certain action *as group members* and *for the group*.

[60] Tuomela's favorite example for this is that at one point in the history of Finland, the Finns used to treat squirrel-pelt as money.

Tuomela's positional account of group structures

Having laid out the groundwork of his theory, we now might investigate what makes the Tuomelian account of we-mode group agents especially apt to explain *institutional* group agency.[61] Because capturing structures is a primary hurdle when attempting to describe the agency of institutional groups, we should have a look at how his theory tends to do so. To this end, Tuomela highlights that his theory of we-mode group agency is a *positional theory* of group agency (cf. Tuomela 2013, 157-158; 163-169). So what makes his theory of we-mode group agency "positional"? And why is such a *positional* theory especially apt to explain institutional group agency?

Recall that according to Tuomela, when an individual acts in the we-mode, she is acting *as a member* or acting *qua being a member*. Such a reflexive capacity for perspective taking of an individual is the fundamental or "rock bottom" sense in which individuals act in institutional contexts, and we can "equivalently speak of institutional acting in a group as acting in the we-mode rather than in the I-mode" (Tuomela 2013, 137). Further recall that if an individual is acting in the we-mode *as a group member,* she may not only accept certain propositions to express group-subjective facts, but she may also commit herself to promote the group's *ethos*, i.e., group-subjective facts about the fundamental goals, purposes, beliefs, and norms etc. of the group. Derivatively, a group member can infer tasks, purposes, and functions required in promoting the group's goals, *viz.* ethos that she and all the other members collectively accepted and are collectively committed to.

Now *structure* enters this picture when acting as a group member is *contextually specifiable*, i.e., when individuals can infer what tasks, purposes and functions they as *specific* group members (and not as group members *generally*) ought to do in a given situation. These individual tasks, purposes, or functions of the group, in turn, are constitutive of a group's structure. They define the *positions* individuals occupy by identifying with the we-mode group, or reasoning *as a group member*.

Such positionally organized groups, on the basis of establishing a structure, exhibit certain key features: First, positions are neutral regarding the particular individual occupying it. Positions are specified according to the group-related tasks, functions or purposes and not in light of any particular individual. In turn (leaving some minutiae aside), any individual occupying a certain position is able to fulfill these tasks, functions or purposes. This is important, as it allows for group structures to remain intact, or to *persist* although the individuals occupying the positions may change:

> "When one position-holder leaves the group the position is filled and group life continues much as before: Positional structure guarantees smooth maintenance of group functions. This

[61] Tuomela's analysis of the we-mode group describes a *mode of cooperation,* whose nature and type is both the same for small, "unorganized, egalitarian collectives built around more or less spontaneous or task-relative cooperation" *and* groups with "high degree of organization and hierarchical structures" (Tuomela 2013, 157). So if, as Tuomela argues, we-mode group agents do not come in any particular size or form, his we-mode account of group agency should - at least *prima facie* - not lead to the problem of upscaling when applied to institutional groups. To the opposite, explaining the agency of potentially large, organized and hierarchically structured groups, according to Tuomela, is an "indispensable part of social theorizing" (Tuomela 2013, 9) that needs to be accounted for.

gives permanence to the group but also flexibility, as the particular members can be changed" (Tuomela 2013, 32).

Next, Tuomela's positional theory includes a differentiation of the member-to-action (MtA) relations of power: Position P_1 might require the individual group member M_1 to x_1 while position P_2 might require the individual group member M_2 to x_2 etc. Taken together, these sub-actions then give rise to the group action, or as Tuomela would put it: the we-mode group agent performing action x. So the first step in explaining institutional group agency, i.e. an explanation of the persistence of institutional groups through the change of membership as well as the division of labour and compartmentalization of tasks seem to be satisfied by Tuomela's positional view.

Up to this point, one might object, this all resembles the above mentioned theory of Kutz. But as I argued above, Kutz ultimately fails to explain how groups can establish hierarchies. So why does Tuomela fare better in this regard? Here, a key feature of Tuomela's positional account must be considered: Tuomelian positions are *interrelated*. Not only do they specify relations of individual group members concerning actions (M_1 to x_1; M_2 to x_2 etc.) but they *also specify the relations of members to one another*, i.e. *member-to-member* (MtM) *relations of power*. To see how a distribution of member-to-member relations of power gives way to establishing hierarchies (and thereby overcome the problems that Kutz's account faces), we may now look at Tuomela's divide between *operative* and *non-operative* members.

At the first and fundamental level, this divide concerns the relation of group members to the collective acceptance of propositions to count as group-social facts in the way described above. Remember that a we-mode group establishes an *ethos*, i.e. "constitutive goals, values, and purposes to which group life is dedicated" (Tuomela 2013, 15) as well as related group reasons guiding the members' actions. On this basis, we-mode group members can assert certain propositions regarding their group's beliefs, goals or intentions. Those assertions may take the form of "We intend to achieve X", "We believe X" etc., applying to all the group members *as a unit* qua their collective commitment.

Tuomela claims that in some situations, the establishment of certain propositions as group-social facts may require the collective acceptance from *some,* but *not all* the members of a we-mode group. Tuomela here refers to cases where the established decision procedures within a group are insufficient to solve a given problem, where group-action may be "too complex without hierarchical organization" (Tuomela 2013, 161), or cases where "the members may simply appoint somebody to be in charge in virtue of her superior capacities" (ibid). In such cases, the distinction is made between *operative members*, whose collective acceptance is necessary for the establishment of group-social facts, and *non-operative* members, whose acceptance of a specific proposition is not necessary. However, the non-operative group members have to - so to speak - "meta-accept" the acceptance of certain propositions by the operative members.

This is the first dimension of the division of member-to-member relations of power: Non-operative members accept the operative members to be *authorized* to accept certain propositions to hold *for the group*. So for a two-level case of (i) non-operative authorization of (ii) operative acceptance of (iii) a proposition to count as a group-social fact, we might schematize this the following way:

i) $M_{non\text{-}operative}$ accepts that
ii) $M_{operative}$ accepts that *p*, where
iii) *p* describes a group-social fact (we believe X; we intend to X, etc.).

The member-to-member relation of authorization in this case is bottom-up: Several "rank-and-file", or non-operative members M_1-M_n authorize $M_{operative}$ to accept certain propositions to be counted as true for the group.[62] Here, the acceptance of a proposition by the operative member is sufficient in order to explain collective, or group-social beliefs. All that is needed by the other, non-operative members is that, in some weak sense, they must tacitly accept or go along with the operative members' collective acceptance of the proposition (cf. Tuomela 2013, 162-164; for a critique of "tacit acceptance" see: Störzinger 2022, Ch. 6.2.; Ch. 8). In this regard, Tuomela's operative/non-operative divide is similar to Kutz's distinction between participatory and group intention: operative members correspond to group members having Kutzian *group intentions*; non-operatives correspond with those who have *participatory intentions*.
Now on a second level, positional group members' collective acceptance of propositions can *turn on the group members' positions themselves*. This inverts the member-to-member relation of power from bottom-up *authorization* to top-down *commanding*. This is the second aspect of how the operative/non-operative divide establishes hierarchies.
Here, the member-to-member relation of power can generate "second-order" member-to-*action* relations of power, as in the case of $M_{operative}$ accepting certain propositions regarding the tasks, functions or purposes of the rank-and-file, non-operative members M_1-M_n. So for a two-level case of (i) non-operative authorization of (ii) operative acceptance of (iii) a proposition to count as a group-social fact about (iv) the tasks, functions and purposes of a non-operative group member, we might formalize this as:

i) $M_{non\text{-}operative}$ accepts that
ii) $M_{operative}$ accepts that *p*, where
iii) *p* describes a group-social fact (we believe that X; we intend to X
etc.)
iv) about $M_{non\text{-}operative}$ tasks, functions or purposes where
v) this includes $M_{non\text{-}operative}$ accepting (i-iv).

So an uneven member-to-action relation can be brought about by mechanisms of acceptance by which *some* (operative) *group members can influence bringing about the action of other* (non-operative) *group members*, which is a member-to-member (MtM) relation of power. This might seem somewhat abstract. But consider the scenario of your boss telling you to be in charge of the Christmas party-committee and you complying with her order. We could schematize this in the following way: i) you accept that ii) your boss accepts that *p*, where iii) *p* describes a group-social fact iv) about your tasks, functions or purposes, where

[62] Again, think of Tuomela's example that the Finns used to treat squirrel-pelt as money. Another version of the story has it that not all the Finns had to come together and collectively accept this group-social fact (that squirrel-pelt is money) for it to actually be the case. Rather, some, namely the operative members of the Finns (maybe their Queen and her court) might have sufficed to do so, while the other peasant Finns just "went along" with these decisions.

v) this includes that i) you accept your boss can (partly) establish group-social facts ii) + iii) about iv) your tasks, functions or purposes qua group member.
Strictly speaking, such a relation constitutes a *member-to-member-to-action* relation: One member (e.g. Anna), occupying a certain position, accepts that another member (e.g. Betty) - by her acceptance as an operative member - establishes group-social facts about the tasks, functions or purposes, *viz.* actions of the position of Anna. Tuomela captures this interrelation between *top-down commanding* and *bottom-up authorization* in terms of a "flow" of authority relations:

> "Authority may flow from the rank-and-file members to the leaders and down again, as when a […] head of department appoints some subset of the staff to prepare the filling in of a vacancy in the department. In each case, we are dealing with a complex task-right system with authority relations that may be omnifarious" (Tuomela 2013, 162).

It might seem mysterious, or circular that a member has to accept that she has to accept certain propositions including her acceptance of this itself. But keeping in mind that we-mode group members act *as group members* holding certain positions can solve this problem. An individual might accept that *as a group members* her position demands her to accept certain propositions *qua her position in the group*. To reiterate this thought rather bluntly: *holding certain positions comes with accepting certain propositions*. And this can include accepting that other members of the group are able to shape, influence and alter what propositions need to be accepted when holding one's own position.
In its simplest form, the establishment of hierarchies along the operative / non-operative divide describes a two-level relation (cf. Tuomela 2013, 161). Here, in any given institutional group, there are two sub-groups of individuals comprising a we-mode group: the operative members and the non-operatives. Crucially, these relations of power can be *iterated* and the authority-system based on the operative/non-operative distinction can therefore be *multi-layered*. Think of the above mentioned division of a military corps into brigades, the division of brigades into battalions; of battalions into companies; of companies into platoons; of platoons into squads etc. as an especially good example for such a multi-layered hierarchy. Now this iteration of power-relations may come in two dimensions: horizontal iteration and vertical iteration.
The vertical iteration of power-relations is especially important for the establishment of a *chain of command*. Let us start discussing vertical iteration by adding just one layer of hierarchy to an institutional group and see what this amounts to. Suppose a group with a 1) top-, 2) middle-, and 3) bottom-layer. At the top-level, the operative members accept certain propositions to be counted as true for the group, where these propositions include propositions about group members in the middle-layer. Here, operative members of the top-layer are authorized to shape, influence and alter what propositions need to be accepted when holding one's position in the middle-layer. Now, and crucially, this can include the acceptance of certain propositions that *provide the members in the middle-layer with an operative status regarding the members of the bottom-layer* (but not regarding the members at the top-layer). On being granted this operative status, members of the middle-layer can in turn accept certain propositions to be counted as true for the group, where these propositions include propositions about group members of the bottom-layer. Here too, the operative members are authorized to shape, influence and alter what

propositions need to be accepted when holding one's position in the bottom-layer. Institutional group members (of some middle-layer) therefore can occupy *both an operative and a non-operative status relative to the layers below, or above them*. This status of authority is level-specific and hierarchically *opaque*, i.e. it travels down- but not upwards. This has certain implications for our understanding of the we-mode, which I will draw out below.

The *horizontal* iteration of hierarchy should be understood as a *task*-, or *domain*-relative differentiation of power. Suppose institutional group *G's* goal *g* sees that a certain action *A* has to be realized, and that the action in question must be divided into smaller sub-actions, e.g., sub-actions a_1-a_3. Realizing these sub-actions is both necessary and sufficient for realizing A (more on this below in 3.1.-3.2.). The horizontal iteration of hierarchy, e.g., in *brigades de cuisine*, among the operative/nonoperative divide now can be seen as *task*-, or *domain*-relative insofar as certain members gain the operative status regarding the performance of *certain* sub-actions, e.g., sub-action a_1 (*Hors d'œuvrier*) but not regarding the performance of certain *other* sub-actions, e.g., sub-action a_2. (*Légumier*), or sub-action a_3 (*Confiseur*). According to Tuomela, this will "result in several *kinds* of authorized operative members. For instance, a member (or type of member, a position) may be an operative one for action (e.g., building a bridge) while also being subject to orders from another operative" (Tuomela 2013, 161) [own emphasis].

Taken together, the horizontal and vertical dimensions of hierarchy can explain how institutional groups can be complexly structured and consequently account for the division of (cognitive and/or manual) tasks. Ultimately, a group's capacity for task-division rests on the multi-dimensional, i.e., horizontal and vertical distribution of member-to-action (MtA) relations of power along the group members' positions or functions.

It's worth noticing the resulting explanatory advantage of Tuomela's positional theory over Kutz's minimalist theory of collective action. While Kutz seems to merely presuppose that institutional groups exhibit hierarchical distributions of member-to-action relations of power along group members, Tuomela's theory can actually account for such uneven distributions of member-to-action relations of power. Why? Because he bases them on the *uneven distribution of member-to-member relations of power*. Hierarchies in the member-to-member relations of power put group members "in their place". Some members can - in virtue of holding operative positions - command other, non-operative members make do certain things, to comply to orders and to accept certain propositions about the tasks, functions and purposes of their positions. Tuomela's *operative members* may be said to correspond to those occupying an executive perspective in Kutz's account. Both settle what the group's goals are. But these *operative members* come to occupy such a perspective on group action not on a random basis, but by *being collectively authorized by the non-operative members*. Collective authorization, then, is the missing criterium for determining whether someone is justifiably acting on such an executive perspective. The authority of operative positions therefore does not stem out of the members randomly choosing to hold an executive perspective.

Nonetheless, I think that *because* Tuomela's positional theory can explain the organizational complexity of institutional groups, his emphasis of the *we-mode* in explaining institutional agency must be rejected as too restrictive. Acting in the "we-mode", I will argue, cannot paint a realistic picture of what it means to *act as a group member* as it would require for individuals to *we-reason* from the *groups point of view*. But an

institutional group's structure might just be too complex, compartmentalized and fragmented to provide such a unified perspective for the individual group members. Let me further explain this criticism.

Critique of Tuomela

Because Tuomela's theory of institutional agency received substantial criticism of its central aspects elsewhere (see, e.g., for an overview Hindriks 2015; Priest 2014; Townsend 2015; see for detailed discussion Preyer & Peter 2017), I will focus my critique on questions regarding the theory's ability to adequately explain the (domain-specific) *compartmentalization* of tasks, characteristic of institutional group agency. For brevities sake, compartmentalization can be understood both as an institutional group's capacity for vertical and horizontal iteration of power-relations, including an institutional group's capacity to break up actions into smaller sub-actions and distribute these sub-actions among the members of a group.[63]

The fact that groups are able to compartmentalize actions entails that it is not necessary for all of the group members to be (at least directly) involved in realizing an institutional action. As argued above, the horizontal, i.e., task- or domain-relative division of hierarchy along the operative/nonoperative divide allows that certain members gain an operative status regarding the performance of *certain* sub-actions (cooking the vegetables) but not regarding the performance of certain *other* sub-actions (creating the deserts). This in turn implies the possibility, especially for large and complexly structured institutional groups, that not every group member has to be involved in realizing a group action, or even has to have knowledge on how the division of tasks is actually designed. This form of compartmentalization motivates my criticism of *the we-mode being too demanding* as a basis for explaining institutional agency. So what are reasons to doubt that Tuomela's positional theory can account for these features of institutional agency?

Recall that for individuals to be engaged in the we-mode, and consequently constitute a we-mode group agent, three conditions need to be fulfilled: there needs to be a *group reason for action*, the *collectivity condition* needs to be fulfilled, and there needs to be *collective commitment* among the group members. Recall that when individuals act in the we-mode, they consequently

> "identify with the group (or rather with being a group member) and […] act as a full-fledged and well-informed member of the group, guided by its goals and norms. To think (believe, want, intend, or feel) and act in the we-mode is to see one's activities essentially as part of what the group is doing. Accordingly, an agent who fully functions in the we-mode is on this ground disposed to cooperate with the others—obeying the group-normative requirement concerning cooperativeness" (Tuomela 2013, 153).

I will now argue that the compartmentalization of group action stands in conflict with these conditions in several, yet critical aspects. Often times, we simply cannot assume that one knows what one's group is actually doing, yet one can contribute to its action. Also, identifying with the group that one is supposed to act for turns out to be to strong of a condition for cases of institutional agency.

[63] For a more detailed explanation, see Ch. 3.

First, consider the condition of group reason: For a we-mode group agent to act intentionally is to act for a reason. Because Tuomela wants his theory of group agency to bottom out in the agency of individuals, he distinguishes between two levels of group-reasons: reasons for action on the level of the group (the so called *group agent's reasons)* and reasons for actions on the level of the members (group reasons simplicitier) (cf. Tuomela 2013, 39). According to Tuomela, the relation between these two levels is established by individuals engaging in *we-mode we-reasoning*, which occurs "when they act as group members […], thereby *reasoning from the group's point of view* and taking the group's directives as authoritative by deriving member-level group reasons from them" (ibid) [own emphasis]. Such we-mode we-reasoning further requires the members to "think and reason in terms of a thick, 'togetherness' notion of 'we' with respect to attitudes, actions, and emotions attributable to the group and its members" (Tuomela 2013, 39).

Now contrast this with cases where an institutional action is split up and compartmentalized along the vertical and horizontal distribution of hierarchy. Here, reasoning from the "group's point of view" and "taking the group's directives as authoritative by deriving member-level group reasons from them" (Tuomela 2013, 39) might not be an available option for individuals.

To make my case, let us consider the *Manhattan Project.* The project was realized primarily (though not exclusively) by one of the world's biggest public engineering agencies, the United States Army Corps of Engineers (USACE), a sub-division of the United States Army. Its goal was to build the first atomic bomb in order to defeat Nazi-Germany and its allies. Building the first atomic bomb was a large-scale, yet highly secretive and heavily compartmentalized collective action with more than 150.000 people involved extending from 1942 to 1945. Of those tens of thousands of individuals, only a handful of people knew what the whole project actually was about (i.e., what the exact goal was) and only *one* person, General Leslie Groves, was said to be knowledgable about the entire project, its sub-actions *and* how they were compartmentalized (i.e., only he knew both what the goal was *and how exactly it was to be realized*) (see for details: Rhodes 1987). In order to operate so-called "Calutrons", which were machines separating isotopes in order to enrich uranium, the Manhattan Project employed a workforce of nearly 10.000 mostly young women, the so called "Calutron Girls" (see: Kieran 2013). The Calutron Girls worked at the *Y-12 Plant* in Oak Ridge, Tennessee. During their involvement in the project, the workers weren't told *anything* about their work other than it would consist out of operating machines: "During processing and training, individuals, no matter the rung they occupied on the information ladder, were given just enough detail to do their job well, and not an infinitesimal scrap more" (Kiernan 2013, 127). Due to the omnifarious secrecy and seclusion, even the machines' monitoring gauges, knobs, and displays were encoded with cryptic letters in order to prevent the operating workers from understanding what they really represented. Yet, the Calutron Girls outperformed STEM-professionals at Berkeley, who, holding PhDs and working under Nobel-laureate Ernest Lawrence, operated the same machines knowing what they were actually for. Apparently, it sufficed that they were doing their part, maybe (but not necessarily) believing that other members were doing their part as well. Yet, the Calutron Girls did not know - and were not even in a position to know - what the parts of the other members were, or what their own sub-actions contributed to on a group level. According to Kiernan (cf. 2013, Ch. 6), one of the interviewed workers, at the time, thought she was producing celluloid-films. Others thought they were mixing paint.

So the Manhattan Project was case of an institutional group action, where there was a strong hierarchical distribution of power. It also figures as a prime example of working under ignorance. Only a few operative members (maybe just one), residing at the top-layer of the hierarchy, planned and consequently divided the group's action. Based on the vertical distribution of member-to-member relations of power, the operative members decided which group members had the relevant, domain-relative member-to-action relations of power regarding the compartmentalized group action. Further, this division of tasks was designed in a way that restricted the lower-level group members in their ability to grasp the extent and scope of the overarching group-action. When such ignorant members asked what goal they were contributing to, they were probably told by their superiors (the operative members residing over their position): "Don't worry, you don't need to know that. Anyway, get back to work and stick to your knitting!" It is also plausible to assume that the Calutron Girls had only vague or false beliefs about the extent or scope of the group action they were contributing to. Their superiors might have intentionally misrepresent the group's goal, leading individuals to act under false pretenses.

Thus, it seems implausible to say that the Calutron-Girls were necessarily *acting on a group reason*, i.e. it is implausible to assume that they were thinking and reasoning in terms of a togetherness notion of "we" with respect to attitudes, actions, and emotions attributable *to the group as a whole.* For the Calutron Girls, there might not have been access to such a unified "group's point of view" to begin with.

Second, and related to the potential unavailability of a group reason, the collectivity condition (CC) can also be regarded as too restrictive. Tuomela sees the collectivity condition as necessary for a we-mode group agent's ability to act "tightly as a unit, and so countervailing private reasons are completely set aside and there accordingly is no incentive to free-ride, and each of the members acts in a solidary way towards the others" (Tuomela 2013, 8). He posits:

> **(CC_i)**: It is necessarily true (based on the group's acceptance of P as group g's goal) that P is satisfied for a member A of g (qua member of g) if and only if it is satisfied for every (other) member of g (qua member of g) (Tuomela 2013, 41f).

Tuomela goes on to explain that this collectivity condition has two aspects to it, an objective and a (inter-)subjective dimension. On the one hand, for a group goal to be satisfied, this requires the objective occurrence of said goal due to the group members' actions. This should not strike one as controversial. The inter-subjective dimension of the collectivity condition, on the other hand, is more problematic. For a goal to be satisfied for the group members in a full *subjective* sense,

> "we must require more than the objective occurrence of the goal-state due to the participants' action. Analogously with the case of intentional action, we must require *that the participants believe that the goal has been satisfied and*, given (CC_i), also *believe that it has similarly been satisfied for the others*" (Tuomela 2013, 41) [own emphasis].

In this subjective sense, the condition seems to imply that individuals understand their individual actions as contributing to a single, monolithic goal, which is further known to all of the participants. If we assume that,

through the process of compartmentalization, an institutional group's overarching goal is broken up into smaller, fragmented sub-goals, some of which - by design or out of sheer complexity - may not be known by all group members, the collectivity condition seems to lose its plausibility.

Again, consider the Calutron Girls. On what basis can such a group member believe that a certain sub-goal (e.g., enriching enough uranium) has been satisfied and - given (CC_i) - also believe that it has similarly been satisfied for all the other group members, given that she does not even know which sub-goals there are and which group members are - qua their position - contributing to any given sub-goals?[64] The subjective dimension does not seem compatible with compartmentalized group actions, as it would require that individuals mutually believe that the goal is satisfied. However, they may not know about the sub-goals of other members and still act *as group members*.

Finally, the condition of collective commitment also tends to collide with the assumption of a division and compartmentalization of tasks. According to Tuomela, collective commitment, the "glue that binds we-mode groups together", serves two purposes necessary for a group to function as an agent. First, it serves to establish the unity and identity of a we-mode group by binding the members together around the groups ethos. Second, it provides the group with the authority to decide about its members' activities in a practically efficient way (cf. Tuomela 2013, 45). Without collective commitment, Tuomela states, "the members could not coordinate their activities or perform together effectively to achieve group goals" (ibid). His analysis of collective commitment is the following:

> **(CoCom$_{Int}$)**: Members A_1, … A_i, …, A_m of group g are collectively committed to performing X jointly as a group if they jointly intend to perform X together as a group, that is, ‚Jointly-intend-qua-members(we, we perform X jointly as a group)' is true of them qua members of g, and this whole sentence, which is about the fact that they so intend, has the mind-to-world direction of fit, while the intention content, viz., ‚We perform X jointly as a group', has the world-to-mind direction of fit. In addition, because they have the aforementioned intention, the members ‚group-normatively' ought to participate in the performance of X together as a group (and thus to we-intend so to participate) (Tuomela 2013, 83).

Again, as with the condition of collectivity, Tuomela's analysis of the collective commitment condition seems to imply that individuals are jointly committed to realize a single, monolithic action, which is known to all of the parties of the collective commitment.

If, however, we can reasonably assume that through the process of compartmentalization and on the basis of positional hierarchies, an institutional group's overarching goal is broken up into smaller, fragmented sub-

[64] I above mentioned a group running up a hill where this is the group's goal (e.g., to run up a hill) which is only satisfied simultaneously and interdependently for all of the members of the group, i.e., if all the members reach the top of the hill. But my point is that, within this example, a group's goal could consist of several hills to climb for different members of the group, unbeknownst to each other. Of course then, for this goal to be objectively satisfied, this necessarily requires the satisfaction of said sub-goals due to the group members' actions of each running up their assigned hill. When each member reaches the top of her hill, the group's overarching goal of running up the hills is accomplished. And this is true even if the members do not know where and how many other hills have to be climbed by others.

goals, some of which - by design or out of sheer complexity - may not be known by all group members, then how should individuals, acting in light of their task-specified positions, be collectively committed to this overarching goal? Is it really plausible to assume that individual group members, being assigned domain-specific tasks by their hierarchically defined position in the group structure, could *not* "coordinate their activities or perform together effectively to achieve group goals" (Tuomela 2013, 45) simply because they do not "jointly-intend-qua-members that (we perform X jointly as a group)"?

While the Calutron Girls may had fully grasped what *their individual positional task* required them to do, they may had only a vague or even false understanding about the extent and scope of the overall group action they were contributing to. Yet, it sufficed for them to simply be *individually committed* to fulfilling this positional task, specified by their position in the group (Do what you're told. Don't ask why!) to still be able to act *as group members.*

To recap, I think that Tuomela is in a good position to explain how institutional groups, when conceptualized as a multidimensional structure of interrelated positions, can compartmentalize overarching actions into compartmentalized sub-actions. This ultimately relies on the establishment of a hierarchy, which his positional theory offers a good explanation for. Tuomela, however, relies too heavily on a notion of collective action which includes individuals engaging in the *we-mode*. Conceptualizing institutional agency in terms of such a we-mode, as I tried to show, is too restrictive. The compartmentalization of tasks can passively circumvent, or even actively prohibit individuals to think in terms of the whole group, i.e., to occupy a group-perspective on the collective action when acting for the group. Yet, individuals can act for the group by performing component-actions without having such a perspective. In a way then, Tuomela's theory of group agency turns on itself in a self-defeating manner. Once we take Tuomela's positional view seriously, the we-mode framework it is building up upon becomes implausible. We then might try to save the positional aspects of Tuomela's theory from Tuomela himself. In the third chapter, I will heavily utilize the concept of institutional roles, which, in important aspects, are like Tuomelian positions, just without the "collective baggage" of the we-mode framework.

2.3. Summary

The main goal of this chapter was to provide the reader with an overview of theories that try to answer the question of how we should understand claims about institutional group agency. To this end, I focussed on two broad avenues for explaining institutional group agency. The first path of r*ealist, non-reductive (or inflationary) theories of group agents* argued that the agency of institutional groups is to be explained by arguing for these groups to be genuine, non-reductive agents; or agents *in their own right*. I tried to show that the examined theories are all vulnerable to reductionistic charges.

French, for example, seems to conflate the possibility that an institutional group's CID-Structure can *represent* intentions to perform an action, with the intentions to perform the action itself. Bratman's theory, on the other hand, by looking only at the downstream effect of rule-based decisions and attributing these outcomes to the institutional groups themselves, seems unable to recognize the influence that -potentially powerful - individuals have on institutional group actions. By focussing merely on the outputs of what he calls an institutional intention, he seems to neglect that -at least often times - it are individual agents, who

provide the only input. List & Pettit argued that certain decision-procedures seem to give rise to group-level attitudes (and thus group agents) which are somehow disconnected from (and irreducible to) the attitudes of the individual group-members. Pace List & Pettit, I tried to argue that this only shows that members of a group may choose decision procedures because of *their individual desire to be collectively rational.* This, however, amounts to nothing more than *individuals choosing a procedure*, and all the intentional states involved are held by individuals. Finally, Tollefsen argued that because our practices of *interpreting* groups as if they were agents are successful, we are warranted to claim that they actually *are* agents. To this end, I tried to show that her argument seems to rely on a small but crucial mis-analogy between Dennett's *physical* and *interpretative* stance, and her *individual* and *collective* intentional stance. It follows, I argued, that she fails to show that something (or rather: anything) is left unexplained when applying the individual intentional stance to groups themselves.

In a next step, and instead of explaining institutional actions in terms of a non-reductive institutional agent, I tried to argue that we should rather think of institutional group actions in terms of *reductive* or *deflationary* theories of *collective action*.

Thus, the second part of the chapter examined theories which argue that the agency of institutional groups is *reducible to* the capacity of the institutional groups' members for *collective action*. I argued, that theories of collective action which start out with small-scale cases of collective action face the so called *Upscaling Problem*. As such direct attempts to apply small-scale cases of collective action to institutional groups lead to frustrating results, I went on to examine two theories that explicitly target the agency of large, and complexly structured institutional groups. Christopher Kutz's minimalist account of collective action claims that the concept of *overlapping participatory intentions* is key to understanding the forms of collective action that characterize institutional groups. Such participatory intentions are formed if an individual agent understands her individual action as *contributing to a collective end* and thinks of her actions as *doing her part* in order to contribute to that end. One of the merits of Kutz's account is that he distinguishes between so called *executive* and *subsidiary perspectives* that members can have on an institutional group's action. To this end, however, his theory seems to rely on some sort of presupposed hierarchy of members concerning their authority to act on behalf of a group. I argued that simply *presupposing* the existence of such a hierarchy won't do the trick, for we have to have some criteria which allows us to say whether a group member is actually justified in occupying either of these perspectives. Next up was the we-mode framework of Raimo Tuomela's positional theory of group agency, which seems to amend this problem. Here, I argued that his emphasis of the *we-mode* in explaining institutional agency must be rejected as too restrictive.

During the course of this second chapter, I hope to also have provided the reader with reasons to think that a catch-all analysis of our capacity to cooperate will eventually fail to acknowledge the fact that cooperation is a heterogenous and multi-facetted phenomenon. Thus, the characteristic features of *institutional* cooperation will require from us a fundamentally different analysis than, e.g., taking a walk or dancing the tango. My conjecture here is that the isolated, individualistic and often alienating character of institutional actions should shape our analysis of institutional agency in two ways: First, to say that institutional actions have an *individualistic* character reveals how ultimately, we should refrain from invoking non-reductive, autonomous intentional states being held on a group-level, in order to explain what is going on in cases of institutional action. Second, the individualistic character of institutional actions should provide us with

reasons to doubt that it can be analyzed in terms of strongly shared, interdependent collective intentional states. Thick notions of acting together in terms of a *we-mode*, a shared *group mindedness*, a *sense of us,* etc. will all fail to acknowledge this individualistic character of institutional actions.

So in the next chapter, I will try to navigate the scylla of non-reductive group agents and the charybdis of theories of collective action unfit to account the actions of institutional groups. I will do so - to inappropriately stretch this metaphor - by tying myself to the mast of role-based explanations of institutional action.

3. Role-Based Theories of Institutional Group Agency

> *It is perhaps because many of us know what it is to spend an afternoon baking biscuits that there is something striking about encountering a company which relies on the labour of five thousand full-time employees to execute the task. Maneuvers which one might briefly have carried out on one's own in the kitchen (readying an oven, mixing dough, writing a label) had at* UNITED BISCUITS *been isolated, codified and expanded to occupy entire working lives. Although all employment at the company was ultimately predicated on the sale of confectionery and salted snacks, a high percentage of the staff were, professionally speaking, many times removed from contact with anything one might eat.*
>
> Alain de Botton: The Pleasures and Sorrows of Work

Both explanatory paths to explain the agency of institutional groups seem to be blocked. In this third chapter, I want to examine an alternative route to explain the agency of institutional groups. This alternative route is offered by *role-based* explanations of institutional agency (or *role-accounts of institutional agency)*. The main line of reasoning of these accounts can be summarized the following way:

> 1) An institutional group action consists of (and consequently can be reduced to) the individual contributory actions of its members, who act in their assigned roles.
> 2) For members of an institutional group to act in their assigned roles is to perform the functions and tasks definitive of these roles.
> 3) To perform the functions and tasks definitive of institutional roles is to act on the deontic powers, i.e., rights, duties, responsibilities etc. that define the (inter-related) roles.

So whenever an institutional group acts, it is because those who have appropriate roles in the institution at that time act (cf. Ludwig 2017b, 260). In the following chapter, I will defend such a role-based explanation of institutional agency. This chapter is building upon and synthesizing existing work in this field of research, e.g., by Katherine Ritchie (2020a), and more importantly by Seumas Miller (2001; 2010; 2019) and Kirk Ludwig (especially 2017b). While the presented accounts were developed independently of each other, and do not call themselves "role-based" theories of institutional action, they fundamentally agree on several key features:

First, they give deflationary (or reductive), and individualistic explanations of institutional agency. The presented accounts therefore explicitly dismiss the non-reductive *group-agent strategy* of institutional agency. Kirk Ludwig's ultimate goal, e.g., is to "understand institutional agency in terms of individual agents and the concepts already deployed in understanding individual agency" (Ludwig 2017b, 237). Ludwig's approach to institutional agency, inspired by Davidson's Causal Theory of Action, including his method of semantical event analysis of action sentences, and John Searle's work on social ontology, is deflationary and reductive at its core. He therefore proposes his so called *multiple agents account* of institutional group

action. According to this account, to say that an institutional group did something is to say that there was an event of which each individual member of the group, and only those individual members, were the direct agents of.

> **Multiple Agents Account:** Collective action is a matter of all the members of a group (and only them) contributing (in the right way) to bringing about some event or state (from: Ludwig 2020c, 81; see especially: 2017a).

So what we are inclined to call a *group agent*, to paraphrase Ludwig (cf. Ludwig 2017a, 296-298), really is (just) a *group of agents.* Institutional group agency, then, involves multiple agents of a single event. It does not, however, involve an event of which there is one single group agent. Still, the group doesn't disappear. It is indeed relevant to talk about group agency, or group action. What Ludwig wants to emphasize, however, is that only the individual members of a given institutional group are the agents of anything. And although he acknowledges the rich and sophisticated practices by which we attribute actions to such groups, he aims to show that talk of institutional agents ultimately is a *façon de parler* (ibid). Upon closer inspection of our talk about the actions of institutions, such as corporations,

> "we find behind the curtains only individual agents. There is no super-agent, no corporate agent as such, and no need for one. There is therefore no need for a super-subject of psychological attitudes distinct from the various role players in a corporation, its shareholders, directors, management, and employees, hovering over them all, with its own mind, doing its own thing" (Ludwig 2017b, 237).

Similarly, Seumas Miller claims that the actions of institutional groups are reducible to actions of individual human persons (cf. Miller 2001, 160). Miller, too, demarcates his theory of institutional group action from the above mentioned theories of group agents, as he does "not accept that macro-entities, such as nation states and corporations, have beliefs and intentions, and consequently are either rational or moral agents. Properly speaking, all social actions are performed by individuals, not social entities" (ibid).
Miller analysis of what he calls *joint action*, too, is deflationary and reductive, i.e., he argues that social actions are reducible to forms of individual, interpersonal actions without the need to invoke any notion of supra-individual entities (cf. Miller 2001, 6). This idea also travels to the actions of institutional groups, which Miller calls *corporate actions*. While Miller also acknowledges the *usefulness* of speaking about the actions of institutional groups, *strictly* speaking, "there are no such things as corporate actions" (Miller 2001, 53) but only individual actions which are related in a certain way.
The second commonality of the presented accounts follows directly from such a deflationary explanatory strategy. The role-based explanations of Miller and Ludwig fall into the camp of promoting a *collective action* account of institutional agency. However, contrary to the accounts of collective action that I discussed in the second part of chapter two, the authors do not centrally focus on the analysis of collective intentional states of the group members (e.g. in forms of Bratmanian shared intentions, joint commitments, or a we-mode) in order to explain the actions of institutional groups. Rather, they invoke the notion of

collective intentional states in an *indirect* way and primarily explain the agency of institutional groups via *individuals performing the functions and tasks of their assigned institutional roles*. Institutional roles then figure in the explanations of the actions of institutional groups front and center.

But what do I mean with collective intentional states being *indirectly* involved in collective action? When acting in an institutional role, I will argue, individuals do not have to possess collective intentional states in order to directly cooperate with other group members. There has to be no reference to an irreducible "we", or the group they belong to, nor a joint commitment, nor collective intentional states in form of shared contents plus mutual beliefs, in order for two individuals occupying institutional roles to cooperate and contribute to a collective action. For individuals to engage in collective action in virtue of the roles that they occupy, it suffices that they carry out the tasks and functions of their roles. Nonetheless, collective intentional states do play a role. They come into play when explaining how such institutional roles viz. their functions come about in the first place.

Katherine Ritchie describes this indirect involvement of collective intentional states with the Wittgensteinian picture of knocking away the "intentional ladder", upon which one has ascended to a level of organizational complexity, where direct cooperation based on collective intentional states is no longer needed:

> "It might be true that a complex representational account with symmetric attitudes is required to create an organized group with a particular structure composed of roles and relations. […] Yet, once an organized group with various defined roles exists, minimal cooperation need not involve anything as mentally complex as that posited by Bratman, Gilbert, or Searle. […] The organizational structure allows for group members to minimally cooperate by playing particular roles. Once a group is ‚built' the ladder of collective intentionality that was used in constructing it can be ‚knocked away'. Once an organizational structure is in place, complex mental work involving mutual knowledge, representations of others' mental states, and formation of joint commitments are not necessary for cooperation" (Ritchie 2020a, 101f.).

For Ludwig, the existence of what he calls "status roles" rests on the collective acceptance of constitutive rules which ultimately rely on his theory of *we-intentions*. For Miller, the central concept in his *teleological* account of institutional agency is that of *joint action,* which, according to him, consist of the intentional individual actions of a number of agents "directed to the realization of a collective end" (Miller 2019). This is the "Collective End Theory" (CET) of joint action (Miller 2010, Ch.1). Miller's core-theory of joint action requires that for two or more individual actions *x* and *y,* performed by agents *A* and *B* to constitute a joint action, there has to be mutual belief on part of both the agents. However, once we move to the level of institutional actions carried out by individuals *occupying roles*, things are different. Here, I will argue that the *non-specifity of role-occupants* does not require that individuals have mutual beliefs or shared intentional states.

The accounts then converge on the claim that institutional group agency can be reduced to, and explained by individuals performing the functions of the institutional roles they come to occupy. They also, as I will try to show, do not succumb to the problems in explaining the compartmentalized and anonymous cooperation

within institutional groups that the accounts of collective action in Ch. 2.2. faced. They are therefore apt to explain how individuals, on the basis of occupying roles, *can cooperate with one another while remaining anonymous to each other.*

A third fundamental similarity, which I will examine in depth in the following section, is that both Ludwig, Miller, and Ritchie share an understanding of what institutional groups and institutional roles are. This might seem trivial, yet I will argue that this common understanding of institutional roles is where the accounts eventually go astray. A detailed analysis of this will be provided below. But let us construct and ascend the ladder first, before eventually knocking it away. For now, I will outline the overall line of reasoning that I want to defend in the next three chapters. It takes the following steps:

(I) Institutional group agency can sufficiently be explained by individuals acting in institutional roles. (Ch. 3)

(II) Institutional roles involve the performance of assigned tasks and functions, which is based on the exercise of deontic powers. (Ch. 3)

(III) An individual's performance of an assigned institutional role cannot be sufficiently explained by the individual's acting on such tasks and functions (Ch. 4.)

(IV) My account of "Role Agency" can explain the shortcomings of the presented theories of institutional agency. It fills the lacuna of the existing theories (III) by highlighting the relation between individuals and their roles. (Ch. 5)

(C) (From III and IV): A role-based explanation of institutional agency needs to involve an account of "Role Agency". (Ch. 5)

Regarding (I): I will try to show that role-based accounts offer perspicuous explanations for the agency of especially large and complexly structured institutional groups. With such role-based explanations, we can see how institutional groups persist through time and survive the change of membership, and they also explain how institutional groups establish hierarchies of power which allow for the division of (cognitive and manual) labour and tasks. But role-based explanations are also apt to explain how cooperation within institutional groups can occur *anonymously* thereby encompassing cases like the above mentioned Calutron Girls.

Regarding (II): The definitions of institutional roles given by Miller, Ludwig and also Ritchie converge on the idea that institutional roles are primarily defined through tasks and functions that the role-occupants need to fulfill, and that the individuals achieve this through the use of deontic powers, such as rights, duties or responsibilities. This way of characterizing roles has clear explanatory merits. We can, for one, explain how groups come to have a structure, how they survive the change of members, and how roles establish specific member-to-action and member-to-member relations of power, leading to the division and specialization of labour. But while I hold this standard way of defining institutional roles to be fundamentally correct, I also

hold it to be *lopsided*. Regarding institutional roles, there really is a tripartite relation between (1) institutional groups, (2) institutional roles, and (3) the individuals who occupy the roles (see: Box 2 in Ch. 4.). The presented theories can be described as focusing on the former relation between groups and roles (1)-(2), while *neglecting*, or *under-theorizing* the latter relation, i.e., the way in which individuals relate to their institutional roles (2)-(3). This, however, leads to certain problems.

Regarding (III): Here, my claim is based on two objections, which I call the *two Problems of Discretion.* The first objection is what I call *institutional stupor*. I will show that a group's capacity for action can break down, i.e., that institutional groups can (partially or fully) lose their agency, although, or especially *because of* all formal aspects of the institutional roles are being fulfilled to the letter. Empirical data, e.g., studies of so called *work to rule strikes* (see, e.g., Scott 1998; Bloch & Moorman 1993; Kühl 2022) shows, seemingly paradoxically, that if individuals perform *exactly* what their roles require them to do, institutional groups stop to function. The second, and related objection is that of the phenomenon of *role-ambivalence*. I argue that even when the deontic powers of a role are formally established, an individual's functioning within a role can be partially impaired or even made impossible because of insufficient instructions, changing circumstances, or demands for spontaneous decision-making in uncertain situations. The overall goal of discussing the two *Problems of Discretion* is to show, how the existing approaches cannot sufficiently explain how individuals can perform the tasks and functions of their roles, even if -or especially *because*- these roles have *discretionary powers* attached to them.

Regarding (IV): Exploring what I call "Role Agency", which is done in Chapter 5, aims to fill in for the shortcomings of the presented theories of institutional agency. It does so by highlighting the relation between individuals and the institutional roles they occupy. Role Agency describes the ability of an individual to reflexively engage with, and act on the role that she is assigned in a group context. To this end, Role Agency describes more than merely exercising one's deontic powers or fulfilling certain tasks or functions (role-*taking)*. It captures the ways in which individuals internalize, interpret and alter their assigned roles and the corresponding tasks (role-*making*). Understanding the concept of Role Agency also allows us to make sense of the two *Problems of Discretion*. My concept of Role Agency ultimately aims to make intelligible the ways in which individuals develop a reflexive self-understanding of, and agential identification with their assigned roles. Such a reflexive self-understanding, in turn, can explain how individuals shield the agency of their assigned roles against the backdrop of institutional rigidity and inflexibility, and how they come to internalize both the formal and informal aspects of their roles. But Role Agency can also explain how role-occupants can prevent their discretionary powers from turning *toxic*. This is because they may engage in externally-fixed, action-guiding role-*idealizations*, which provide regulative standards for them to navigate through ambiguous usages of their discretionary powers.

So let us go through each step, starting with a role-based explanation of institutional agency. As I mentioned above, the presented role-accounts all share a similar understanding of institutional groups. So let us start with the question of which type of groups we are talking about when considering role-based explanations of institutional groups.

3.1. A Structural View of Institutional Groups

The starting point here is the above introduced definition of institutional groups as organized groups. Recall again the above introduced definition of institutional groups.

> **INSTITUTIONAL GROUPS:** Institutional groups are organized groups that consists of an embodied (or realized) structure of differentiated roles. Institutional roles are defined in terms of their interdependency, and in terms of tasks that individual members have to perform, as well as rules regulating how these tasks are to be performed.

As already indicated in the introduction, Katherine Ritchie (2013; 2015; 2020a; 2020b) adheres to this view of institutional groups as organized groups. Organized groups, again, are entities that have organizational *structure* (cf. Ritchie 2020a, 95; also: 2013; 2015). Organized groups, however, are not *identical* to these structures. They are structures that are, or have been *realized* by individuals. Similarly, Seumas Miller here also talks of organized groups being *embodied* structures (cf. Miller 2001, 28; see also: 2001, Ch. 5; 2010, Ch. 1; 2019, Sec. 1). Organized groups are thereby not abstract entities but social groups consisting of individual members *realizing,* or *embodying* a structure.

This definition of institutional groups as structures of roles corresponds with Kirk Ludwig's account of institutional groups (see especially 2017b). Ludwig draws a basic distinction between groups formed under so called ∈-membership and groups that are formed under so called ε-membership relations, which matches Ritchie's distinction between unorganized and organized groups. Groups formed under ∈-membership relations, which Ludwig calls "natural groups" (Ludwig 2017b, 160), are (nothing more than the) mereological sums of their members. So for any individuals I_1 I_2 I_3 - I_n, there is a ∈-group consisting out of exactly those individuals (ibid). To "think of such a group is to think of some individuals either by enumeration or as those gathered under some concept, and to be a member of such a group (an ∈-member) is simply to be one of them" (ibid). Membership in such ∈-groups is "free" (ibid), i.e., there are no further membership criteria.

Ludwig's second type of group are groups with ε-membership relations (Ludwig 2017b, 161-162). Here, things are different. To be a member of such an ε-group, individuals have to fulfill specific, socially-constructed conditions of membership. Such ε-groups can be further divided into unorganized ε-groups (such as mobs and crowds) and *organized* ε-groups (such as organizations and corporations). Now institutional groups fall into the latter camp, i.e., they are *organized* ε-groups.

To say that an ε-group is organized, is to say that the members of an ε-group have membership-conditions which can be explicated by what Ludwig calls *status roles*. ε-membership in a group that involves having such a status-role requires for there to be *collective acceptance* of an arrangement of *constitutive rules* in a relevant community (details will be provided below). According to Ludwig, an "organized differentiation of roles directed toward joint action, which may be occupied successively by distinct individuals" (Ludwig 2017b, 2) then is the "hallmark" (ibid) of organized institutional groups.

Institutions (or institutional groups), like the *Supreme Court of the United States*, the *British Parliament*, the *World Bank*, the *Chinese Communist Party*, and the *Times Corporation*, are groups that "are united by an ε-

membership role that organizes (at least some) ε-members for *joint action* in those roles, and so which require of those who act in those roles acceptance of them" (Ludwig 2017b, 161). Ludwig thus defines institutional groups (or just: institutions) as a "set of transferrable roles inter-defined in terms of their functions established to collectivize behavior for one or another purpose" (Ludwig 2020a, 181). As such, an institutional group exhibits some remarkable features, as it is

> "designed for persistence through change in members. It can bridge periods in which it has no members. It can undergo reorganization. Subunits can be assigned and reassigned. Its identity does not seem to be tied to the identities of the individuals who at any time constitute its membership. It can carry out tasks over periods of time in which its membership changes and through periods of time in which its membership changes completely" (Ludwig 2017b, 1).

Ludwig states that institutional groups cannot simply be equated with their members. This is because they have the feature of being *structured* in virtue of the status roles being *inter-defined* in terms of their functions (cf. Ludwig 2020a, 181) (more on this below). For now, we can explicate that this inter-definition of status-roles grounds the *structure,* the possibility of hierarchies and the diachronically robustness of institutional groups. Further, status roles can be iterated and multiply realized. One status role can be realized and occupied by different individuals at different times, so that status roles can remain diachronically intact throughout the change of occupants. Also, one individual can occupy multiple status roles within an institutional group. Further, status roles are *recursive*, i.e., the assignment of status roles can be based on the status roles of other individuals. So individuals occupying certain status roles can, via having such a role, create (or alter) other types of status roles. This allows for institutional groups to exhibit complex and extensive structures, which arise out of such simple features.

Institutional roles

On the presented accounts, institutional roles are agreed to be both *relational* and *functional*. Ritchie defines roles through "tasks that role-players are allowed or required to carry out" (Ritchie 2020a, 95f.) and "powers, norms, or responsibilities pertaining to [these] particular tasks" (ibid). She further defines roles in terms of "relations to other roles, tasks that role-players are allowed or required to carry out, and in some cases specific features a role-player must have" (ibid). Ritchie states that the inter-relations between roles can be hierarchical, or non-hierarchical depending on whether the deontic relations between roles encompasses relations of power. Roles that involve deference to and (or) authority over the actions of other group-members are hierarchically related. For instance, a role might allow one role-player to give orders to individuals playing another role (ibid). In turn, relations of seconding a motion or reporting on a project involve relations between role-players that are non-hierarchical (ibid).

For Miller, too, organizational roles are functionally defined in terms of tasks that an agent is to perform and rules, procedures and/or conventions that regulate how an agent is to perform these tasks (Miller 2019 Sec. 1; 2010, 47).[65] Consequently, Miller characterizes roles in the following way:

> "I suggest that a role is simply an abstraction from procedure-governed tasks, and the actions that constitute the undertaking of those tasks. Qua abstraction, the role is simply a specification of the task or tasks, and the procedures that govern it. To say that the role is filled is just to say that some individual is undertaking that task or tasks, and following those procedures" (Miller 2001, 172).

Miller describes roles to be relational in that they are *interlocking*. By this he means that they are related to one another in terms of hierarchy and interdependence. The central idea of role-interdependence is that one role can be defined by a task (e.g., packaging pins) that can only be carried out if one or several other role-occupants undertake their actions specified by *their* tasks (e.g., wire-drawing; -straightening; -cutting; -pointing; -grinding etc.). Just like in the above mentioned accounts, Miller emphasizes that interrelated and interdependent roles in organized groups create *structures*. Here, too, roles can be hierarchically and/or non-hierarchically related; and involve different levels of status and different degrees of authority (see Miller 2019). Because roles are defined in terms of tasks, and *not* in terms of the *specific* individuals performing these tasks, it follows that one and the same set of tasks can be performed by different individuals throughout time.

Regarding institutional roles, Ludwig establishes the above mentioned concept of *status roles.* To this end, he draws heavily on Searle's (1995; 2010) concepts of *constitutive rules*, *collective acceptance* and *status functions*. I will clarify these notions in detail below. For now, and leaving some details aside, this idea can be broken down the following way: Just as with objects, the assignment of status functions to agents is defined by constitutive rules for joint action types. So status roles are status functions that are assigned to individual agents, which are to exercise their agency in fulfilling the functions definitive of the role (cf. Ludwig 2020a, 185). While there is an abundance of examples of such status roles, e.g., judges, CEOs, clerks, police-women, soldiers, managers, workers, call-center operators, kindergarten teachers, spokespersons, corporate risk-assessment strategists, nurses, lawyers, regional sales managers, assistant (to the) regional manager, etc., such instances of status roles have several key features in common. First, any individual that comes to occupy a status role is, on this basis and in virtue of this being collectively accepted,

i) to give or accept directions or permissions to or from others, given their status roles, or
ii) to play certain roles or do certain things in joint activities or types of social transactions
    a) on certain conditions obtaining or

65 Miller differentiates social institutions from organizations because for him, the former, but not the latter necessarily have a normative dimension: "Organisations with [a] normative dimension are social institutions. So institutions are often organisations, and many systems of organisations are also institutions" (Miller 2001, 29). As my definition of institutional groups does not encompass such a normative dimension, I will proceed to take Miller's talk about organizations to be applicable to my definition of institutional groups, and hence, talk about institutional groups.

b) at their discretion or
c) upon the exercise of their judgment about certain matters

which all parties to the arrangements are to act in conformity with, in accordance with their own status roles (cf. Ludwig 2020a, 186; also: 2017b, 138f.). Several things follow from this description.

First, status roles are *relational* in the sense that they specify how the individual occupying a status role is to interact with *others* in virtue of *their* status role (cf. Ludwig 2020a, 188). Status roles are *interrelated*, i.e., some status role can only fulfill its assigned functions in virtue of there being *other* corresponding, or interlocking status roles. Status roles may give the individual certain powers in regard to the relation to other members, i.e., they may encompass a member-to-member relation of power. Second, status roles are task- or action-specific. There are things that one has, needs, or is allowed to do when occupying a status role. So they encompass a member-to-action relation of power as well. Status roles are in this sense *agency-directing*, by which I mean that the bearer of a status role can make use of her powers only in some definite context, or institutional arrangement. Status roles are also *agency-enhancing*, by which I mean that individuals that occupy status roles possess certain (member-to-member; or member-to-action) powers and on this basis come to have certain agential abilities, which they would not have outside of the institutional context. How can we further analyze this? According to Ludwig, the powers that individual agents may have in virtue of their status roles include:

i) directing or
ii) giving permissions to others,
iii) exercising rights,
iv) making findings that have an official status that others must conform their behavior to,
v) issuing rules (which are generalized directives), and
vi) conferring status roles on others or status functions on things (Ludwig 2020a, 186).

Regarding these powers, two further distinctions can be made: First, one can distinguish between two *modes* of power, i.e., *positive power* and *negative power*. Positive power can be defined as *enabling* an agent to do something in virtue of her status role which she could not do in virtue of her mere physical properties. In turn, negative power *requires* the role-occupant to do something. Ludwig then distinguishes between *types* of status roles which accord to these two modes of status power. A role that primarily involves directing others in virtue of their roles, i.e., a role which is based primarily on positive power, is a *command* role. A role that primarily encompasses the exercise of negative powers is a *compliance* role. According to Ludwig, status roles "may and often do combine both of these elements, but need not involve either" (Ludwig 2017b, 139).

Those powers, which can be expressed through deontic modals such as rights, responsibilities, obligations, duties, privileges, entitlements, penalties, authorizations, permissions, etc. (cf. Searle 1995, 100f.) are "real powers, which make a difference to the causal evolution of the world, and effect constitutive changes to the fabric of social reality" (Ludwig 2017b, 141f). But they also supervene entirely on the relevant community

collectively accepting that the individuals occupying such status roles actually have them (ibid). Ludwig emphasizes the importance of collective intentional *activity* for understanding these powers. On the one hand, he claims that these powers rest on there being collective acceptance of the relevant community. And on the other hand, he connects the concept of collective acceptance to so called *we-intentions.* We-intentions are intentions held by individuals directed at the group doing a particular thing together (cf. Ludwig 2017b, 22). So these deontic powers, linked to status roles, then boil down to this: When we confer on someone a status role that involves the exercise of deontic power, w*e are committed to engage as appropriate in collective intentional behavior* in which that person plays the relevant role (cf. Ludwig 2017b, 141f.).

To gain further clarity on the concept of status roles, Ludwig develops a taxonomy of status roles and their corresponding forms of collective acceptance regarding the occupancy of such roles. First, he differentiates between *agent status roles* (or *agent roles*) and *subject status roles* (or *subject roles*) as well as hybrid forms of the two:

> "Agent status roles presuppose that the person to whom they are assigned is party to the collective acceptance that imposes the role on her. The second sort, I will call subject status roles, in contrast, are assigned to agents independently of whether they accept the role. Being a university professor is an agent status role. Being a prisoner of war or *persona non grata* are subject status roles. Even functions of subject status roles involve their possessors exercising their agency in those roles in the sense that they are supposed to recognize the roles assigned to them and recognize that roles as such involve behaviors on their part in various circumstances. However, they need not participate in the collective acceptance that assigns them the role (e.g., prisoners of war), and there is no presumption that they will willingly fulfill the role assigned, hence, the need to make provisions for various forms of coercion" (Ludwig 2020a, 187).

Second, and related to this, is Ludwig's refinement of status roles being accepted either *formally, or substantively.* Individuals who *substantively* accept their assigned status role are "assumed to be sincerely engaged, when appropriate, in the kinds of joint activities defined in part by the someone performing the functions of the role" (Ludwig 2020a, 188f). Here, and in light of the problem of alienated group members, Ludwig argues that substantive acceptance of a status role inevitably comes "under pressure from the real-world conditions under which human agents make as if to sign on to them" (ibid). By this, he means that - especially for large-scale institutional groups - "it becomes more important that organizations can rely on someone not to capriciously opt out" (ibid). This is captured by status roles being (merely) *formally accepted,* where formal acceptance of a status role "is a public acceptance of the role that represents oneself as committed to fulfilling the role's functions" (ibid) without possibly substantively accepting the role. However, individuals that only formally accept their status role can be kept in line, as institutional groups may install arrangements that incentivize such individuals to fulfill the duties of their roles either by forms of or rewards for doing so (e.g., receiving a paycheck) or punishment for failing to do so (disciplinary dismissals, written warnings etc.), or a mixture of both (ibid).

A useful, summarizing view of institutional groups as consisting of a structure of differentiated roles is provided by Garcia-Godinez. According to Garcia-Godinez, institutional roles can be viewed as "job descriptions", that "specify both the requirements for role-occupancy and the deontic powers attached to it“:

> "Think e.g., of a corporation. A corporations consists, roughly, in shareholders, a board of directors, officers and employees. To be a shareholder, an officer (e.g., a Chief Executive Officer or CEO) or even an employee, one needs to satisfy certain conditions. These conditions as well as the deontic powers attached to the roles are (however vaguely) established in their associated job-descriptions. For example, a CEO is responsible for making major corporate decisions, managing the day-to-day operations and resources of the company, being its ‚public face', etc. So, whoever takes on the role of CEO will hold such responsibilities […] By generalizing, we can say that the institutional roles and relations that correspond to a certain formal group structure are normatively determined by the whole network of job-descriptions" (Garcia-Godinez 2020, 49f.).

So while the view on institutional roles examined may differ in detail, their commonalities can be summarized as following: Institutional roles are defined through the tasks and functions that an individual occupying such a role must fulfill. The tasks and functions of institutional roles, which Ludwig calls the "design specifications" (Ludwig 2017b, 149), can be characterized through the exercise of (positive or negative) deontic powers, such as rights, duties and responsibilities. Institutional roles are *related* and *interdependent* insofar as they rely upon each other for the exercise of the deontic powers and for the fulfillment of the assigned tasks and functions. Institutional roles can be occupied by different individuals throughout time, so that if one individual stops to fulfill the tasks and functions of her role, another individual can chime in and take over this role.

Collective acceptance as an adequate basis for institutional roles?

To occupy an institutional role and carry out its functions is not something that one can do in private, as roles are both interrelated and defined through deontic, i.e., interpersonal powers. One does not become, e.g., a policewoman by individually deciding that one has the right to arrest people, or the duty to regulate traffic. The seminal way of explaining how institutional roles come about, is by arguing that individuals have to *collectively agree*, or *accept* that a given role has these deontic powers (Searle 1995; 2010).[66] A worry we

[66] Now, the term *collective acceptance* or *recognition* is used by two philosophical schools in rather different ways. On the one hand, the term recognition is used by political philosophers in the tradition of critical theory, which dates back to a Hegelian tradition of using the term, i.e., *Anerkennung* (cf. Hegel 2018). The critical theorists' use of the term (see, e.g., Honneth 1994, 2007; Fraser & Honneth 2003; Schmidt am Busch & Zurn 2010) applies primarily to problems in practical philosophy, e.g., the fair distribution of political power, the "struggle for recognition" of political minorities, or the relation between individual personhood and socially recognized forms of identity. I will leave this debate aside and focus on the use of the term in the contemporary, analytical tradition of social ontology. For a comparative overview of the usage in both traditions, see: Ikäheimo & Laitinen 2011.

need to address is that the concept of *collective acceptance* may not provide an adequate basis for institutional roles to begin with.

To see why, note that the core concept of collective acceptance was initially introduced into the debate by John Searle (1995). Collective acceptance, for Searle, is a special form of collective intentionality. Searle conceptualizes institutions as a collectively accepted system of rules (procedures, practices) that enable us to create institutional facts. And according to Searle, what makes collective intentional states, including collective acceptance, special cannot be analyzed with regards to their *content*, nor to a special, collective *subject*. Instead, Searle initially proposes to analyze collective intentional states in terms of the above mentioned *mode-account* of collective intentional states, according to which collective intentional states are held by individuals who are engaged in a special mode of "we-intentionality".[67] Regarding this mode, Searle initially made the following negative claim: "[W]*e-intentions are a primitive form of intentionality, not reducible to I-intentions plus mutual beliefs"* (Searle 1990, 407).[68] However, Searle later revised this view on the irreducibility of collective intentional states in the case of collective acceptance or recognition. After criticism of his account,[69] Searle gave in to the possibility of there being collective acceptance or recognition that can indeed be reduced to individual "I-Intentionality" plus mutual belief. This is because, as a "much weaker form of collective attitudes", collective acceptance, or recognition can be conceptualized without there being the feature of *cooperation* occurring (cf. Searle 2010, 56f.). A couple, who is planning marriage, Searle says, "accept[s] the institution of marriage prior to actually getting married. This is not a case of cooperation in a form of behavior but simply going along with an institution" (Searle 2010, 57). So collective acceptance, according to Searle, can occur with all that is required for it to be the case is "that each participant accepts the existence and validity [of the institution] in the belief that there is mutual acceptance on the part of the others" (ibid). He concludes that the existence of an institution "does not

[67] It is not just a special mode of intending that makes intentionality genuinely collective, as collective intentionality for Searle has certain pre-requisites that must be in place in order for it to come about. Searle says, that intentional states do not function in isolation and for any particular form of collective intentionality to exist, this will acquire what he calls the "background" by which he means a "set of capacities, abilities, tendencies, habits, dispositions, taken-for-granted presuppositions, and ‚know-how' generally" (Searle 1998, 107f). Also, Searle says that, additional to the "background", human individuals have a biological capacity "to recognize other people as importantly like us, in a way that waterfalls, trees and stones are not like us", a capacity he calls the "*preintentional sense of ‚the other'* as an actual or potential agent like oneself in cooperative activities" (Searle 1990, 413, own emphasis). Second, this sense of others as a candidate for cooperative agency comes with what Searle calls the "sense of us", i.e. the coalescence of this "preintentional sense of ‚the other' by multiple people to form a feeling of collectivity" (ibid). Searle has been both criticized for leaving the Background and the "sense of us" at rather vague terms. Scholars have since tried to make intelligible how this "sense of us" could be further analyzed (see, e.g., Martens 2018; Schmid 2014; Crone 2021).

[68] He takes his analysis to be consistent with two further conditions of adequacy. The first condition states that an explanation of intentionality must be consistent with the fact, that all conscious states (including intentional states) must be located in individual minds, i.e. individual brains. Second, intentional states (individual or collective) must be fallible in the sense that "the structure of any individual's intentionality has to be independent of the fact of whether or not he is getting things right, whether or not he is radically mistaken about what is actually occurring" (Searle 1990, 406).

[69] Searle's own theory of collective intentionality is far from decisive and has been criticized on many different occasions. See especially: Meijers 2003; Andersson 2007. For *Special Issues* and anthologies see, e.g., Grewendorf & Meggle 2002; Koepsell & Moss 2003; Tsohatzidis 2007.

require cooperation but simply collective acceptance or recognition", which itself can be cashed out in term of such I-Intentionality plus mutual belief (ibid).

But if we are to accept this view, we come to face a problem: If the existence of certain institutional facts, including the existence of institutional roles, can be cashed out in term of individual "I-Intentionality" plus mutual belief, then collective acceptance is likely to be an *inadequate basis* for explaining the existence of institutional roles within large, and complexly structured institutional groups. Within such groups, we cannot simply assume that all individual members accept that every given role within the group has certain features, comes with certain tasks, or has certain deontic powers attached to it. More importantly, we also cannot simply assume that all the members have symmetrical mental states in the form of mutual belief (i.e., beliefs about the beliefs of others, etc.). After all, these groups can involve tens of thousands of individuals, who are unknown to each other.[70] So what we need is a theory of collective acceptance that provides us with a basis for institutional roles which does not rely on such a condition.

In the next sections, I will look at Ludwig's theory of we-intentions, and the role we-intentions play in the collective acceptance of institutional status roles. The summary of what I aim to show in these sections is this: Ludwig's *shared plan account* of collective intentions provides us with a theory of collective acceptance which can explain the existence of institutional roles without invoking the concept of mutual knowledge, or recursive belief-cascades. These requirements, Ludwig argues, are neither necessary nor sufficient for collective intentional *behavior* (cf. Ludwig 2017a, 221). This, in turn, saves the account from being inapplicable to large, and complexly structured institutional groups. Understood this way, collective acceptance can indeed provide an adequate basis for institutional roles, on which we then may proceed to climb up, and eventually knock away the intentional ladder.[71]

Kirk Ludwig's theory of shared intentions

Let us begin with Ludwig's account of shared intentions. Ludwig pleads for a distributive account of shared intentions, according to which shared intentions can be analyzed by explaining *we-intentions* of individuals. So we-intentions are intentions that individuals possess and shared intentions are we-intentions which are distributed among those individuals. We-intentions are not, in any sense, possessed by groups or supra-individual entities as such. The term we-intentions is only meant to pick out whatever sort of individual

[70] Shapiro (2014), e.g., sees it as a necessary condition of any explanation of large-scale cases for collective action (which he calls cases of "massively shared agency") that it pays tribute to the fact that members of a group are unknown to each other. Shapiro states: "When multitudes work together on a project, it is unlikely that any participant will know the identity of all the other participants. Most of the employees at Microsoft, for example, are unknown to each other. Because these workers don't know the identity of many of the other workers, the intentions of the latter cannot affect the intentions of the former. If Abel does not know that Baker works at Microsoft, Baker's plural intention cannot affect Abel's plural intention. Baker cannot, in other words, intend that the group of employees work together because he cannot settle the matter for them. *In cases of massively shared agency, therefore, interdependence, even of the weak kind* […]*, must fail*" (Shapiro 2014, 26) [own emphasis].

[71] I beg the reader to be patient here, as Ludwig establishes his arguments against the backdrop of Searle's initial theory. So I will only be able to draw out Ludwig's ideas by contrasting them to Searle's in the first place. Readers who are familiar with the theories of Searle and Ludwig may skip these sections and commence with 3.2.

intention is "distinctive of participating in collective intentional action" (Ludwig 2020b, 15). The question of the distributive account of shared intention is what makes these we-intentions *special*.

Ludwig here proposes a reductive account of we-intentions, which he calls the SHARED PLAN ACCOUNT. According to the account, what makes we-intentions special is their *content*. We will see below, what the SHARED PLAN ACCOUNT amounts to. But let us first look at Ludwig's description of individual intentions, as it offers a useful analogy to understand his concept of *collective* we-intentions.

Ludwig here ascribes to Davidson's causal theory of action, according to which actions can be explained in relation to *reasons*, i.e., beliefs and desires. Beliefs and desires mutually restrain each other in *practical reasoning* regarding the ranking of our preferences on the one hand, and the degrees of confidence about the likelihood of an outcome on the other hand. This sort of practical reasoning then results in the formation of *intentions* to do something, which, according to Ludwig, are a distinctive sort of pro-attitude whose causal–functional role is to bring about actions (cf. Ludwig 2017a, 44f).[72]

Suppose that I intend to end world hunger. What is required for me to *carry out my intention*? It cannot be simply the case that, by miracle or divine intervention, world hunger suddenly ends. In a case of divine interference, it would be false to say that I carried out my intention to end world hunger when world hunger ended while I was sitting at home and watching TV. In such a case, I did nothing to contribute for this event to come about. And events that occur while I did nothing to contribute to their coming about are not actions of mine.

So let us look at the conditions under which my intention is (or would be) satisfied. As just established, the mere fact that world hunger suddenly ends is not sufficient to satisfy my intention. Instead, my intention must refer to *my bringing it about* that world hunger ends and my intention is not satisfied if its end of resolving world hunger is achieved accidentally, or by divine command. Rather, as Ludwig concludes (2007, 368), my intentions needs to embody a *plan* and I need to do something which - in accordance with this plan - *brings it about* that world hunger ends. So the important thing about intentions is that Ludwig conceives of them as *commitment to a plan of action*, to change the world in accordance with one's pro-attitudes by undertaking specific actions (see: Ludwig 2017a, Ch. 4; see for an original planning-theory of intentions Bratman 1987).

What constitutes having a plan in the first place? Plans can be thought of as a series of actions, which are to be carried out in a particular order to bring about an event, or state of affairs. Also, plans can specify actions to bring about other, composite actions and they usually involve multiple steps to achieve a goal. If an agent has an intention to *F,* then she is committed to a plan, where such a plan can be complex or simple. Complex plans consist of multiple, intermediate steps, where those steps themselves may involve further plans that involve further steps, etc.[73] Ludwig calls this a series of "nested plans":

---

[72] If one forms the intention to do something in the future, this describes a *prior intention*, and if one one forms the intention to do something at the present time, the resulting intention is an *intention-in-action*.

[73] Primitive action can be planned as well. If I plan to move my index finger, and insofar as moving one's finger is a primitive action, I do not need to break down this plan into smaller sub-actions. So, while plans may or may not involve the breaking-down of actions into smaller steps, they can be best characterized as *specifications of actions* which are to be performed (Ludwig 2017a, 213)

> "Thus, plan A to F may involved steps B, C, and D, each of which involves some further steps B1–B4, C1–C3, and D1–D5, where each step itself has a particular end in view which contributes to the pursuit of the overarching goal to F. Thus, the plan A to F may be represented as A [B[B1, B2, B3, B4], C[C1,C2,C3], D[D1, D2, D3, D4, D5]]. Here the plan A to F represents the highest level of planning, the steps B, C, and D the next highest level, and so on. Steps that specify primitive actions are the lowest level of planning because these involve things we do but not by doing anything else. In a complex plan that involves steps, the plan is not merely to do the sequence of steps, as it were, independently of their being parts of the overarching plan, but to do them as parts of the overall plan, so that in, for example, executing B1 as part of the plan A to F, one executes it as the appropriate part of the overall plan" (Ludwig 2017a, 92).

Let us circle back to my intention to end world hunger. Ludwig arrives at a general rule [R] on how to represent the *content* of individual intentions with respect to acting in accordance with a plan:

> **[R]**: Where ‚*p*' is replaced by a sentence that does not have *a* as its subject a full rendering of ‚*a* intends that *p*' would be represented by ‚*a* intends that *a* bring it about in accordance with a plan he has as a result of so intending that *p*.' (Ludwig 2007, 369).

Applied to the case of my intention to end world hunger, this would translate to: "I intend that I bring it about in accordance with a plan that I have as a result of so intending, that world hunger ends". This action-plan that I have may include specifications of certain actions, which I need to undertake in order to realize, or bring about, my intended action. If I then instantiate that action plan successfully, i.e., if I bring it about, in accordance with the plan I have as a result of so intending, that world hunger ends, I end world hunger.
Now, let us move to his analysis of we-intentions. Assume that Bob and I intend to end world hunger. Understood loosely (or read "distributively"), this could mean that I intend to end world hunger and Bob also intends to end world hunger, independently from my intention to do so. But understood in a more substantial sense (or read "collectively"), this describes not a case of parallel individual intentions but of a case of collective, or shared intention: Bob and I intend to end world hunger *together*.

How should we understand our shared intention? According to Ludwig, what makes shared intentions special is that each agent involved has so called *we-intentions*.[74] We-intentions are intentions held (only) by individuals directed at the group doing a particular thing together (cf. Ludwig 2017b, 22). Leaving some further details (especially cases of mutual deception) aside (see Ludwig 2017a; 2017b, Ch. 2.7; 2007, Sec. IV; see for (critical) reviews: Gunnemyr 2017; Blomberg 2018; Bratman 2018; Miller 2021; Tuomela 2017), Ludwig arrives at an analysis of shared intentions of the following sort: We intend to *J* iff each *x* of us intends that we *J* in accordance with a shared plan (cf. Ludwig 2017b, 26). Considering the we-intention of individuals, this is Ludwig's account:

> **SHARED PLAN ACCOUNT:** *x* we-intends that we *J* if *x* intends that *x* contribute (in accordance with a plan *x* has at the time of acting) to there being a plan in accordance with which each of us makes our contribution to our *J*-ing (at the time of acting) (Ludwig 2020c, 82; see also: Ludwig 2017a, 289ff.).

So if *we* (which here means: Bob and I) intend to end world hunger, then both Bob and I *we-intend* that we contribute to there being a plan in accordance with which each of us makes our contribution to our ending world hunger. For these we-intentions to be satisfied, it is required that we intend to *implement a collective action plan* for ending world hunger (cf. Ludwig 2020b, 16). Because the content of my and Bob's intention must refer to *our bringing it about* that world hunger ends, both Bob's and my intention is not satisfied if its end of resolving world hunger is achieved accidentally or by divine command. But both Bob's and my bringing our collective action about is also constrained by the requirement that our contributory actions are in accordance with the plan that is so shared by Bob and I.
So how can plans be shared? And what does having the same plan come down to? What if, e.g., Bob and I have very different and non-compatible ideas, i.e. plan-conceptions on how to end world hunger? Do we then fail to act together intentionally? To this end, Ludwig states that

---

[74] As the account is developed in the light of plural action sentences, Ludwig's semantical analysis of shared intentions (SI) and we-intentions (wi) is the following:

[SI]

$[\Gamma_x]$(x we-intends to Φ with Γ iff
x intends that Γ Φs in accordance with a shared plan
iff

[WI]

(a) (∃*s*)(intention(*s*, *t**, *x*) and content(*s*, [s canonically brings it about that:
(b) (∃*e*)(∃*p*)(∃*t* ≥ *t**)(∃*f*)(primitive-agent(*f, t, x*) and brings-it-about(*f, e*) and
(c) [only *y* ∈ G(*p*)](∃*t'*)(∃*f'*)(primitive-agent(*f',t', y*) and brings-it-about(*f',e*) and (∃*p'*) (accords(*f', e, t', p', y*)) and
(d) accords(*f, e, t, p, x*) and
(e) becoming(*e*, that Γ Φs in accordance with a shared plan))]))
iff [fully expanded]
(a)–(d) and
(f) becoming(*e*, that (∃*e*)(∃*p*)$[\Gamma_x]$[∃*t*: *t* ≥ *t**](∃*f*)(primitive-agent(*f, t, x*) and R_Φ(*f, e*) and accords(*f, e, t, p, x*) and [only y ∈ Γ]
(∃*t'*)(∃*f'*) (primitive-agent(*f, t, y*) and R_Φ(*f'*, *e*)) and Φing(*e*))))])))) (Ludwig 2017a, 204).

> "all that is required above is that to have a we-intention one intend that members of the group will do the thing in question in accordance with a shared plan. […] The key to seeing why this [divergence of plan-conceptions] is not likely to be a problem is that the requirement is only that there be *a* plan that they share for carrying out what they intend, not that *every* plan any has for carrying out what they intend be shared. For any genuine intentional joint action, there will be one at a sufficiently *general* level of specification that they share, for example, that each should coordinate as necessary with the others in the light of information about how things are going to achieve their intended end" (Ludwig 2007, 372f.).

So crucially, according to Ludwig, a plan can be shared in a rather loose sense. For Ludwig, a shared plan can leave out "many things, and even most, for later development" (Ludwig 2007, 372). This way to think about shared plans allows for three characteristic features, which amend problems of the account being too restrictive:

1) Different participants may have different conceptions of how actions are to be carried out, i.e., they may have - to some extent - distinct plans and still act together intentionally. Ludwig explains that, just as in individual cases, shared plans can have a *margin of error*:

> "When I am trying to kill someone, I may aim at his head but shoot him in the torso. I did not intentionally shoot him in the torso, but I did intentionally shoot him, and I did intentionally kill him, even if not every detail of the plan to kill him by shooting occurred in accordance with the way I envisioned it" (Ludwig 2017a, 214).

Because plans allow for variation and thus a margin of error, they should be understood as *prototypical* or *canonical plan descriptions*, which can encompass a range of variations, or "ways that the end can be achieved that are close enough, for it to count as coming about in accordance with my general plan for doing it" (ibid). Anything close enough to the prototype then counts as falling under the canonical plan concept in question.[75]

2) Participants may not know more about the plan than simply what *their* part in it is. Ludwig gives the example of members of a sales team, which may be given their instructions by their regional sales manager

[75] Ludwig gives the example of a group of six coffin-bearers, who are split up into two sub-groups (The R-group of A, B and C; and the L-group of D, E, and F). The two sub-groups are to carry their side of the coffin from the hearse to the gravesite, but are given different instructions on how to take their positions doing so, resulting for them two have two different plans. The plan for the R-group says that the group carrying the right side are to take positions on the right from front to back in the order A, B, and C and their plan also says that the L-group are to take positions on the left from front to back in the order D, E, and F. The plan for the L-group, however, says that the L-group are to take positions on the left from front to back in the order E, F, D, while members of the R-group are to take positions on the right from front to back in the order B, C, A. Although the plans of the L- and R-group differ from another, when each sub-groups carries out *their* plan, the coffin is carried from the hearse to the gravesite. Ludwig says that the two plans of the sub-group are sufficiently overlapping, or matching in order for their actions to be collectively intentional: "[T]he plan that they have has a canonical specification but the plan is to do that or s*omething close enough* for the main aim. Then we can simply treat the case in which the L and R group in the example above are given different canonical specifications but where the extension of the prototype plan concept associated with each is coextensive" (Ludwig 2017a, 215) [own emphasis].

who wants to boost revenues. In this example, the individual members of the sales team "each of them, with the exception of the manager, knows only what *his own role* in the plan is" (Ludwig 2017a, 212) [own emphasis]. Now the requirement for a shared plan here simply amounts to the requirement

> "that one *may know enough about the plan to do one's part* but not know the details of the other parts. It is clear that we cannot make sense of the shared plan requirement unless the individual participants in it all have the same determinate plan. Still, having the same determinate plan need not imply that they all know all the details about it" (Ludwig 2017a, 214) [own emphasis].

So in the case of the sales team, "each knows his part in the regional sales team plan as laid out by the regional sales manager and knows that it is his part in the plan they are all executing laid out by the regional sales manager" (ibid). So, even if each member of the team cannot specify what all the other members contribute, "they are all on the same page about what the plan is" (ibid). As long as the members have a way of thinking about their own role in that plan, this can extend to plans in which "the participants don't even know who the other participants are or how many there are" (ibid).

3) Participants can intend to act together in accordance with a shared plan without having yet settled on a *specific* or detailed plan of action. Bob and I, for example, do not have to have a specific plan with all the minutiae and intermediary steps worked out. According to Ludwig, individuals can share a plan without having a detailed plan-specification on how to proceed, because such situations correspond to a general feature of plans, i.e, that they can be complex and consist out of multiple, intermediate steps. Those steps themselves, in turn, may involve plans that involve *further* steps etc. Complex plans allow for there to be flexibility of *specific* actions to be undertaken and plans can evolve and be refined through time. What is important for Ludwig is that a plan is shared at least on the *outset*, i.e., that individuals share a composite, *overall* plan to do something together, at least on a sufficiently general level of plan-specification. The more detailed specification of sub-plans comes subsequent to there being a general plan.

So again, according to Ludwig's account, a plan can be shared in a rather loose sense. What is important is that the individuals involved have a way of understanding *how to individually contribute to a plan* that they suppose is to be shared by everyone else. So while everybody has to have "way of thinking about the relevant group and the shared plan and know their own role in that plan" (Ludwig 2017a, 214), the account does not suppose that everyone has *exactly the same conception* of what they are doing, or *who is involved* or *what the specific steps* in the shared plan are (for further details see: Ludwig 2017a, Ch. 14.3). Plan concepts only need to specify a *prototypical or canonical plan description* and "anything close enough to the prototype counts as falling under the plan concept in question" (Ludwig 2017a, 214).

As a last point, it should be noted that individuals might have a weaker form of we-intentions, i.e., *conditional we-intentions* (see especially: Ludwig 2015b; 2017a, Ch. 5). Conditional intentions are "the upshot of contingency planning—that is, planning about what to do upon (finding out about) various contingencies obtaining that impinge on our interests" (Ludwig 2017a, 46f.). They can be expressed via: X intends to *A if C*. The conditions under which conditional intentions can be satisfied, in turn, can be expressed via: "X intends to make it the case that if C, then X *As*" (Ludwig 2017a, 60f.). As to conditional

shared intentions, Ludwig states that a group has a conditional intention to *A if* C when its members have conditional we-intentions directed toward their A-ing given that C obtains (ibid). So two individuals might we-intend to go jogging in the park *if it is sunny*, or have a beer *if the pub is open*. In those cases, individuals have conditional joint intentions, which is just a matter of each of them having conditional we-intentions directed at their doing something *on a certain contingency obtaining* (cf. Ludwig 2017b, 35). Those joint conditional intentions may likewise be *generalized*, that is, individuals can have what Ludwig calls *general policies* directed at joint action in certain circumstances (ibid). Generalized conditional we-intentions, in turn, are analyzed in terms of conventions (see below).

Having assembled Ludwig's theory of we-intentions, let us now see how they ultimately relate to status roles. To this end, we have to look at two further concepts: constitutive rules and collective acceptance.

Constitutive rules

The canonical way of describing constitutive rules starts by distinguishing them from another set of rules, namely *regulative* rules. According to Searle, the latter "regulate antecedently existing activities" (Searle 1995, 27; further see: Searle 1964; 1969, 32ff.) whereas the former make certain types of activities possible in the first place.[76] An example of a regulative rule is the rule of driving on the right side of the road, where this rule regulates an activity (driving) that existed prior to positing this rule. Searle's example of constitutive rules are the rules of chess: without these rules (e.g., that the rook may move to any square along the file or the rank on which it stands), there would be no game of chess, as the game of chess is - among other things - constituted by its rules.[77]

Searle's formula of constitutive rules is that of "X counts as Y in C", where the X-term describes some already existing entity (e.g., an object, activity, person or state of affairs) which provides the basis for the Y-term. In a constitutive rule, the Y-term assigns a "status function" to the object functioning as the X-term. Importantly, the X-term itself, e.g. in virtue of its physical properties, cannot perform this function prior to

[76] The distinction of regulative and constitutive rules can be traced back to John Rawls (1955), who distinguished between what he calls the "summary view" of rules on the one hand, which describe rules that (in certain cases and within certain circumstances) are followed because of their utility, e.g., so called "rules of thumb". Those summary-rules can be conceptualized as generalizations that are being inferred from decisions in the past. On the other hand, then, there is his conception of rules as a *practice*. Such rules of practice, Rawls states, are "stage-setting" in the sense that one cannot perform certain types of practices without these rules being established in the first place. The rules of practices are *logically prior* to particular cases of this practice, which means, that "given any rule which specifies a form of action (a move), a particular action which would be taken as falling under this rule given that there is the practice would not be described as that sort of action unless there was the practice" (Rawls 1955, 25). So rules of practice define certain types of actions logically prior, i.e., *before* they are instantiated in specific cases, or to put it in Searlean terms: they are *constitutive* of the practice that they govern. A philosophical precursor to the concept of constitutive rules has been developed by Anscombe (1958) who distinguished "brute" from "institutional" facts. See also: Hart (1994) for the related distinction between "*primary*" and "*secondary rules*".

[77] The distinction between these two types of rules, however, has been thoroughly contested. Giddens (1984) pointed out that constitutive rules can also regulate the behavior and action of those who follow them, and that regulative rules can have constitutive aspects or, when taken together, indeed constitute a certain practice, such as the rules of etiquette. Hindriks (2009) argues for an extension of the distinction in order to introduce a third type of rule, i.e., "status rules", whereas Guala (2016) posits the possibility to ultimately reduce constitutive rules to regulatory ones, leading the distinction to collapse altogether.

being assigned the Y-status. The "C"-notion in the formula describes the context, i.e. the properties and conditions that must be fulfilled to successfully ascribe the status function to the X-term.
Now Ludwig advances on Searle's definition of constitutive rules. For Ludwig, Searle's formula of constitutive rules is about stating *conditions* under which something is to be called, or falls under the concept of, an *intentional action type.* However, Ludwig says that "stating the conditions under which something is to be called one of these things is not to state a constitutive rule" (Ludwig 2017b, 102). According to Ludwig, Searle's formula of "X counts as Y in C" just does not have the right form for a rule "the intentional following of which brings into existence a type of activity" (ibid). He considers Searle's example that under certain circumstances (C), crossing a line with a ball (X) counts as scoring a touchdown (Y):

> "Clearly this does not provide a rule one can follow intentionally with others which contributes to constituting a play of the game of football. It is not about what anyone who is a member of either team is to do in order to play football. It rather specifies a definition for a term that itself will figure in a description of what it is to play football" (Ludwig 2017b, 102).

Rather than linking constitutive rules to conditions under which something is to be called, or falls under the concept of, an intentional action type, Ludwig focusses on the *agency-enabling* dimension of constitutive rules. For Ludwig, constitutive rules enable certain *activity types or activity patterns*. Constitutive rules then are *only* constitutive relative to a *particular activity type*. A constitutive rule,

> "is not a rule of a special type or form, or a rule with a special content. Nothing intrinsic to a constitutive rule makes it a constitutive rule. Constitutive rules are constitutive *relative to* an action type because the concept of the type of action determines that any action that comes about by way of the rules being followed intentionally (or close enough) is of that type. It follows that for any set of rules describing a pattern of activity, there is an action type relative to which they are constitutive rules. It is the content of an action type concept that determines that following rules intentionally is constitutive of that type of activity" (Ludwig 2017b, 93f).

A way to test whether a rule is constitutive of an activity or not, is to ask whether there is a *distinctive pattern* associated with this intentional activity or not. Not every intentional action requires there to be rules that govern such activity: Waiting for the bus can be described as an intentional action, however, there is no distinctive pattern of activity one must engage in, in order to be waiting for the bus. Now some of the patterns of activity governed by constitutive rules concern the actions of individuals (e.g., playing solitaire). And some are collective in a weak sense because they can be performed *either* by an individual *or* together (e.g., singing, or playing the piano). But some types of activity are *essentially* collective: Dancing a *Pas de deux*, playing hide and seek, or chess-boxing, for example, are *essentially* collective intentional activities, i.e. activities that could - in principle - not be performed by just one individual. And just as patterns of individual intentional activity require individual, or I-intentions directed at the activity type, when such patterns of activity are collective, the participants in those actions must have we-intentions of the above mentioned form directed at the activity type.

The upshot is that constitutive rules can enable *essentially intentional collective action types*, i.e. activities that can only be performed if several people jointly intentional follow a pattern of activity, like orchestral performances or playing a game of baseball. When turning to collective acceptance, we will see how exactly we-intentions figure in the functioning of rule-governed activity types and how the collective acceptance of a rule is connected to individuals having we-intentions.

Status functions and collective acceptance

Status functions are functions that objects or individuals have which are governed by constitutive rules. As mentioned above, Ludwig draws on the work of Searle in oder to modify his concept of status functions and to develop his concept of *status roles.*

Searle posited that functions in general are intentionality-relative phenomena, i.e., they are assigned to entities relative to the interest of users and observers. So you can figure out the function of an entity, e.g., a remote control, by asking what purpose the entity is meant to serve (i.e., its design-based function) or actually serves, i.e., (its use-based function). Functions which are defined in relation to human purposes can be called *agentive functions*. According to Searle, all functions are agentive functions (see: Searle 1995). Now some entities, e.g., a two-levered draisine, possess agentive functions which they can only perform in the context of *joint* actions. These functions that an object can perform only in virtue of joint action are *joint agentive functions*.

Status functions are a subclass of joint agentive functions (see: Ludwig 2017b, Ch. 8.2). What makes status functions special, and distinguishes them from functions of tools like the draisine, is that their existence rests not on their physical properties and/or design. Rather status functions are (implicitly or explicitly) defined by constitutive rules for joint action types, which are *collectively accepted* by the relevant individuals involved. Status functions are derived from constitutive rules of intentional action types that multiple people aim to realize:

> "This is where both the idea of an assignment of a function to an object or type of object and of collective acceptance comes in. Something is recruited to play a role in a social interaction by members of a group, which is for it to be assigned that role in a collective activity they anticipate, and the mode of recruitment is collective acceptance of its serving that role" (Ludwig 2017b, 116).

So to explain status functions, Ludwig draws on Searle's idea that they require collective acceptance. Crucially, Ludwig's modified account of collective acceptance does not face the problems of being applicable to institutional groups which I sketched above. Ludwig and Searle both agree that it is a necessary requirement for status functions (and consequently, institutional facts) that the members of a relevant community have certain attitudes toward it. But on Ludwig's account, these attitudes are not collective *representations* in Searle's above mentioned sense, but rather *we-intentions* or *conditional we-intentions*.

Consider the example of the game of tic-tac-toe. What makes a given three by three grid a board of tic-tac-toe? Of course, not just any three by three grid automatically counts as a board of tic-tac-toe. For a three by three grid to have the *status function of being a board of tic-tac-toe*, two individuals have to decide, at least for the time and space of the game, that this grid will count as such a board (cf. Ludwig 2017b, 130f). This particular grid therefore has the status function of being a board of tic-tac-toe by being collectively accepted to count as such a grid. But what does this collective acceptance of the players in such a situation amount to? According to Ludwig,

> "it is their adoption of *certain commitments with respect to the token [grid] in question*, commitments with respect to how to behave in the anticipated play of the game with respect to it. That at once constitutes its having a certain social status in the group constituted by the players and its both having and being able to perform the function it does in a certain social transaction, the play of the game. Since the *commitments are commitments with respect to how to act, however, they are intentions.* Thus, the crucial thing is that they should be committed to treating it, i.e., intend to treat it, as having the relevant function relative to their play of a game of tic-tac-toe. *Since these intentions are directed at their doing something together in accordance with a common plan, they are we-intentions*. Their we-intentions coordinate on the same things for roles specified in the action type they aim to instantiate" (Ludwig 2017b, 132) [own emphasis].

Finally, this turns on Searle's assumption that for collective acceptance to occur, there must be mutual belief involved:

> "Social facts are grounded in the conditional we-intentions of members of a community, and while these may and will typically be associated with beliefs about the corresponding conditional we-intentions of others and so beliefs about the existence of objects or types with the relevant status functions, they are concomitants to what constitutes the social facts, and conceptually neither necessary nor sufficient for them" (Ludwig 2017b, 184).

Ludwig's central claim is that for objects to be assigned status functions, this amounts to enough members of the community being prepared to *use* these objects (or treat these persons) as if they had those functions when they engage in the relevant types of activities (cf. Ludwig 2017b, 133f).
However, the collective acceptance of status functions does not necessarily specify, *which* objects should be assigned the function. Therefore, those individuals are faced with a *coordination problem* (cf. Ludwig 2017b, 117f). In chess, for example, there are many things to which the individuals could assign the pawn-status to (walnuts, peanuts, hazelnuts, etc.). According to Ludwig, individuals solve these coordination problems by establishing certain *policies* to use the same objects or types of objects on repeated occasions. He then

argues for such policies to be *conventions.*[78] Ludwig argues for the idea that a group adopts a convention when its members have accepted a solution to a coordination problem that leads them to *act jointly intentionally* in accordance with the agreed upon policy:

> "for a thing to play the relevant role […] it must be *used by all the participants in the action involving it intentionally in that role*. That falls out in turn of its being implicitly defined by constitutive rules understood as governing an essentially intentional collective action type. This explains the source of the conventional in connection with status functions, from the treatment of certain bits of paper or metal as money and certain objects as chess pieces, to the treatment of certain rituals as conferring on a pair of individuals the status of being married to one another, or on a particular individual, say, the role of being the President of the United States" (Ludwig 2017b, 130) [own emphasis].

What we've now assembled is an account of status functions under which status functions are essentially intentional collective action types governed by we-intentions. And some of these status functions, i.e., *status roles*, are not assigned to objects, but to individual agents. Again, it is worth emphasizing that Ludwig's *shared plan account* can explain the existence of institutional roles without invoking the concept of mutual knowledge, or recursive belief-cascades. These requirements are neither necessary nor sufficient for collective intentional action, which he bases on his theory of we-intentions (cf. Ludwig 2017a, 221; 2017b, 132).

3.2. Role-based Institutional Action

With the concept of institutional roles being established, we can start to investigate how they figure in explanations of institutional agency. How exactly do roles provide an adequate basis for explaining the *agency* of institutional groups?

The main goal of this section will be defending the claim in (I), according to which institutional group agency can sufficiently be explained by *individuals acting qua institutional roles*. The main tenet of a role-based explanation of institutional agency then is this: An institutional group action consists of (and can be reduced to) the individual contributory actions of its members, who act according to their assigned roles. Clarifying

[78] The standard analysis of conventions is given by Lewis (1969), according to which conventions are solutions to a coordination problem, which have the features of being *stable*, *arbitrary*, *reciprocal* and *social* (see Ludwig & Jankovic 2022). In contrast to Lewis, Ludwig relies on the "Collective Acceptance Account" of conventions (see for further elaboration: Ludwig 2017b, Ch. 9). This account, according to Ludwig, diverges from Lewis' analysis in certain respects: "The most important difference between Lewis's account and the Collective Acceptance Account is that on the latter a convention involves constitutively the group that faces a coordination problem *collectively accepting* a solution to it. This implies that when people follow a convention, they do so as a group intentionally, because collective acceptance of a solution to a coordination problem implies that when the members of the group coordinate in the relevant way, they are doing so on the basis of that prior joint commitment. On the Collective Acceptance Account, then, convention involves collective intentionality essentially. In contrast, this is not a requirement of Lewis's analysis" (Ludwig 2017b, 125).

this will require first to keep track of, and ultimately overcome certain problems that reductive explanations of institutional group agency encounter.

The first problem, let me call it the *Problem of Action Integration*, stems out of the seemingly complex relation between an individual role-occupant's contributory actions and the group-level actions of institutional groups. Institutional actions can exhibit varying degrees of complexity. They can reach from a spokesperson making a single announcement on behalf of the company she's working for, to military armies fighting battles through the integration of the various branches of their forces into a single, unified command; with different combat arms and sub-units achieving mutually complementary effects at the same time at different places. A problem here, on the one hand, is to explain how seemingly small and unimportant individual actions could be said to be integrated into these complex actions. How do such individual actions relate to the actions performed on the group-level?

On the other hand, institutional groups seem capable of performing certain actions without every member being involved in bringing about those actions. But if that is the case, how could we make sense of reductionist explanations, according to which group actions can be reduced to the actions of their members? Recall that, according to Ludwig, to say that an institutional group did something, is to say that there was an event of which *each* member of the group, and only those members, were the direct agents of. The phenomenon of institutional groups being capable of performing certain actions, without every member being involved in bringing about those actions, seems to provide a direct counter-example to such a claim. Here, the question is how single individual actions can be said to constitute the actions of the whole group, i.e., how group actions can be performed through only one member.

Second, and related, is the problem of *Diachronic Group Constitution*. This problem can be best exemplified by looking at the diachronic endurance of group action: institutional actions exhibit diachronic robustness and can seem to be quite long-lasting. Building the first atomic bomb, for example, was an institutional action that took multiple years to be carried out. And during these years, not all of the initial members of the Manhattan Project were involved throughout the entire time, with some being only involved at the initial stages of the project, while others joined only at the very end. The relation between institutional groups and their members therefore seems to be rather loose. Now if a group consists out of nothing more than its members, and those members changed throughout the performance of an institutional action: Why should we say that the *group* persisted, i.e., that it remained identical throughout the entire time? And why should we speak of the *group's* action and not only of the actions of individuals that constituted this group at any particular time?

I will now argue that role-based explanations can solve these problems by highlighting critical features of institutional roles in relation to institutional agency. I will examine each feature and consequently explain how complexly structured and potentially large institutional groups can perform actions through their members performing the tasks and functions of their institutional roles.

Acting *qua assigned role* or *as a role-occupant*

A core claim of role-based accounts of institutional actions is that they are reducible to forms of individual actions which are related to one another. Properly speaking, "there are no such things as corporate actions"

(Miller 2001, 53), but only individual actions related to one another in a certain way. If we take this at face value, then an analysis of institutional group action, cashed out in terms of collective action, could be given in the following way:

> **(G)**: a group of which *A, B,* and *C* happen to be the sole members performs a joint action *j* just in case *A, B,* and *C* jointly perform *j* (Schmitt 2003, 148).

Let us say that the institutional group in question is a *party committee* consisting out of three individuals I_1, I_2 and I_3. Their collective action *j* is to throw a party, and the individual actions taken to realize this joint partying consist of I_1's *x*-ing, which is decorating the room, I_2's *y*-ing, which is mixing the drinks and I_3's *z*-ing of playing the DJ. Let us assume that taken together, these actions constitute the party committee throwing a party at *t*. If we follow a role-based way of explaining institutional agency and think of institutional action to be constituted by (and to be nothing more than) individual actions of group members, we come to face a problem. For we need a way to tell apart the actions of these individual group members acting *for the group* from the actions of these individual group members not acting this way.

Here, one problem for individualistic accounts of collective action is that of *co-extensive* membership, which is also called the *Counting Problem* (see Gilbert 1987 for an early discussion of the problem; also see Ludwig 2017b, Ch.11.3; Schmitt 2003). The Counting Problem tackles the idea that institutional action can sufficiently be reduced to contributory actions of its individual members.

To see why, imagine that, while I_1, I_2 and I_3 may be the only members of the party-committee, they also are the sole members of the *bereavement committee for recently deceased family pets*. If we are to think of the actions of the bereavement-committee to be constituted by (and to be nothing more than) the individual actions of I_1, I_2 and I_3, then it would seem that by I_1's *x*-ing, I_2's *y*-ing and I_3's *z*-ing at *t*, both the party- and the bereavement-committee threw an (inappropriately cheerful) party. So because distinct groups can have the same individual members, there must be a criterion for the actions of those groups to be distinguishable. There seems to be some sort of context missing to sufficiently tell which group acted when the party took place. But how could we specify such a context? To this end, Miller offers a useful insight by his notion of individuals acting *qua assigned role:*

> "The notion of acting *qua* occupant of a role is simply that of performing the tasks definitive of the role (including the joint tasks), conforming to the conventions and regulations that constrain the tasks to be undertaken, and pursuing the purposes or ends of the role (including the collective ends). [This] can be supplemented by recourse to concepts of conventions, social norms, and the like, and especially by recourse to the explicitly normative notions of rights, obligations, and duties that are attached to, and in part definitive of, many organizational roles […] It is not simply that organizational role occupants *regularly* jointly act in certain ways in preference to others, or in preference to acting entirely individualistically; rather, they have institutional duties to so act and – in the case of hierarchical organizations – institutional rights to instruct others to act in certain ways" (Miller 2010, 54).

Miller's quote then can be read as summary of the core claim of (II). An individual acts *as a role-occupant,* or *qua occupying a role* (or *qua role-occupancy*) if she performs the tasks definitive of her role. These tasks and functions, in turn, are based on the exercise of certain, definite deontic powers.
For Miller, an interesting feature of roles here is that they go along with certain mental attitudes that the individual must adopt or realize. Miller states that these attitudes can either be *associated* with certain roles, or they can be *constitutive* of these roles:

> "I say these attitudes are associated with the roles rather than constitutive of them since it is possible to occupy the role without adopting the attitudes. Of course, there are some beliefs and intentions that are constitutive of the role, but these are simply those beliefs and intentions that are constitutive of the actions one performs in occupying the role" (Miller 2001, 172).

Similarly, Ludwig states that institutional roles come with what he calls "*role-based reasons*" for action, which describe reasons individuals may posses to perform certain actions in virtue of them occupying a status role. He defines these role-based reason to perform some action *A* in a context *C* as follows (Ludwig 2017b, 149):

> For any *x, x* has a role-based reason to *A* in *C* at *t*
> if and only if *x* has a status role at *t* which obligates *x* to *A* in *C*
> if and only if it is part of the design specification of a status role *x* has at *t* that *x A* in *C*

With this notion of acting *qua assigned role*, or *for a role-based reason*, we can begin to see how institutional actions can be differentiated from one another in the case of the set of members of two institutional groups being co-extensive. It is *not* the case, that I_1's *x*-ing (decorating the room), I_2's *y*-ing (mixing the drinks) and I_3's *z*-ing (playing the DJ) constitute the action of the *bereavement-committee*, because the individuals are not acting in order to perform the tasks and functions definitive of their roles in the bereavement-committee. Also, they are not pursuing the purposes or ends because of the role-based reasons for acting in such a way. In return, in order for I_1's action to be attributable to the bereavement-committee, I_1 has to act *qua* her role in the committee, which consist of performing the tasks definitive of the bereavement-role, obeying its duties, exercising its rights, pursuing its ends, etc.
Another cogent example of *acting qua role-occupancy* is given by Michael Schmitz (2017), who argues not only for understanding role-specific attitudes, reasons but also for corresponding *forms* of reasoning as to be linked to a so called "role-mode" (to which I will turn back later in Chapter 5.1.). The canonical representation of individuals acting in the role-mode, i.e., on role-specific attitudes and reasons is expressed through locutions like "As [role]", "In my role as [role]", or "qua the power invested in me as [role]" (cf. Schmitz 2017, 62). Schmitz gives an example of an individual having role-based reasons for action, or acting on role-specific attitudes *as a policeman*:

> "The crucial point for present purposes is that because attitudes are in some cases role-specific, so are reasons and the corresponding forms of reasoning. That somebody is smoking a joint may be a reason for the policeman to arrest him, though as a private person the bearer of this role may have no objection to it. So the policeman may reason deductively from his belief *as a policeman* that a certain man has smoked a joint and his (let us assume) general obligation *as a policeman* to arrest people who do such things, to the particular obligation to arrest this man. It is necessary that this belief be one that the man holds as a policeman because if, for example, his personal belief was based on inadmissible evidence – say, obtained through illegal wiretapping – it could not provide a legally valid reason to arrest the man even if it was true" (Schmitz 2017, 62).

The general point of Schmitz is that institutional roles not only provide role-specific attitudes and reasons but also role-specific *forms* of reasoning, i.e., they shape an individual's perspective on her capacity for action. Schmitz describes this as institutional roles providing "vantage points" on the world "that can differ from our merely personal, I-mode ones, both with regard to our practical and to our theoretical attitudes" (ibid).

One of the merits of such role-specific attitudes and reasons can be fleshed out in comparison to the *group agent* explanation laid out in the previous chapter. Recall here the theory of List & Pettit. According to List & Pettit, the *doctrinal paradox* (List & Pettit 2011, Ch. 2; see also Pettit 2003) seemingly showed that groups can be genuine agents and have "psychological autonomy" (Pettit 2003, 167) which stems out of what List & Pettit call the "imposition of discipline of reason at the collective level" (Pettit 2003, 175), or the above mentioned *collectivization of reason* in the sense of aggregating intentional attitudes of group members into a "single system of such attitudes held by the group as a whole" (List & Pettit 2011, 58ff.).

I already argued above that this rests on a mischaracterization of group belief, individual belief and the relation between those two by pointing to Gilbert's (1987) distinction between the general notion of Individuals *believing that p* and an individual *personally* believing that p, where every instance of an individual *personally* believing that p is an instance of an individual believing that p, but not the other way around. By invoking the concepts of acting *qua role-occupancy*, and having the related *role-based attitudes and reasons*, we can now make further sense of Gilbert's claim that individual group members expressing the belief of a group view themselves as speaking "in their capacity as group members," or "as a member of this body," etc. We can do so by pointing to these individuals occupying institutional roles. Acting *qua role-occupant* can encompass having corresponding, or associated attitudes and role-based reasons for expressing attitudes or externally acting on such role-based reasons. For Schmitz (2017), it also encompasses role-specific *forms of reasoning*. In a next step, we can then explain away the apparent "psychological autonomy" of List & Pettit's group agent by understanding the aggregated intentional attitudes of group members not as giving rise to a "single system of such attitudes held by the group as a whole" (List & Pettit 2011, 58ff.), but by being attitudes of individual group members that they have *qua role-occupancy,* or *qua reasoning as role-occupants*.

The upshot is that occupying an institutional role can both drive a wedge between an individual's personal, i.e., non role-based attitudes and actions, and her role-based attitudes and actions. We *also* can explain (in

a reductionist way) just *how* this wedge can explain the attitudes held on a group-level being apparently "autonomous" or "holistically supervening". It is not the case that the "attitudes" of the group differ from that of the individuals' and thereby are somehow autonomous, or metaphysically distinct. Rather, the attitudes of individuals can differ *according to whether they occupy a role or not*. Hence, no non-reductive group agent is necessary to explain the divergence between group-level attitudes and individual attitudes once we invoke the concept of institutional roles.

Specialization

Let us continue with the feature of the role-specific *specialization of tasks and functions*. The fundamental idea here is to connect the concept of institutional roles with the division of (manual or cognitive) tasks along the unequal distribution of member-to-action relations of power, which characterizes institutional group action. To gain clarity on the complex forms of specialization occurring in institutional groups, we might begin with a rather simple example:

Miller asks us to imagine a group of three individuals who want to achieve a collective end.[79] In order to achieve this end, each individual (A, B and C) have to each fulfill a special, non-general task (x, y and z), which are individually necessary and jointly sufficient to bring about the collective end. Then, Miller asks us to assume that:

> "*A* cannot perform *y* or *z* (or at least cannot perform *y* or *z* without difficulty); *B* cannot perform *x* or *z;* and *C* cannot perform *x* or *y.* Assume finally that if *A* dies or leaves, *B* and *C* will identify some *D* to replace *A;* similarly if *B* or *C* dies or leaves, then some *E* or *F* will be found as a replacement" (Miller 2001, 29).

Miller concludes, that this constitutes an - albeit primitive - institutional group (or as he calls it: organization) engaged in *specialized* action. Let me home in on this by connecting the *specialization* of tasks to the above mentioned organizational structure, that roles give rise to.

Recall that institutional groups are defined as embodied structures of roles, which, in turn, are defined in terms of tasks that an agent is to perform, and rules, procedures and/or conventions that regulate how an agent is to perform these tasks (cf. Miller 2001; 2019 Sec. 1; 2010, 47). Roles and the structures in which they are embedded "dictate" (Miller 2001, 172) certain actions which are to be performed for the purpose of the collective end. For some institutional actions, the actions performed by individual members might all be the same, generic *type* of action. However, the actions performed by individual agents do not necessarily have to be of the same type and each agent might have to perform a *different type of action* in order to fulfill the defined tasks of their roles and to contribute to the realization of a collective end (cf. Miller, 2001, 62f.).

So while the central idea was that an organizational role can be defined via tasks, these tasks can also be defined by *specific* types of actions these tasks are supposed to realize.

[79] A further analysis the notion of "collective ends" and Miller's Collective End Theory (CET) of Joint Action will be provided in Ch. 3.3.

Because Miller defines roles as "simply an abstraction from procedure-governed tasks, and the actions that constitute the undertaking of those tasks", to say that "the role is filled" is just another way to say "that some individual is undertaking that task or tasks, and following those procedures" (Miller 2001, 172f). The somewhat vague notion of "*undertaking" a task* here is meant to explicate that individuals perform "a relatively complex set of differentiated actions directed at some given end" (ibid). So while institutional group actions can be complex, (think, e.g., of a company that produces cars and all the intermediary steps and procedures that it takes to realize this) the differentiation of roles encompassing the specialization of tasks can explain not only how institutional actions can be "split up", or segmented into smaller sub-tasks, but also how the complexity of action on the institutional level can thereby be reduced to simpler actions on the member-level.

As with the above mentioned accounts of Tuomela (2013) and Kutz (2001), the role-based specialization of tasks and functions leads to a differentiation in the member-to-action relation of power. Similarly to Tuomela's positional theory, a role-based explanation of institutional group action accounts for such uneven distributions of member-to-action relations of power in virtue of basing them on the uneven distribution of member-to-member (MtM) relations of power. *Pace Tuomela*, these MtM-relations do not reside in the unitary divide between operative and nonoperative members, but in the structural interrelation between roles. So, e.g., role R_1 might require the individual group member m_1 to *x* while role R_2 might require the individual group member m_2 to *y* etc. And as with the account of Tuomela, those relations of power can be *iterated horizontally* as well as *vertically*. The vertical iteration of role-based task-specifications is especially important for the establishment of a *chain of command*, which allows to include non-egalitarian groups in the analysis of institutional group agency. The horizontal iteration of hierarchy institutional roles can also be seen as *task-*, or *domain*-relative insofar as certain roles require the occupants to carry out the performance of *certain* sub-actions (e.g., sub-action *x*) but not the performance of certain *other* sub-actions (e.g., sub-action *y*). Taken together, the horizontal and vertical dimensions of hierarchy, based on the role-based specification of tasks and functions can explain how institutional groups can be complexly structured and consequently account for the division of (cognitive and/or manual) tasks.

Now, if the specialized actions of individuals can vary vastly from one another, why are they still the actions that constitute a group-level action? Whereas the initial question was under which circumstances the actions of individuals could be said to belong to, or be constitutive of a group action, the question now arises how group actions, qua being constituted out of individual actions, could be *individuated*. So, if a joint action *j* is performed just in case *A, B,* and *C* jointly perform *j as group members*, how can multiple, specialized individual actions of A, B and C constitute *one and the same group action*? What *unites* those actions?

Layered structures

The ways in which we ordinarily talk about institutional group action often seems to suggest that these actions are the single, unitary actions of such institutional groups. Also, talk about institutional action often comes in the form of actions that institutional groups can perform, but that individuals cannot. Not simply because they are too complex, but because they are conceptually impossible to be performed by

individuals. Whereas, e.g., the *merging* between institutional groups is a common thing to occur, individuals have yet failed to do so. Military units can be *disbanded*, venture capital funds can *dissolute*, and companies may *launch at the stock markets*. Individual (human) agents can do none of these things. This apparent exclusivity of institutional actions can lead to the impression that they are performed by a single, ontologically distinct group agent: Why *should* we think of institutional actions to be nothing more than individual actions, if individuals cannot perform these types of action? I will argue that this problem does not compel us to require an explanation of institutional group agency which includes an ontologically distinct group-agent.

In order provide for a reductive explanation of institutional group actions, we must point to the *layered structure* of institutional actions. The basic idea is to think of institutional (or as Miller calls them: corporate) actions to be constituted by sub-, or component-actions, which are embedded, or functionally integrated into a *layered structure*. This is the way in which individual actions are integrated into, and thereby constitutive of an institutional, group-level action. In turn, it allows for the complex, seemingly independent actions of an institutional group to bottom out in the specialized component-actions of its individual members. Strictly speaking, there then is no such thing as a group action *per se*. So talk about the unity of institutional action is a *semantical* problem, that a move beyond the surface of our grammar can resolve. To see how this can be further analyzed, we might first consider the ways in which actions in general can be individuated.

So let us start with the question of how to individuate actions in the first place.[80] Think, e.g., of small, every-day actions like deleting your spam-mail. Even such seemingly mundane actions can be described differently, i.e., *coarse*- or *fine*-grained. Your deleting the mail can be described as one single action, but it also can be described more precisely: opening your mail-program, clicking on the new mail, briefly reading the first sentence, your right index-finger clicking on the mail-icon and dragging it towards the spam-file, etc. What initially was described as a rather mundane, single action can in fact be broken down into (infinitesimally) smaller parts. But now the question arises as to which action (or actions) did take place. Did you delete your spam mail *and* did you open your mail-program *and* did you click on the new mail, briefly reading the first sentence, etc.? Or, if you did not do anything *more* than open your mail-program, click on

[80] In classical (individual) action theory, there are broadly two camps: one is to allow for coarse-grained individuation of actions and the other one is to advocate to fine-grained individuation of actions. Both theories are endorsed and both theories accuse the other of being implausible or inconsistent. See for coarse-grained individuation e.g., Davidson 2001h; for fine-grained individuation e.g., Alvin Goldman's A Theory of Human Action (1970); or McCann (1983). For an intermediate position e.g., Ginet 1990. Paul (2020, 47f.) argues for the irrelevance of this debate altogether. There are downsides to both approaches: According to the fine-grained individuation of action, each of the above described smaller parts are to be conceived of as actions themselves; clicking, dragging, moving one's fingers etc. each describe one distinct action that - taken together - constitute the deleting of the mail. Coarse-grained individuation postulates that while one may executed all of the above things, one only performed a single action, i.e. deleting your mail. The point proponents of coarse-grained individuation like Davidson make, is that once one has subscribed to fine-grained individuation, one also need to have a criterion for when to stop the regress into splitting up actions into more fine-grained actions, and splitting these actions into even more fine-grained actions, etc. If, however one subscribes to a coarse-grained individuation, one might run into problems of describing when exactly an action started and when exactly it might have ended (See for discussion: Ludwig 2010, 44f.; Piñeros Glasscock & Tenenbaum 2023, Sec. 5.2., Bratman 2006).

the new mail, briefly read the first sentence etc., then maybe there was no such action of you as to delete the spam mail *as such*?

Here, Davidson's work on the individuation of action (2001e; 2001h) draws on the helpful distinction between action *types,* i.e. sorts of actions and action *tokens*, i.e., instances of actions. Action types, e.g., *greeting someone* can be realized in different ways, i.e. by different token-actions (e.g., by waiving one's arm, by saying "Mahlzeit!", or by nodding at someone). So first, individuating actions can then be done by individuating action *types*. In a next move, we can distinguish *complex* action types from *basic* action types. Complex action types can be defined as the summation of *basic* action types (or in Davidson's terms: *primitive action types*). Take the example of building a house as a complex action type that *is* the sum of all the more primitive action types that constitute the complex action: framing walls, hammering nails, and laying shingles, etc.[81] Complex actions, understood as actions that *are brought about by doing something else*, can therefore be further decomposed into primitive actions, which can be understood as actions that are brought about *without bringing about something else* (see Davidson 2001e). Building a house then is not a primitive action. It is brought about *by* hammering, framing, laying shingles etc. Rather, our talk about action in these cases is *elliptical*: To say that an agent caused some event *Y* is to say that the agent caused *Y by causing some other event X* which caused *Y.*[82]

Now this idea of distinguishing complex from basic actions is picked up by Miller in his description of a *layered-structure* of institutional action. Miller asks us to consider the example of a military unit fighting a battle. Here, a number of joint actions of each specialized sub-group of individuals is severally necessary and jointly sufficient to realize the collective end of this military unit, e.g., taking a strategic hill:

> "Thus the ‚action' of the mortar squad destroying enemy gun emplacements, the ‚action' of the flight of military planes providing air cover, and the ‚action' of the infantry platoon taking and holding the ground might be severally necessary and jointly sufficient to achieve the collective end of defeating the enemy; as such, these ‚actions' taken together constitute a joint action. Call each of these ‚actions' level-two ‚actions,' and the joint action that they constitute a level-two joint action" (Miller 2010, 48).

Now the actions of these specialized sub-groups can further be reduced to the severally necessary and jointly sufficient actions of the individual members of those subgroups:

---

[81] This way of conceptualizing actions is well established, but by no means uncontested. Scholars argue whether this view of reducing complex action types to primitive ones is the right way to describe actions or whether such primitive actions exist in the first place (see for discussion: Lavin 2013).

[82] If this is the right way to look at actions, what then are basic or primitive action types? Davidson famously argues that for humans, the most fundamental form of action consist out of - generously interpreted - bodily (and/or mental) changes (cf. Davidson 2001e, 47ff.; 57). Such bodily changes human beings can bring about primitively include movements of limbs and various other changes in our bodies brought about by movement of one's muscles. All we can ever really do, Davidson famously claimed, is move our own body, the rest being up to nature (ibid). Like almost everything else in this debate, this is contested too. Especially the characterization of basic actions as changes of *mental states* ultimately depends on which theory of mental states one endorses. If one thinks of mental states as being identical to states of the body, then one could just say: primitive action types consist out of bodily changes.

> "Thus the individual members of the mortar squad jointly operate the mortar to realize the collective end of destroying enemy gun emplacements. Each pilot, jointly with the other pilots, strafes enemy soldiers to realize the collective end of providing air cover for their advancing foot soldiers. Further, the set of foot soldiers jointly advance to take and hold the ground vacated by the members of the retreating enemy force" (ibid).

At the level of the sub-groups, or "level-two" actions then, we find three distinct ends, that each specialized sub-group aims to realize (destroying the gun emplacement; providing air cover and advancing on the battlefield). However, as these specialized sub-group actions can be subsumed under the collective end of the military unit, which comprises each of these sub-groups, they constitute the single, joint action of fighting a battle in order to defend an enemy. Why? Because this macro-action in question is *brought about by the individual members doing something else*.[83] Fighting a battle *by* taking hills, *by* providing air cover, and *by* destroying gun emplacements etc*.* is not a basic, or primitive institutional action. It can be further decomposed into (more) basic, or primitive actions of the individuals acting *qua assigned role*, i.e., performing the assigned tasks and functions definitive of the role and pursuing its ends. So, Miller concludes, "from the perspective of this level-two joint action, and its collective end, these constitutive actions are (level-two) individual actions" (ibid).

Turning back to the problem of the apparent exclusivity of institutional actions, we are now in a position to resolve it. Regarding actions of institutional groups that seemingly none of its individual members can perform, we do not have to commit ourself to a non-reductive, independent group-agent, since the individual contributions, *nested into a functionally integrated, layered structure*, suffice for explaining the occurrence of the institutional group action. What was missing was the distinction between basic, or *primitive action types* and *complex action types* on the one hand, and the concept of *essentially collective action types* on the other hand. Crucially, the above mentioned types of actions, such as *merging, disbanding*, or *launching at the stock market*, are all instances of what Ludwig calls *essentially collective action types.* Essentially collective action types, as Ludwig defines them, are action types that no individual could *in principle* perform (cf. Ludwig 2017b, 261) on his or her own.

But although we tend to express essentially collective activity types through the use of singular verbs, the fact that collective action types can be *essentially collective* does by no means exclude these actions types from being *complex* in nature. And once we realizing e that essentially collective action types can consist of complex, composite actions, we start to see how they can be constituted out of more basic, and functionally integrated individual sub-actions embedded in a layered structure. Ludwig claims that this applies to other forms of *essentially collective types of action*, too, such as meeting someone or singing a duet:

> "That the students met in the library requires the participation of all of them, and no one of them alone could have done what they did, even in principle. But it does not require a super-agent, only their individual contributions to their coming together in the same place. Similarly,

[83] See also Copp (1979) for an early version of this argument.

> the Supreme Court making a ruling does not require a super-agent over and above all of them, but rather each justice making his or her individual contributions to their making a ruling by voting in his or her capacity as a justice on the Supreme Court" (Ludwig 2017b, 262).

With a role-based explanation of the layered structure of institutional actions, the problem of the apparent exclusivity of institutional actions then seems to vanish. It can be identified as a semantical obfuscation stemming out of the surface grammar of our ordinary talk, i.e., the verbs through which we express *essentially collective, complex activity types.* It is not the phenomenon itself but rather the talk which aims to capture it, that leads to the problem of exclusivity.

Proxy agency

Another critical feature of institutional roles, which is related to the specialization of tasks and functions, is that they allow for what Ludwig calls *proxy agency* (see especially Ludwig 2014; 2018a; 2018b). Proxy, or *representative* agency describes the phenomenon of institutional groups acting through only a segment, or just one of their members. Examples of proxy agency are a state declaring war *via its parliament passing a resolution to declare war*, a corporation declaring bankruptcy *through the corporation's lawyers filing papers*, or a jury sentencing someone *via the jury's foreman announcing the verdict* (cf. Ludwig 2017b, 188). Ludwig acknowledges the importance of explaining such cases of proxy agency, and it being a "pervasive feature of institutional action" without which "scarcely any modern institution's workings can be understood" (Ludwig 2017b, 186).

Proxy agency addresses the flip-side of the above mentioned *Counting Problem*. The Counting Problem casted doubt on the idea that an institutional action can sufficiently be reduced to contributory actions of its individual members. The problem of proxy agency, in turn, seems to cast doubt on the claim that it is *necessarily* the case that institutional group action can be reduced to the contributory actions of the group's members. Recall that on Ludwig's multiple agents account, to say that a group did something is to say that there was an event of which *each* member of the group were the direct agents of. But if this was correct then how should we explain cases were the *group* does something via a proxy while most members (or all but one) contribute nothing at all to it?[84]

To explain the phenomenon of proxy agency, we can draw back on Ludwig's idea of status roles being *collectively accepted* status functions assigned to individuals. Proxy agency is essentially social, i.e., the basic idea is that, in cases of proxy actions, those actions only *appears* to be an action of just one individual, while yet, it is in fact all members who are contributing to bringing this action about. But how are non-proxy individuals contributing to bringing the action about while just sitting around and doing nothing? The reason that the actions of proxy-agents are only *apparently* actions of single individuals, is that the contributions of the other group members lies in their *collective acceptance* or *authorization* of the individual actions. Ludwig argues that the actions of those individual proxy-agents are the *culmination* of contributions of all the members of the group, only that the contributions of the other members are

[84] For a discussion of proxy-agency that draws an *inflationary* conclusion concerning a group's capacity for assert false statements, i.e., a group's capacity to lie, see Lackey (2021).

obscured, "because they contribute in *very different ways* to the group action of which the proxy agent's actions are the culmination" (Ludwig 2017b, 189) [own emphasis]. The way in which the non-proxy members of a group action contribute to the proxy-agents action is by conferring a certain authority (or accepting that this proxy has a certain status regarding her member-to-action dimension of power) upon the proxy agent in virtue of *collectively accepting* that the proxy-agent has a certain status role which they comply to. This can be captured as a relation of authorization that links the individual actions of the proxy-agent to the group that she is acting *for*. In turn, the actions of just one individual can be attributed to the whole group. The proxy represents the whole group "because other members accept the arrangements […] by which the proxy is assigned" (Ludwig 2020a, 192). To then say that the group did something is to say that there was an event of which *each* member of the group were the agents of: the proxy-agent performing the action and the non-proxy members authorizing her to do so.

Ludwig discusses proxy agency primarily in terms of spokespersons (see: Ludwig 2017b, Ch.13.2) who are authorized to perform speech acts, i.e., to make announcements or assertions, which count as the announcements or assertions of the group they are speaking for (see especially for these cases: Ludwig 2018a). Yet, proxy agency can take different forms than that. Proxy agents may possess powers to act in the name of the group that include the assignments of *other* proxy agents, or they may formulate a group's (general or domain-specific) policy directed at joint actions, issue orders, or make judgments on the group's behalf, undertake investigations, etc. (cf. Ludwig 2020a, 191). The authorization of proxy agents here is to be understood as *transitive*, i.e., if a group authorizes an agent to assign further proxies, the group also authorizes those further proxies and a group that authorizes proxy agents to make judgements on the group's behalf, endorses these judgments, too (ibid).

On a last note, for Ludwig, the collective acceptance of proxy-agents must by no means be understood as an active endorsement, and it can occur *without* the individual agents being aware of which exact individual is occupying a given proxy-position. Rather, individuals can simply accept the institutional group structures, i.e., the *institutional arrangements* of status roles and corresponding responsibilities under which proxy-agents come to possess their representative power. Such arrangements can be complex and individuals might only have a vague understanding of all of the institutional arrangements. Here, Ludwig argues that "one doesn't have to know all the details of an institutional arrangement to agree to them, just as one doesn't have to read the fine print in a contract to take on the commitments they entail in signing it" (Ludwig 2017b, 199). Still, if an individual accepts the institutional arrangements of an institutional group, e.g., by voluntarily joining it, this can be regarded as an form of authorization of proxy-agents, who act as on behalf of the institutional group. Partial or complete knowledge of the institutional arrangements, including knowledge about the specialization and division of roles, or about which specific individuals fill the roles, is not needed in order to agree to the arrangements (ibid).[85]

[85] See for several objections and replies to the concept of proxy agency: Ludwig 2017b, Ch. 13.3

Now might be a good moment to briefly pause and recapitulate: In institutional groups, individual role-occupants get assigned specialized tasks and functions, which, on the one hand, correspond with interpersonal rights and duties. On the other hand, the specialization of such tasks and functions functionally integrates such tasks and functions into a layered structure of institutional action. When combined, the specialization of tasks as well as the integration of individual action into layered structures constitute the level of organizational complexity, where cooperation no longer seems to be necessarily based on the collective intentional states of the individuals involved in bringing about the institutional action.

We might first look at this from a top-down perspective: What we are inclined to view as a solitary institutional action can in fact be broken down into smaller sub-actions, which are assigned to individual role-occupants in the form of specialized tasks or functions. When individuals are assigned such tasks, they do not need to know about the tasks of *other* role-occupants in order to fulfill them. This way, the institutional-level action can come about without the individual involved having collective intentional states, e.g., by Tuomelanian *we-intentions*, or Gilbertian *joint commitments*.

This resulting *individualistic character of institutional sub-actions* justifies Ritchie's claim that, once a level of organizational complexity is reached, "minimal cooperation need not involve anything as mentally complex as that posited by Bratman, Gilbert, or Searle. […]" (Ritchie 2020a, 101f.). Thus, the collective intentional ladder can be knocked away.

Such a top-down perspective reveals another remarkable feature of institutional action: i.e., that it can be *planned*, or *pre-designed.*[86] This feature of institutional action is directly connected to the features of institutional roles both corresponding to deontic, interpersonal powers and the feature of roles being interdependent. The two features combined explain how member-to-member relations of power of institutional roles can influence the member-to-action relations of power. Strictly speaking, institutional roles then encompass a *member-to-member-to-action* relation of power. Let me quickly digress on this point: Because roles are hierarchically structured, the tasks and functions which role-occupants are assigned (the member-to-action relations of power) can be determined, changed and re-directed by *other* role-occupants (in virtue of the other role-occupants' member-to-member relations of power). Occupying a role that is hierarchically connected to other roles then not only accounts for the ways in which roles can be "superior", or "subordinate". The hierarchical connection of roles also explains how one can be ordered certain actions by one's superiors, or, in turn, how one can *plan* to explicitly alter, change, or re-design the design-specifications (specifying the tasks and functions) of the institutional roles that are subordinated to one's own role. Note that this feature is not confined to a "one-on-one" relation, but that the institutional actions of several (and indeed, many) role-occupants can be *pre-planned* in such a way by one, single role-

[86] Further below (Sec 4.1), I will argue that the pre-design of institutional roles' design-specifications leads to problems if cannot take into account the contingent, changing, or ever-evolving circumstances under which such pre-designed tasks and functions must be *applied*, or *carried out*. Just because a pre-designed division of labour seems able to bring about a component-action "on paper", this does not mean that it can be perfectly executed under actual, and contingent circumstances. Planning actions *in vacuo* is something different than performing these actions *in situ*.

occupant. We nowadays tend to call such role-occupants *managers*. Finally, the tasks and functions of multiple role-occupants can be pre-designed by one individual in a way which functionally integrates them into a *layered structure*. Again, in such a pre-designed *division of labour,* the individuals which are assigned pre-designed sub-tasks do not need to know about the tasks of other role-occupants in order to fulfill them. Such pre-designed component-actions are *individualized,* i.e., they are actions that individuals can carry out on their own, without the need for direct cooperation with others. As such, they can come about without the individual involved having collective intentional states.

Facing the initial challenges

Thus far, we have gathered the key components of giving a role-based explanation of institutional group action, i.e., an understanding of acting *qua role-occupant*, which encompasses performing *specialized actions* which are embedded in a *layered structure*, where some actions of individual role-occupants are *representative*, or proxy-actions, backed up by the collective authorization by other members. It seems as these features can explain characteristic ways in which institutional groups perform actions. But can they answer the initial problems of *Action Integration* and *Diachronic Group Constitution*?

First, the problem of *Action Integration,* stemming out of the seemingly complex relation between individual role-occupants and institutional actions, can be straightforwardly explained by pointing to institutional actions being comprised out of *layered structures* of individuals performing role-based, specialized sub-actions. Within large institutional groups, the myriad of individual sub-actions, benign or not, can be described as constituting a group-level action, insofar as these individual actions are layered, and thereby functionally integrated into in a system of institutional roles to bring about the group-level action. However, it is not the case that just *any* individual action therefore contributes to, or is constitutive of a group-level action. This is because roles "*dictate*" (Miller 2001, 172) certain types of individual actions by which such a layered institutional action is performed. So only *those* actions which are related to the tasks or functions of the role get "counted" as contributing to institutional actions. In turn, such individual actions being severally necessary and jointly sufficient to bring about a group-level action, the resulting action is constituted by nothing more than the actions of individuals, i.e., group-level actions bottom out in the single, (basic or non-basic) actions of individual members. It is also possible that the actions of one single individual are sufficient to bring about a group-level action, i.e., in cases of acting *as a proxy*. Proxy agency, via its dual mechanism of *authorization* and *representation*, as well as it being *transitive*, can discharge the objection that it is not necessarily the case that a group acts if and only if all its members act, while the concept also aligns with a reductive explanation of group actions.

The second, related problem of *Diachronic Group Constitution* can also be explained by referring to the above mentioned features of institutional roles, as well as their relation to institutional group agency. With Ludwig, we come to understand the importance conceptualizing membership-relations as *time-indexed,* and the need for specification of temporal intervals when talking about those groups. Explaining how we could make sense of the sentences like "The Supreme Court ruled in 1896 that segregation is constitutional but in 1954 it reversed itself and ruled that segregation is not constitutional" (Ludwig 2017b, 260), Ludwig states that:

> "at each time the Supreme Court is entirely constituted by its ε-members at the time—even if there being a Supreme Court requires an institutional context and so other people playing appropriate roles in other institutions. It may at a different time have different ε-members because some lose the status role or cease to exist and others acquire it. This does not make the Supreme Court something over and above its ε-members, subsuming but not being constituted by them. Taken as a persisting entity, it merely means that it is a group consisting of all the people who were ε-members of it *at one or another time* […] Its persistence is a matter of there being a temporal ordering of groups of ε-members under the being-a-justice-of-the-Supreme-Court at *t* relation" (ibid) [own emphasis].

Relating this aspect of time-indexation to institutional actions, Ludwig then states that when talking about what an institutional group does, one always talks about it doing something *at a particular time*:

> "When we talk about what an institution undertakes and does over a period during which its ε-membership changes, we invoke a distributive quantifier over ε-members of the group during that time interval. When we talk about a group doing something at time t and something else at a time t', we have in mind that the ε-members of it at t did the one thing and the ε-members of it at t' did the other" (ibid).

Thus, whenever an institutional group acts "it is because those who have appropriate roles in the institution *at that time* act" (ibid) [own emphasis]. So if we take time-indexed membership-relations into account, the challenges posed by the diachronic endurance of group action can be overcome. So it might *seem* that these actions which are attributed to groups cannot be the actions of any group that constitutes its ε-members at any particular time. But as one and the same status role can be successively occupied by different individuals throughout time, the problem resolves: "Sentences which appear to assert that a single institutional entity did things at different times with different ε-membership resolve into claims about those with the relevant status roles at the one time doing something and those with the relevant status roles at the other time doing something" (Ludwig 2017b, 261).

3.3. Anonymous Cooperation and Compartmentalized Action?

In this section, I will argue that role-based explanations of institutional agency can account for two critical features the accounts of collective action (recall Ch. 2.2.2 and 2.2.3) failed to make sense of, i.e., the anonymity and compartmentalization of institutional action. I think that, ideally, any account of institutional group agency should be able to explain cases like the Calutron Girls, where compartmentalization and anonymity are involved. Now I hold that a role-based account of institutional group agency can indeed account for such cases. However, I still carry with me the burden of proof to substantiate this claim. The rather simple thought that I want to develop in this section is this: Two individuals, A and B, may occupy roles which require them to individually do something, the result of which is a collective action. Now for A

and B to do this together, it doesn't matter to A *who B is*, as long as A knows *what B's going to do* (and vice versa). And in order for A to know *what B's going to do*, it suffices that A knows what actions *the role* of B entails. In order to fully develop this argument, certain obstacles need to be overcome.

To do so, I will now discuss a general challenge for giving a role-based explanation of anonymous cooperation and compartmentalization within institutional action, which is posed by Katherine Ritchie (2020a). Ritchie's argument for a role-based account of institutional action starts with the claim that for members of institutional (or as she puts it "organized") groups to be minimally cooperative, this requires two things: 1) them playing roles in an organizational structure and 2) them having a common goal (cf. Ritchie 2020a, 93). The challenge here is to explain how the condition of having a common goal can be harmonized with the feature of anonymity. Can two individuals work towards a common goal when they don't know each other? The main aim of this section is to answer this question affirmatively.

Before assessing her example, let me quickly clarify Ritchie's use of the term "cooperation", so that we do not misinterpret her claims here: Ritchie marks off her use of the term primarily against competing usages in the field of collective action, stating that "cooperation involves collectively intending to ϕ with others" (Ritchie 2020a, 97). While ultimately, she wants to argue for an account of cooperation that does not involve collective intentional states, she discusses her approach mainly in demarcation of Bratman's account of shared cooperative activity (e.g. Bratman 1993), Gilbert's analysis of collective action based on joint commitment (e.g. Gilbert 2006), as well as Searle's account of irreducible we-intentions to ϕ (e.g. Searle 1990). So, as I interpret Ritchie, she wants to explain cooperation as being a form of (or maybe even term for) collective action. Collective action that does not, however, necessarily involve collective intentional states.

So let us take a closer look at Ritchie's main example of minimal cooperation within organized groups. Ritchie here describes a case of specialized, role-based, and layered individual actions that realize a collective outcome. Her example is a consulting firm, whose task is to determine whether the merger between two companies would be beneficial to the former:

> "In order to determine what to recommend, thereby meeting their goal, the firm puts together a consulting team with various roles. *Suppose further that the team has many members who are located across multiple offices and that many team members do not know of one another.* Roles involve responsibilities and obligations that normatively bind role-players in various ways. The consulting team includes roles that require role-players to research similar past mergers and pass findings on to members who will include them in a report. Other members have roles that require analyzing a merger's impact on stockholder and customer perception. They too report their findings to members tasked with writing a final report. And so on. *Through many members playing their assigned roles—that is, carrying out tasks and interacting in ways team roles require—the team concludes that Company A should merge with Company B as it will benefit Company A to do so*" (Ritchie 2020a, 102) [own emphasis].

The question that Ritchie now wants to answer is whether such a case involves individuals *cooperating*. Why should we assume that the individual employees of the firm *worked together* if they do not necessarily

know each other; or if they are unaware of each other's contributions? At the heart of the challenge to explain anonymous cooperation, then, lies Ritchie's second requirement of minimal cooperation, which states that cooperation can only occur if institutional roles have a common goal. In order to draw out this problem more clearly, she considers the case of a spy network, where several individuals, who are unbeknown to each other, perform individual actions, and where the combined actions of said individuals contribute to a certain end, but where no individual is aware of this end. The individual spies in the network, Ritchie explains,

> "might not know anyone else in the network. Moreover, they might not understand what their roles are, what end they are helping to work towards, or how they are contributing to that end. Nevertheless, the combined efforts of those in the network might fulfill [a particular] end" (Ritchie 2020a, 105).

Ritchie's question here is whether the combined efforts of the individual spies constitute an instance of cooperation. The first requirement of each individual performing their assigned tasks and functions seems fulfilled.[87] But according to Ritchie, it is unclear whether the spies actually *cooperate*, because it's unclear whether they fulfill the second requirement, i.e., whether they are working towards a common goal. She states that intuitions about whether such a case counts as an instance of cooperation are mixed, yet her account could be adapted to accord to both stances on the issue:

> "If one takes the spies *not* to be minimally cooperating, the account could be adapted to require some mental requirements on having a common or shared goal. For instance, one might argue that minimal cooperation in ϕ-ing requires playing roles in an organized group structure that they work *towards a goal that all know*. In the spy network case, not all know the goal, so the case does not involve minimal cooperation. In contrast, if one takes the spies in the network *to be minimally cooperating*, *one could take having a common goal to require less*. It might require that roles are functionally integrated to achieve an end. Or, one might require just that someone with authority over the group know the goal for it to be a common goal" (Ritchie 2020a, 105) [own emphasis].

So let us discuss these options. The first option would be that individuals occupying roles and fulfilling the corresponding tasks only cooperate if all of the group members know what the goal of the group is that

[87] Note that her description of the spy network is somewhat ambiguous and self-undermining, as the example could see it that the individuals involved "might not understand what their roles are" (Ritchie 2020a, 105). This can be understood in several ways. If it was the case that the individuals did not know what their tasks and function are, it would indeed be puzzling why, on her own account, these individuals should be cooperating. In fact, without them knowing their roles *understood in this way*, it remains unclear how (and why) the spies in the network would contribute any efforts to fulfill a particular end at all. So I interpret this to mean that the individuals involved might not understand what *function* their roles play in the overall group's goal to bring about a collective outcome, and that they do not know how their role relates to other roles within the group. This, in turn, does not imply that they don't know what their role *requires them to do*.

they are contributing to by playing their roles. This would rule out the spy case, but also other cases of cooperation like the above mentioned Calutron Girls. Recall that the Calutron Girls were (deliberately kept) in partial or complete ignorance of the end of their contributions in producing the first atomic bomb. The Calutron Girls may exhibit all features of Ritchie's definition of occupying a role, i.e., they may perform their prescribed tasks, their tasks may be interrelated, and they may be subjected to specific membership conditions (not spying for the nazis being the most important condition). The skeptic, however, would doubt that this was a case of the Calutron Girls *cooperating* because they lack common knowledge about the groups overarching goal or goals. I'll explain down below why I think that the case of the Calutron Girls is a case of cooperation.

But first, consider the consequences of the claim that cooperation *necessarily* involves common knowledge of a shared goal. The standard analysis of common knowledge (Lewis 1969) sees that for a group of individuals to possess common knowledge (e.g., that p), this implies not only that each individual knows that *p*, but also that each individual knows, that all the other individuals know, that *p,* etc. (see for a discussion of the different analyses of common knowledge: Vanderschraaf & Sillari 2022; for a defense of Lewis' notion see: Paternotte 2011). The resulting belief-cascades not only pose a heavy cognitive burden for individual group members, but seem too strong of an assumption when applied to institutional groups. This concept of common knowledge presupposes that all individual members in institutional groups are aware of even the *existence and group-membership* of all the other individual members. But while this might certainly be the case for *some* institutional groups, it is far from clear how this would be the case for *all forms of institutional groups,* especially for large and complexly structured institutional groups like e.g., multi-national cooperations or trade-unions. If common knowledge was indeed necessary for cooperation to occur within such groups, Ritchie's account would run into the very *Upscaling Problem* she criticizes the other accounts of (cf. Ritchie 2020a, 102ff).

So we need an argument that common knowledge of a goal is *not necessary* in order for the cooperation of individuals occupying roles to achieve to occur. Can we find such an argument? My strategy here is to argue that the functional integration of roles indeed can suffice to speak of cooperation occurring in light of a achieving a common end. And this is because institutional roles *are action-specific but agent-ambiguous.* I will explain what I mean by this in a bit.

To make my case, it is useful to draw on the theory of collective action provided by Seumas Miller (2001; 2010). Choosing Miller to discuss this point is helpful in several respects: First, Miller argues for a role-based explanation of institutional action which is reductive and deflationary in nature. As mentioned above, Miller takes institutional actions to be the specialized, and layered actions of individuals who fulfill the tasks and functions of their assigned roles. The basic idea by which Miller integrates his CET into an account of *institutional* action is that institutional actions are constituted by secondary, or sub-actions of individual role-occupants, which are functionally integrated into a *layered structure*. For Miller, an individual end that each role-occupant has in fulfilling her tasks and duties then depends on, and can be subsumed under the *collective* end of the institution that she is part of, and on the basis of which she occupies her role. Miller offers arguments for the weaker conditions of functional integration and authority-according division of labour to be actually sufficient in order to guarantee the capacity for individuals to cooperate towards a

collective end. The account ultimately then does not rely on symmetrical intentional states such as common knowledge. Applying Miller therefore seems to be compatible with what we have established so far.[88]
I will try to show that Miller's account is apt to deal with the problems the skeptic poses for Ritchie's account of minimal cooperation. Because this is the notion that we want to clarify, his theory allows us to directly tackle - and consequently overcome - the skeptical challenge.

Seumas Miller's teleological account

Seumas Miller's teleological account of social institutions (see especially 2001; 2010) centers around the explanation of institutional group action (or as he calls it: corporate action) as *joint action* aimed at the realization of collective ends. For Miller, the central concept in his *teleological* account of institutional agency is that of *joint action,* which, according to Miller, consist of the intentional individual actions of a number of agents "directed to the realization of a collective end" (Miller 2019). The central concept in the teleological account of institutional agency is that of *joint action,* which, according to Miller "consists of (1) a number of singular actions and (2) relations between these singular actions. Moreover, the constitutive attitudes involved in joint actions are individual attitudes; there are no sui generis we-attitudes" (Miller 2010, 39). According to Miller, a joint action comes about when the intentional individual actions of a number of agents are directed to the realization of a collective end. This is the "Collective End Theory" (CET) of joint action (Miller 2010, Ch.1). Fundamentally, the CET sees that joint actions are actions, which are directed to the realization of a collective end. A collective end, however, is not the end that a collective (or group) itself has, but rather "it is an end possessed by each of the individuals involved in the joint action. [I]t is an end that is not realized by the action of any one of the individuals; rather, the actions of all or most realize the end" (Miller 2001, 24). So collective ends are interdependent individual ends, i.e., one comes to have one's own collective end only if others have it, and vice versa. Let's look at Miller's core analysis of joint action. Miller states that, "individual actions *x* and *y,* performed by agents *A* and *B* (respectively) in situation *s,* constitute a joint action iff:

(i) *A* intentionally performs *x* in *s* (and *B* intentionally performs *y* in *s*);
(ii) A *x*s in *s* if and only if (he believes) *B* has *y*ed, is *y*ing or will *y* in *s* (and *B y*s in *s* if and only if (he believes) *A x*s or is *x*ing or will *x* in *s*);
(iii) *A* has end, *e,* and *A x*s in *s* in order to realize *e* (and *B* has *e,* and *B y*s in *s* in order to realize *e*);
(iv) *A* and *B* each mutually truly believes that *A* has performed, is performing or will perform *x* in *s* and that *B* has performed, is performing or will perform *y* in *s.*
(v) each agent mutually truly believes that (2) and (3)" (Miller 2001, 57).

[88] I hold it that Miller's account of collective action occurring anonymously on the basis of institutional roles having a common goal can be consequently harmonized with Ludwig's shared plan account. Recall that on Ludwig's account, joint action occurs on the basis of *shared plan*. Now for Ludwig, the requirement for a plan to be shared simply amounts to the requirement "that one may know enough about the plan to do one's part but not know the details of the other parts" (Ludwig 2017a, 214) and that "having the same determinate plan need not imply that they all know all the details about it" (ibid) nor that the participants "know who the other participants are or how many there are" (ibid). As Ludwig indicates, he holds individuals to be able to share a plan to achieve a collective goal *anonymously*, i.e., without knowing who the other participants of the shared plan are.

One might worry that this initial definition of joint action is too restrictive to explain institutional actions. This worry might be partially motivated by the fact that this basic form joint action captures small-scale interactions such as two people dancing; or people building (or robbing) a house together (cf. Miller 2001, 53). And as I have argued, accounts that rely on interdependent mental states (such as common knowledge or mutual belief) of individuals in small-scale settings encounter implausible consequences when "scaled up" to larger forms of collective action. Note that the conditions of mutual belief of Miller's core analysis seems especially problematic for our purposes.

Could we explain the actions of large institutional groups by appealing to these kinds of collective ends of all of the members? If Miller's view requires mutual knowledge or belief of collective ends, this seems implausible. How could individuals have mutual beliefs about each other's goals if they do not know each other or interact with each other in the first place? To this end, Miller qualifies that the CET of joint actions is supposed to be understood as a "core theory" (Miller 2010, 44ff.) and that both the conditions of collective end and mutual belief can be modified to be applicable to other forms of social action.

First, collective ends can be held and pursued either *consciously* or *unconsciously*: "An example of a collective end that those who pursue it are not fully conscious of might be some end pursued by members of a particular social class, say to exploit workers" (Miller 2001, 63). Second, collective ends can also be held either implicit or explicit and they can be actively or latently pursued. Third, joint action does not necessarily require that all participants directly communicate with each other, as communication can occur indirectly: *A communicating with B and B with C but not A with C* etc*.* (ibid). Fourth, individual contributions to a joint action can occur separated from each other in time and location, resulting in joint actions to be long lasting, even intergenerational joint projects, where individuals can be separated by "thousands of miles and/or hundreds of years" (Miller 2001, 63) as in the example of building the Great Wall of China.

Crucial to endeavor, however, is that Miller argues that the requirement of mutual belief can be replaced by "something weaker than belief" (Miller 2001, 57ff.). Participants in joint action may solely think that "it is likely," or even "thinks it is quite possible" that other agents do (or will do) their part in a joint action (ibid). This will be important later. However, some form of doxastic attitude about the actions of other agents must remain, although it does not necessarily have to be mutual belief: "For if the first agent did not think there was any chance the other agent would do his part, then how could the first agent be thought to have the end in question? After all, the end can only be achieved if both agents do their parts" (ibid).

My goal now will be to provide arguments for the following claim: I will try to show that it suffices for such a condition that individuals acting in institutional groups only have *to expect other agents to perform the actions constitutive of their roles.* This, however, can be achieved while the actual individual occupying the role remains unknown or unspecified. The *agent-ambiguity of roles* can explain how joint action may occur anonymously, while it also allows to demarcate such cooperation from (mere) instances of parallel individual action. In order for an individual to have expectations that some other, unspecified individual will perform the tasks and functions of her institutional role, it is not necessary to presuppose mutual belief on behalf of these two individuals. It suffices that of both have a doxastic attitude about the *actions* of another agent. The clue here is that roles don't specify *who* is going to perform certain tasks, but rather *which tasks are going to be performed.* Roles are *agent-ambiguous* but *action-specific*.

Remedy for the skeptic objection

Let's turn back to the initial challenge of this section, i.e., to explain whether individuals performing the functions of their roles could be said to be cooperating even though they are not aware of each other's existence and do not have mutual beliefs about each others goals. Answering this affirmatively could overcome Ritchie's skeptic challenge, and thereby allow us to capture cases of collective action where role-occupants remain anonymous to each other. For Ritchie, this means that we need a way of explaining collective action which sees as a sufficient condition for cooperation that roles are either functionally integrated to achieve an end, or that someone with authority over the group knows the goal for it to be a common goal (cf. Ritchie 2020a, 105). I hold it that the above introduced theory of joint actions helps us to explain how institutional roles can be functionally integrated to achieve an end, and therefore sufficiently explain collective action without mutual belief of a common goal. These cases, then, are cases of actual cooperation, and not just parallel individual action.

Let me start to develop my argument by pumping our intuitions. Let's imagine the following scenario: Two painters, who don't know (of) each other, drive by an abandoned, ugly house each day on their way to work. The painters drive by the house from different directions, so that the first painter only sees the back and right side of the house, and the second painter sees only the front and left side of it. One day, each individual painter, without knowing about the existence of the other one, and in a rush of professional ethos (or perhaps because of strong aesthetic preferences), decides to stop after work and paint the part of the house which is visible from her commute, so to not endure the eyesore any longer. The first painter starts to paint the back of the house and the other starts to paint the front, both working clockwise. After each individual has painted their half of the house, they leave the site, still in complete ignorance of each other. The house, then, is fully painted.

Now one might be hesitant to call this an instance of collective action. Rightly so. After all, there was no shared intention to paint the house, nor was there a collective commitment, nor a shared plan or collective end to be realized. Just like two drivers crashing into each other in an intersection, an outcome to which more than one individual contributes to does not suffice to talk about a collective action.

But how could we model anonymous cooperation so that it does not simply amount to such parallel individual action as in the case of the two painters? Two core features of institutional roles are in a good position to explain how individuals can cooperate anonymously, and how this cooperation can still be an instance of a *collective* action: roles are both *agent-ambiguous* and *action-specific*. So without presupposing mutual belief, two individual actions can still constitute a collective action in virtue of being functionally integrated actions performed *qua* an assigned role.

To see why this might be plausible, consider a second, contrasting case of two painters being hired by a third party to paint the house. They each independently receive a call from the contractor telling them that she wants to have the house painted. The contractor also tells each painter that she is hired to only paint one half of the house, while the other half will be taken care of. However, the contractor tells each painter that she will get paid only after the full job is done. Each painter, without knowing of the other, drives to work and proceeds to paint one half of the house. The first painter takes care of the back and right side of

the house and the other painter paints the front and left side. After each individual has painted their half of the house, not having seen or heard of the other painter, they each leave the premise. The house then is fully painted and both receive their paycheck.

Let us assess this example. This (somewhat primitive) case exhibits all the above mentioned features of an (institutional) action based on institutional roles. The individuals act *qua role-occupancy*, i.e., they perform the tasks definitive of their assigned roles (and not, e.g., as private individuals). Further, these tasks are minimally *specialized* (to paint only one half of the house) and *layered* (insofar as both individual actions are severally necessary and jointly sufficient to realize the collective end of the house being fully painted). Finally, it also seems to be a case of two individuals having a common goal (or collective end) while remaining anonymous to each other: Whereas in the first example, each individual acted out of individual (aesthetic) interests, the individual end of each painter in the second case partially dependents on the end of the other painter. So although unbeknown to each other, they seem to have an collective end, an "end that is not realized by the action of any one of the individuals; rather, the actions of all or most realize the end" (Miller 2001, 24). Note, that the realization of the individual ends of receiving the paycheck dependents on the collective end of the house being painted. It is, however, not necessary that each painter has a collective end (the house being fully painted) that is *not* a means to an individual end (receiving the paycheck).[89] So why is the second case of the painters a case of *cooperation* occurring anonymously? To see why, let's have a closer look at the relation between individual action and institutional roles.

Here is my idea: although it's implausible to assume that the painters have mutual belief about each others goals and actions, they might very well "think it is likely" or "thinks it is quite possible" that *some other, unspecified* or *anonymous* agent will engage in the collective (or joint) action. It is implausible to attribute mutual belief here, because neither painter could plausibly address some *specific* individual with whom she might share such a symmetrical, or recursive state, i.e. a belief about the other's belief (about the other's belief etc.). However, each might think of it as likely, or quite possible*,* that the *task* of painting the other half will be performed by someone else. After all, this is what the contractor told each of them.[90] But having a belief about *the likelihood that an action could or will be performed* is something different than having a belief about *another individual's mental states,* even if said performance would require the existence of another individual with mental states.

---

[89] See for a similar case of two painters Shapiro (2014, 21ff.). Shapiro, in an attempt to criticize Bratman's theory of joint action, however, focusses on the fact that the two painters can be *alienated* from their task, not that they may carry out their task anonymously. Nevertheless, Shapiro draws a similar conclusion, granting that two alienated painters can paint a house together in the absence of a commitment to do so jointly. Shapiro states that "Baker and Charlie [the two painters, M.G.] are both alienated from the project of painting the house. They don't care a wit about painting the house, only in getting their money. Indeed, they may hate Abel [their contractor, M.G.] and not want him to have a nicely painted house. Yet, they can still paint the house together, and do so intentionally, even though neither of them intends that they paint the house together. They paint together, in other words, despite the fact that neither of them possesses a plural intention, let alone share that plural intention" (Shapiro 2014, 23). It's also worth mentioning that, as just demonstrated, philosophers *really* like to talk about painting houses.

[90] Each painter might even draw individual conclusions as to what the contractor meant by saying that the other half "will be taken care of". They do not have to have a specific understanding, such as when, how, or where the procedure-governed task of painting the other half of the house will be performed. They even could think that the other half will be painted by specially trained chimpanzee in an attempt to drive down their wages.

So what makes this impersonal performance conceivable is that roles are an *abstraction,* or *placeholder of procedure-governed tasks*. Each painter might be aware of the *task* that is going to be performed while being unaware of *who* is going to perform it. Still, if it wasn't for the other role-occupant's *action*, neither would have the motivation to paint half the house in order to get compensated. So the interdependence between the two agents is not an interdependence of attitudes, but rather it is an *interdependence of actions*, or, to be more specific, an interdependence of procedure-governed, role-based tasks. And while institutional roles are - often highly - specific about what actions, tasks, functions etc. are to be performed, they are *non-specific* with regard to which particular individual is to perform the actions, fulfill the tasks, or exercise the functions etc. Institutional roles *are action-specific but agent-ambiguous.*[91]

Thus, because roles are constituted by procedure-governed tasks and functions, all that is needed is that both painters think of it as likely, or *quite possible* that the interlocking tasks of the role will be performed and the corresponding actions carried out. But roles are interchangeable, and thereby do not specify *who* performs a task.[92] It could be about anybody capable of doing so.[93] So in virtue of the feature of *non-specificity,* neither painter has, nor needs some doxastic state about *who* will be performing the constitutive

[91] Hans Bernhard Schmid makes a similar point when talking about the normative *expectations* we have regarding the performance of actions of role-occupants. Schmid, talking about *social* roles in general, claims that they come with normative expectations of behavior attached to them. Interestingly, he discusses an example of individuals performing certain actions within *professional* roles of butchers, brewers and bakers (an example he derives from Adam Smith's theory about the division of labour): "But when we buy our meat, beer, and bread, we do not *usually* speculate about any *particular* agenda of the people who produce it or sell it to us, beyond expecting them to be motivated to enter the exchange. We're not thinking about the motivational role of their income, their job satisfaction, or their concern for their customers. The fact of the matter is that we simply expect them to be *doing their job* (selling meat, beer, bread) and to be doing their job reliably or robustly rather than just by incidence. We're not thereby ignoring that people whom we expect doing their jobs are *agents* rather than just cogs in some production, distribution, and consumption machinery. We do know that they are doing what they do based on their guise of the good therein, and if *this* is 'their regard to their own interest,' nothing seems wrong with Smith's positive claim. We know that *they would not be doing it if it made no sense to them*, under any 'description,' but again, this is just to say that we see them as subjects, or *agents*" (Schmid 2023, 231) [own emphasis].

[92] Again, Shapiro (2014) has something very similar in mind when he talks about impersonal relations of authority opening up the possibility of what he calls "offices", which I take to be analogous to institutional roles. Offices, Shapiro states, "are relatively stable and persistent positions of power where turnover in occupancy is not only possible but expected. The Presidency of the United States, for example, is an office because it persists from term to term and its normative character does not change merely because one president vacates and a new one assumes power. Presidents come and go but the Presidency remains. *Impersonal authority relations allow for the possibility of offices because the normative relations are not tied to any particular holder of offices, but rather to the offices themselves.* Someone can accept a plan committing himself to follow the orders of anyone who satisfies the qualifications appropriate to the office (e.g., was elected by a majority in a national election). This relation persists across turnover in office-holders and hence *does not require participants to figure out who satisfies those qualifications and to reestablish their commitment to them*" (Shapiro 2014, 43) [own emphasis].

[93] Depending, of course, on whether the role has some membership-criteria attached to it, such as having a certain age, or gender.

actions of the role and what (contingent) attitudes this individual will or could have while doing so.[94] If we take anonymity be a social relation between an anonymous person and others "where the former is known only through a trait or traits which are not coordinatable with other traits such as to enable identification of the person as a whole" (Wallace 1999, 23), then such non-specificity, or impersonality is indeed a form of anonymity. So if this seems adequate, then we can indeed think of the two painters as a case of anonymous cooperation.

Can we account for ignorance, too?

It is noteworthy that my proposal seems to stand at odds with the claim that an agent's ignorance of what she is doing undermines the status of what she does as an intentional action. This claim, dating back to Anscombe (1963), has often been dubbed the *knowledge condition* of intentional agency. Anscombe characterizes intentional action "as that to which a special sense of the question ‚Why?' applies" (Piñeros Glasscock 2023). Consequently, this question "has application only inasmuch as the agent recognizes himself as acting under the corresponding descriptions expressed by his answers" (ibid). Thus, an agent's understanding of what she is doing is therefore taken to be intrinsic to the intentional action that she performs. If, in turn, an agent does not grasp the action that she is performing under such a description, proponents of the knowledge condition argue that we should refrain from positing an intentional action being performed by the agent. Action, then, necessarily involves knowledge by the agent about what she is doing in order to count as intentional.

Given this, the Calutron Girls' ignorance of the fact that they are contributing to building an atomic bomb could be said to ensure that this is not an intentional action of theirs. To this end, one might also argue that it is relevant whether the causal theory of action is actually the correct way to describe the concept of

---

[94] This argument also resembles a similar claim of Bermúdez, who argues that understanding other agents in terms of "frames, social roles and social routines" (Bermúdez 2005, 203-205) does not require one to actively model the other agent's mental states, i.e., using one's mind-reading capacities. To be more specific, he argues that "[o]rdering meals in restaurants and buying meat in butcher's shops are such routine situations that one need only identify the person approaching the table *as a waiter*, or the person standing behind the counter *as a butcher*. Simply identifying social roles provides enough leverage on the situation to allow one to predict the behavior of other participants and to understand why they are behaving as they are. There is no need to make any folk psychological attributions. There is no need to think about what the waiter might desire or the butcher believe – any more than they need to think about what I believe or desire. The point is not that the routine is cognitively transparent – that it is easy to work out what the other participants are thinking. *Rather, it is that we don't need to have any thoughts about what is going on in their minds at all.* The social interaction takes care of itself once the social roles have been identified (and I've decided what I want to eat). One lesson to be drawn from highly stereotypical social interactions such as these is that explanation and prediction need not require the attribution of folk psychological states" (Bermúdez 2005, 203) [own emphasis]. In such cases, "our understanding of individuals and their behavior is parasitic on our understanding of the social practices in which their behavior takes place" (Bermúdez 2005, 204). For Bermúdez, a "surprisingly large amount of our social interaction is carried out by means of social scripts and routines. Whereas it is usual for philosophers to think that social coordination requires forming beliefs about other people's beliefs, desires and intentions we considered the possibility that many social interactions are sufficiently standardized to be successfully negotiated by identifying the relevant social roles and acting accordingly. Once the relevant social roles are identified the script takes over and the interaction runs according to rule (Bermúdez 2005, 222f).

action or not. If not, one could say that the Calutron Girls causally contribute to building the atom bomb, but they are not (intentionally and collectively) building the atom bomb.

One option to reply to this objection would be to just accept the knowledge condition and thus to exclude institutional groups from acting intentionally. Here, one might point out that if individual action necessarily involves knowledge by the agent about what she is doing in order to count as intentional, then *collective* action might necessarily involve a *collective*, or *interactive* form of knowledge by the involved agents about what *they* are doing in order to count as intentional. And because such interactive knowledge is usually modeled in terms of mutual beliefs or other symmetric epistemic states of the individuals involved, this condition will be unfeasible to assume in cases of *institutional* action. So *if* collective action actually requires collective knowledge in the form of mutual beliefs or other symmetric epistemic states of the individuals involved in order to count as intentional, *and if* mutual beliefs or other symmetric epistemic states are implausible to assume in cases of large and complexly structured institutional groups, then institutional groups will (most likely) not be able to act intentionally. (see my discussion of the upscaling problem in Ch. 2.2.1.) Chant, in her critique of what she calls an "isomorphism between group- and individual-level concepts with respect to their explanatory relations" (Chant 2017, 16), argues, that standard approaches to explain collective action typically use "the most basic actions of individuals as its model for collective action, despite the fact that basic actions are the least similar to collective actions" (Chant 2017, 19). These accounts, according to Chant, therefore run into similar problems as my case of the Calutron Girls in modeling the actions of collectives as *intentional* collective actions (see Chant 2017, 17-18).

As I argued above, I do not wish to argue that one should think of institutional groups (simply) as an extension of small-scale egalitarian groups with high interdependency among the members, i.e. to think of them as *large-scale* egalitarian groups involving such high interdependency and psychological symmetries. Institutional groups differ in critical aspects from such exemplificatory cases of small-scale cooperation because they are groups that have a *structure*, and on this basis survive the change of membership, and because they can establish hierarchies allowing for task-differentiation and the division of (cognitive and/or manual) labour and structural sub-division of groups into specialized units.

So instead of excluding institutional groups from the class of agents capable of acting intentionally, another option would be to adopt a *weakened* version of the knowledge condition which is compatible with the characteristic features of institutional groups. Such a weakened version might be found in the works of Davidson. According to his account (see especially: Davidson 2001e), agents do not need to know what they are doing under *every* description under which they act intentionally, but they must know what they are doing under at least *one* such description (his example being the intentional action of spelling tea thinking it was coffee). This idea could then be applied to the case of institutional group action involving the differentiation and compartmentalization of tasks on the basis of institutional structures and hierarchies, e.g., in the case of the Calutron Girls: Even if we accept the idea that an agent's ignorance of what they are doing can undermine the status of her action being intentional, this doesn't necessarily apply to the case of the Calutron Girls. For one, because the Calutron Girls, whose roles saw them to operate machines as part of the Manhattan Project, weren't ignorant of their *immediate* actions. After all, *they knew they were operating machines* and therefore could be said to perform this task intentionally. Rather, they were ignorant of the ultimate purpose of their actions which was to produce the plutonium necessary for atomic

bombs. This distinction, then, might be critical. The Calutron Girls could be said to be not ignorant of their immediate, role-related actions; they were only ignorant of what their role-related actions of operating the machines contributed to. And if such ignorance of the collective goal toward which one contributes is different from ignorance of one's own contributory actions, then their operating the machines could be said to have been an intentional action of theirs, even if they didn't fully grasp the broader implications of what they were contributing to. The resulting picture of this would see that the Calutron Girls were intentionally contributing to building the atomic bomb, because there is at least one description under which they acted intentionally.

Adopting such a weakened version of the knowledge condition has clear benefits for capturing characteristic features of institutional groups, such as *hierarchy* and the *epistemic division of tasks and labor* because it allows us to explain the collective intentional action of institutional groups in cases where even most, or nearly all members are ignorant about what their actions actually contributed to. In the case of, e.g., the Manhattan Project, the partial or complete ignorance about the group's overall goal by most or nearly all members of the project does not stand in conflict with an explanation of the building of the atomic bomb as a collective intentional action, insofar as at least one role occupant, residing on top of the institutional group's hierarchy, could be said to have been knowledgable about the institutional group's intention to build the bomb. If we assume that this role occupant had the right kind of authority over the sub-actions of the group's members, we can explain how the institutional group's action can be described as intentional although most members were ignorant about its full scope. Scott Shapiro, e.g., analyses the submission to such a form of authority in terms of "*intending* to take the content of another's directives as one's subplans" (Shapiro 2014, 17), i.e., to adapt or alter one's role-based tasks and functions in virtue of another group member's *planning authority.* A member, residing at the top-layer of the hierarchy and being granted such a form of planning authority, thus may be in a position to plan and consequently divide the group's action, and to decide, based on the vertical distribution of member-to-member relations of power, which group members had the relevant, domain-relative member-to-action relations of power regarding the compartmentalized group action. The hierarchical nature of institutional groups, then allows for an explanation of group action which both makes room for the specialization and compartmentalization of institutional actions, as well as for the fact that such actions can be *intentional*, at least under one description.[95]'[96]

Upshot

Let's recap what we've established in this section: According to Ritchie, members of institutional groups can be minimally cooperative in virtue 1) them playing roles in an organizational structure and 2) them having a

[95] I do think, however, that ultimately, we ought to strive for a more nuanced understanding of the relation between knowledge, ignorance, and intentionality in cases of institutional action which would allow us to more effectively differentiate between the immediate intentional actions of agents and the unintended broader consequences to which those actions might contribute.

[96] I want to thank Eva Schmidt for extensively discussing this matter with me and for her helpful comments on the relation of intentional action and Anscombe's knowledge condition. I also thank her for pointing out that the case of the Calutron Girls bears structural similarities to the above discussed spy-case of Katherine Ritchie.

common goal (cf. Ritchie 2020a, 93). The challenge was to explain how the condition of having a common goal could be harmonized with the feature of anonymity. If having a common goal includes symmetrical mental states, e.g., mutual knowledge or beliefs about each other's mental states, then cooperation most likely cannot occur anonymously. But if having a common goal means that the actions of individual role-occupants are *functionally integrated to achieve an end*, then cooperation may indeed occur anonymously. To show how such functional integration of role-performances can be achieved, I drew on Miller's theory of joint action.

The upshot of this section is that role-performances can be functionally integrated in virtue of roles being *action-specific* despite being *agent-ambiguous*. Roles are agent-ambiguous because they are interchangeable, and thereby do not specify *who* performs a task. This feature of *non-specificity* then is key to understand how the functional integration of roles to achieve an end consists out of an *interdependence of actions, but not out of an interdependence of attitudes* (see also: Miller 2010, 108-109). Mutual belief among individuals therefore does not seem necessary for collective actions, at least in cases where these collective actions are constituted by functionally integrated role-performances. What secures the realization of the collective action is the performance of tasks and actions, but not the performance of tasks and actions *by specific individuals having beliefs about other, specific individuals*. Just like the case of two spies communicating anonymously through a dead drop is still a case of two spies communicating, two individuals cooperating on the basis of their roles is a case of cooperation, if they do not know the other role-occupant.

3.4. Summary

In this chapter, I examined role-based explanations of institutional agency. With the label "role-based", I referred to theories that, at their core, argue for the claim that an institutional group action consists of (and consequently can be reduced to) the individual contributory actions of its members, who *act in their assigned* roles, or *qua role-occupancy*. To say that members of an institutional group act in their assigned roles, is to say that they perform the functions and tasks definitive of the roles they occupy.

Chapter 3.1. started with the concept of institutional groups and institutional roles employed by such accounts. The upshot of this section was that institutional groups are best viewed as interrelated structures of institutional roles, which in turn are defined through tasks and functions, that an individual role-occupant must perform. In order to theoretically underwrite how institutional groups may come to establish such complex structures, I examined Ludwig's SHARED PLAN ACCOUNT of collective intentions and his related concepts of *constitute rules*, *collective acceptance* and *agent status functions*. To this end, Ludwig argues -pace Searle- that collective acceptance can provide an adequate basis for institutional roles that does not rely on the notion of mutual belief.

In Chapter 3.2., I investigated how the concept of institutional roles figures in explaining institutional agency. This required us to overcome two problems that reductive explanations of institutional group agency encounter: The problem of *Action Integration* and that of *Diachronic Group Constitution*. I argued that role-based explanations can resolve these problems by highlighting critical features of institutional roles in relation to institutional agency: First, institutional roles provide the individual with so called *role-*

based reasons for action. Second, institutional roles *specialize* the actions of group members, leading to a differentiation of tasks and a division of labour. Third, the actions of individual role-occupants are functionally integrated into a *layered structure*. Finally, institutional roles allow for *representative*, or *proxy action*. With such a role-based explanation of institutional agency at hand, I turned to the question whether role-based explanations can capture the *anonymity* and *compartmentalization* of institutional action in Chapter 3.3. There, I discussed a skeptical challenge posed by Katherine Ritchie, according to which anonymous cooperation based on institutional roles must require that the participants involved all know what the goal of the group is, that they are contributing to by performing the tasks of their roles. In order to overcome this challenge, I drew on Seumas Miller's "Collective End Theory" (CET) of joint action and argued that role-performances of individuals can be *functionally integrated* to achieve an end, without common knowledge being involved in this process. It is on this basis, that cooperation within institutional roles can occur *anonymously*. I argued that such functional integration is possible, because institutional roles are *action-specific* but *agent-ambiguous*.

Let me finish this section with a reflection on those cases that my analysis especially wants to target, i.e., cases like the Calutron Girls. Cases like this are especially interesting, because they include not only an unequal distribution of power within institutional groups, the establishment of hierarchies and the specialization and division of tasks. They also pick out instances of individuals working together on heavily compartmentalized group actions, which no individual member alone can neither comprehend, nor execute. Further, it includes scenarios where individuals have to coordinate on such complex activities while potentially remaining anonymous to each other, and being in partial or complete ignorance about the institutional group's overall action.

Recall that the Calutron Girls' work contributed to the Manhattan Project, which was realized primarily (though not exclusively) by one of the world's biggest public engineering agencies, the United States Army Corps of Engineers (USACE), a sub-division of the United States Army. There were about 10.000 Calutron Girls, all of which operated the so-called Calutron Machines in order to enrich uranium. The Calutron Girls' actions were a necessary part of building the first atomic bomb in order to defeat the Axis powers, yet the individuals were unaware of the overall goal they were contributing to. With more than 150.000 people involved, and extending from 1942 to 1945, building the first atomic bomb was an institutional group action of almost unmatched breadth and scale.

Let us take a brief exercise in both empathy and imagination, and picture us working as a Calutron Girl. Each morning, one arrives at the place of work to operate machines with unknown symbols and incomprehensible knobs and gauges. Next to one's own machine is another one, operated by an individual which one might or might not know the (real) name of, but who seems to be engaged in the same task as one is. One's direct neighbors in the machine room vary, and one is often assigned to work on a different machine in another room or building. Each day, after operating the mysterious machines for a while, another individual passes by to do something in the backroom of the machines. That individual might be the same as the day before, but their shifts seem to follow an incomprehensible pattern. Regularly, one is asked to report to a superior, who gives new orders on how to operate the machine, or on which machine one is to operate. These superiors, too, change, as do the questions they ask and the directives they give. At the end of every month, the paycheck arrives and so one continues to show up to work. The degree of

secrecy, anonymity, and compartmentalization that the Calutron Girls were confronted with is, still to this day, infamous. And from the perspective of an individual Calutron Girl, the realization of the whole project must have seemed astonishing.

But a role-based explanation of institutional group action can give us insight into how the action of building the first atomic bomb could have come about. The upshot of what we have established so far is the following: Institutional groups are structures of interrelated institutional roles. Institutional roles can be thought of as collectively accepted status functions which are assigned to individual agents, where the collective acceptance can be cashed out in terms of individuals having Ludwigian we-intentions aimed at realizing a shared plan. Ludwig's central claim here was that individuals being assigned status roles amount to enough members of the group being prepared to *treat* these individuals *as* role-occupants if they engage in the relevant types of activities, which govern the institutional roles.

For the individual Calutron Girl, this can simply amount to joining the institutional group's arrangements by accepting one's role in the institutional group structure, which remains mostly unknown to the individual. As we saw with Ludwig, while such arrangements can be complex and individuals might only have a vague understanding of all of the institutional arrangements, "one doesn't have to know all the details of an institutional arrangement to agree to them" (Ludwig 2017b, 199). Sticking to Ludwig's account, we can also see how strongly shared collective intentional states don't creep in from the backdoor into our individualistic account. If collective acceptance was to comprise an irreducible notion of "we", a "sense of us", a "we-mode" etc., we would again be faced with the question of how the Calutron Girls would be able to *collectively* accept the institutional arrangements while being unaware of each other's existence in the first place. Still, if a Calutron Girl accepted the institutional arrangements of the group, e.g., by voluntarily joining it, this can be regarded as an *indirect* form of authorization and acceptance on her part. Partial or complete knowledge of the institutional arrangements then, including knowledge about the specialization and division of roles and tasks as well as which specific individuals fill the roles, is not necessary for a Calutron Girl in order to agree to the arrangements that she is now contributing to by performing the functions of her role.

With regard to institutional group *agency*, I tried to give reasons that led us to see institutional agency as being constituted by (and reducible to) individuals acting in institutional roles. These institutional roles involve the performance of assigned tasks and functions, based on the exercise of deontic powers and which can be expressed by the notion of acting *qua role-occupancy*, or *as a role-occupant*. To say that an institutional group acted is to say that the members of such a group acted according to their roles. This is made possible by several features of acting in an institutional role. The notion of acting in institutional roles allows us to think of institutional actions to be *specialized*, *layered*, and performed by *proxy*. Institutional roles *are action-specific* but *agent-ambiguous*, meaning that cooperation between individuals can occur on the basis of their roles, without these individuals knowing each other, or having mutual beliefs about each others mental states. Although they themselves were probably unaware of it, the institutional roles of the Calutron Girls were functionally integrated to achieve a common goal. On this basis, they could cooperate *anonymously* with one-another by simply doing their tasks. This is key for understanding how a individual Calutron Girl contributed to the overall group action: By doing "her part", i.e., by fulfilling the tasks and functions of her individual role, she realized her specialized sub-actions which - by being integrated into a

layered structure - was necessary for realizing the overall group action of building the bomb. The institutional role of the Calutron Girls was *action-specific* but *agent-ambiguous*. Each worker, e.g., through breaching security protocol, could be replaced by another individual occupying this role. Yet, the outcome, i.e., the performed actions of each replaceable worker, remained consistent.

Cases like those of the Calutron Girls are important, not only because they reveal something about the nature of institutional group agency. They also highlight an essential and ubiquitous feature of the way in which institutional groups tend to organize themselves in terms of the relations between individuals and their surroundings. The case of the Calutron Girls is a prime example of what is often called the phenomenon of alienation, or *Entfremdung* (See for different discussions of alienation: Bratman 2022; Shapiro 2014; Ludwig 2017b). Being early scholars of alienation, Marx and Engels stressed that being alienated tends to go along with feeling powerless, subordinated, and that it is accompanied by the subjective experience of being worthless and replaceable. For Marx and Engels, the alienated worker is deprived of any meaningful relation to the work that she executed. She suffers under the anonymity, interchangeability, and pettiness of her own position, and is degraded to a mere "appendage of the machine" (Marx and Engels 1948/1955, 16). The alienated work of the proletarians, they wrote in their *Communist Manifesto*, "has lost all individual character, and consequently all charm for the workman" so that it is "only the most simple, most monotonous, and most easily acquired knack, that is required of him" (ibid). It has long since been established that being "just a cog in the machine" can deeply impact an individual's life, particularly physical and mental health (see, e.g., Hochschild 1983; Adibifar & Monson 2020). Alienation abets the objectification by, and of other individuals. It can thereby undermine individual autonomy. And without a doubt, working as a Calutron Girl must have been a strange and demanding thing to live through.

But alienation is janus-faced. On a positive spin, the alienation of individual agents seems to correlate with an enhancement of agency on behalf of institutional groups. Scott Shapiro, e.g., claims that large-scale actions of institutional groups and the phenomenon of alienation "usually go hand in hand" (Shapiro 2014, 4). But I do not want to simply argue that institutions yield great power because of alienated members, nor that institutional power is *per se* a good (or bad) thing. This might depend on factors which I do not speak to. Rather, the alienation of group members, understood as the loss of individual influence, makes institutional actions - at least in principle - less prone to chaos and capriciousness. Dividing and compartmentalizing work, specializing tasks, representing (or being represented) by proxy allow for institutional action to be pre-planned, regulated, steadfast and subject of oversight. One can retrace and comprehend the interdependent and nested steps of institutional action, and what is expected of individuals becomes manageable and predictable. Results can be calculated and "smart" processes can cancel out the imperfections and stupidities of fallible individuals. So the division of labour and role-specialization of tasks shields the actions of institutional groups - at least to some degree - from individual contingencies. Roles moderate the influence that individuals, with their personal peculiarities and selfish ends, have on institutional actions. Or they can at least mitigate these influences and make them less arbitrary and random. Thus, roles also shield institutional actions from being capricious and even despotic due to the erratic behavior of single individuals. In healthy democracies, for example, roles and offices and are - at least in principle - open for everyone to occupy. No one individual is to govern positions of power

just because of her family, heritage, race, sexuality etc. Institutional organization can thus counter cults of personality.

This page was intentionally left blank

4. A Critique of Role-based Theories

Today it is almost heresy to suggest that scientific knowledge is not the sum of all knowledge. But a little reflection will show that there is beyond question a body of very important but unorganized knowledge which cannot possibly be called scientific in the sense of knowledge of general rules: the knowledge of the particular circumstances of time and place […] We need to remember only how much we have to learn in any occupation after we have completed our theoretical training, how big a part of our working life we spend learning particular jobs, and how valuable an asset in all walks of life is knowledge of people, of local conditions, and special circumstances.

Friedrich August von Hayek: The Use of Knowledge in Society

In this chapter, I will defend the third claim of my overall argument, i.e., that (III) an individual's performance of an assigned institutional role cannot be *sufficiently* explained by the individuals acting on tasks and functions definitive of their institutional roles. But let us take one step back and look at the broader image. On a more general level of abstraction, we find that the debate so far has put strong emphasis on the relation between institutional groups on the one hand, and institutional roles on the other. But if we think about what *hasn't* been paid attention to, we find that the debate about institutional group agency involves in fact a *tripartite* relation. It involves not only the relation between institutional groups and institutional roles but also the *relation between institutional roles and the individual agents who occupy these roles* (see: Box 2).

Institutional Groups ↔ Institutional Roles ↔ Individual Role-Occupants

Box 2: The tripartite relation of institutional group agency

This relation between institutional roles and the individuals who occupy them, I think, is underdeveloped by the existing literature on institutional agency. I will now try to show how this relation has implications for our understanding of a role-based explanation of institutional group agency. To draw out these implications, Chapter 4.1. will give reasons for why role-based explanations must eventually point to an individual's capacity for *discretion*. Here, I will argue that the tasks and functions of institutional roles cannot exhaustively prescribe an individual role-occupant's conduct. I will then try to show how role-occupants might come to face situations where their tasks and functions do not specify a particular course of action. In such situations, individual role-occupants have to "make up their own mind" on how to proceed. This, I will argue, can be modeled in terms of an individual's capacity for *discretion*, or in terms of so called *discretionary powers*. Although this capacity for discretion is - at least to some extent - acknowledged by the existing literature, its full implications have not been drawn out yet. What I ultimately aim to show is, how the discretionary interpretation, adaptation, and performance of institutional roles all point towards a

problem that the existing literature fails to make sense of. Accordingly, Ch. 4.2. will argue that both the individual's *suspension and use* of discretionary powers can be detrimental to, and even lead to a breakdown of an institutional group's capacity for action. These cases, I think, cannot be accommodated by the existing role-based approaches to institutional agency. So we have to move beyond those approaches in order to explain what is going. My own account of Role Agency, which I will present in the fifth chapter, is to be understood as such a move beyond the established view of institutional roles.

4.1. Agent-ambiguity and Discretionary Powers

As I argued in the third chapter, the outcome of an institutional group can remain consistent throughout the replacement of some, or even all of an institutional group's role-occupants. It doesn't seem to matter that much which *particular* individual comes to occupy an institutional role, at least to the extent that this individual fulfills certain requirements. What actually matters isn't *who* performs the tasks, but rather *which tasks are being performed*. Recall, for example, the above mentioned painter who is contracted to paint the front side of a house. For this painter, it doesn't really matter whether it's *Anne* or *Betty* who paints the other side of the house. What matters to her is that the job gets done. Alternatively, imagine a Calutron Girl operating her assigned machine. To the individual Calutron Girl, it doesn't matter which particular individual passes by to do something in the backroom of the machine that she is operating. What matters is that the workflow remains undisrupted. In institutional group action, so it seems, every role-occupant can be replaced, at least to the extent that the replacing individual meets certain membership-conditions (e.g. having certain educational degrees signaling competence, or being motivated to carry out the tasks to a reasonable degree, etc). On this view of institutional agency, no one is *irreplaceable*. And thinking about institutional groups as structures of differentiated, *agent-ambiguous* roles that can be occupied by different individuals leads to the impression that the individuals tend to "disappear" behind their role-performances. This way of framing the agent-ambiguity of institutional roles is intriguing. However, I also hold it to be lopsided. Why?

In order to theoretically motivate my line of reasoning, let us - again - reflect on some critical features of institutional roles. As mentioned, roles are *agent-ambiguous* in the sense of being *interchangeable* (or *transferable*) and *iterable*. Regarding *interchangeability*, this feature of institutional roles lies not on the fringes, but at the very heart of a role-based explanation of institutional agency. According to Ludwig, an "organized differentiation of roles directed toward joint action, *which may be occupied successively by distinct individuals*" is the "hallmark" of institutional groups (Ludwig 2017b, 2) [own emphasis]. Roles are further taken to be *iterable*: One and the same institutional role-*type* can be multiply realized and occupied by different individuals at the same time.

But agent-ambiguity does not only imply *non-specificity* about who will be performing the constitutive actions of a given role. Crucially, it also implies that institutional roles cannot be identified with, or defined through any *particular* individual occupying the role. Roles being agent-ambiguous in *this* sense comes with them being subject to what we might call *impersonality,* or *depersonalization*. They are *standardized,* or "run-of-the-mill" placeholders for activities, rather than bespoken or custom-tailored for *any particular* individual. Individuals, so it seems, have to accommodate to the roles they occupy, and not the other way

around. So roles are *general placeholders for activities* which different and multiple individuals (i.e., individual persons) can occupy. As such, they do not take into account any particular idiosyncrasies (e.g., personal traits) of the individuals occupying them.

But this *non-specificity* (understood as impersonality) of institutional roles, I will now argue, comes with a price. As institutional roles are non-specific *in this sense*, they can only give any particular role-occupant *generic instructions* or *generalized directives* on what to do.[97] They do not, however, specify how to execute the tasks and functions *in situ*, i.e., given *a certain context*, *changing circumstances* or *in a specific situation*. Seumas Miller, e.g., states that "[a]lthough the structure, function, and culture of an institution provide a framework within which individuals act, they do not fully determine the actions of individuals" (Miller 2010, 91).[98] Likewise, sociologist Niklas Luhmann stated that formalized expectations (i.e., the formally established tasks and functions and their corresponding rights and duties) of role-occupants are always just *tendency expectations* (Luhmann 1964, 311).

[97] Consider, for example, the inherent vagueness and open-endedness of job-descriptions for fire-fighters. According to the United Stated Department of Labor's *Occupational Outlook Handbook* (OOH), which lists information about the nature of working conditions for hundreds of different occupations in the U.S., firefighters "control and put out fires and respond to emergencies involving life, property, or the environment" (United States Department of Labor 2023). These tasks correspond with a fire-fighters duties to "respond to emergencies, drive firetrucks and other emergency vehicles, put out fires using water hoses, fire extinguishers, and water pumps, find and rescue occupants of burning buildings or other emergency situations" or to "treat sick or injured people" (ibid). The handbook further includes statement like: "When firefighters are not actively responding to an emergency, they often participate in other activities related to their work. For example, they must maintain a high level of physical fitness" (ibid).

[98] I will examine Miller's use of the word "institutional culture" below.

But why is that? At its core, the underlying problem is an old philosophical chestnut concerning the nature of rules.[99'100'101] But for brevities sake, I will focus on only two aspects of this: First, the formally defined tasks and functions of institutional roles have to be *interpreted* and *applied* in order to match concrete circumstances, as they cannot exhaustively cover every possible contingency that might arise (cf. Miller 2010, 91f.). The constitutive rules as well as the tasks and functions governing institutional roles do not - and may in principle not - fully determine every possible course of action. Theorizing about role-dependent actions *in vacuo* is something fundamentally different than actually performing such actions *in situ*. Also, different tasks and functions may not be realizable at the same time, making it necessary that these tasks are to be weighed against each other, so that some functions and tasks might be prioritized other others.

[99] Recall the debates in the philosophy of language regarding the concept of rule-following. Here, Wittgenstein (PU, §§201-208) famously argued that the grammatical and semantical rules that govern language cannot - in and of themselves - generate meaning, which he locates in the *application* or *use* of language in a given community (for a discussion see: Miller & Sultanescu 2022). It is still contested which exact conclusions is to be drawn from his arguments. What is less contested is that Wittgenstein wanted to get away with the picture "that following a rule – a rule for the use of a word, say – is a matter of traveling along rails which are already laid down and determine its application in new cases, and so on" (Hale 2017, 619).

[100] Another way to explain this underlying tension between explicitly stated rules and their inherent vagueness is to look at the problem of contextuality of meaning in the philosophy of language. Searle (1978) here notices that language possess an inherent ambiguity regarding the *literal* meaning of sentences. Literal utterances, Searle notes, can always be interpreted in a certain way, and their interpretation can be detrimental to what the speaker tried to convey with them. He gives the example of someone ordering a hamburger by saying "Give me a hamburger, medium rare, with ketchup and mustard, but easy on the relish" (126). Now if this person was, as a result of said speech act, to receive a hamburger encased in "cubic yard of solid lucite plastic so rigid that it takes a jack hammer to bust it open" (ibid), we wouldn't say that her order was fulfilled. Because clearly, that is not what she *meant*. Searle then concludes that the literal meaning of sentences is only intelligible "against a set of background assumptions about the contexts in which the sentence could be appropriately uttered" (Searle 1978, 117). I thank Kirk Ludwig for pointing me to this example.

[101] Yet another discussion of this can be found in the work of H.L.A. Hart (1994). Regarding legal rules and their textualisation in bodies of law, Hart argues for rules to be inherently vague or indeterminate regarding the exact possibility of their application (see especially 1994, Ch. 7): "All rules involve recognizing or classifying particular cases as instances of general terms, and in the case of everything which we are prepared to call a rule it is possible to distinguish clear central cases, where it certainly applies and others where there are reasons for both asserting and denying that it applies. Nothing can eliminate this duality of a core of certainty and a penumbra of doubt when we are engaged in bringing particular situations under general rules. This imparts to all rules a fringe of vagueness or 'open texture`" (Hart 1994, 123). Hart's much discussed example is the simple rule "No vehicles in the park" (ibid). Now someone, who is given the task to enforce this rule may face certain situations in which she is unsure whether this rule applies to a given case or not, e.g., in the case of an electrically propelled toy motor-car (ibid). In the legal realm, such "hard cases", i.e., cases in which "no settled rule dictates a decision either way" (Dworkin 1975, 1060) require the use of *judicial discretion:* "There will always be certain legally unregulated cases in which on some point no decision either way is dictated by the law and the law is accordingly partly indeterminate or incomplete. If in such cases the judge is to reach a decision […] he must exercise his discretion and make law for the case instead of merely applying already pre-existing settled law. So in such legally unprovided-for or unregulated cases the judge both makes new law and applies the established law which both confers and constrains his law-making powers" (Hart 1994, 272). For further elaboration see Bix (1991). I want to thank Michael Bratman for pointing me to this example.

Further, different tasks and functions of one and the same institutional role may stand in conflict with each other, so that there may be trade-off scenarios between their fulfillment.[102]

So focussing solely on the tasks and functions of role-occupants does not tell us how individuals function within their status role when the exercise of those tasks and functions is i) conflicting or ii) not sufficient to deal with new situations emerging from changes in the environment or iii) up for interpretation of the individual due to indeterminacy. Let me discuss these -interrelated- aspects by highlighting three exemplificatory cases (CONFLICT, NOVELTY, and INDETERMINACY) where the formally established tasks and functions of an institutional role are inapt to allow an individual to successfully function within her institutional role.

The first case described by CONFLICT arises in situations where: 1) occupying a role provides the individual with tasks and functions (including rights *R* and duties *D*) regarding certain types of activities in 2) a situation *S* where the exercise of one of those tasks and/or functions (including rights *R* and duties *D*) stands in direct conflict with the exercise of some *other* tasks and/or functions (including rights *R* and duties *D*) of the role-occupant, where 3) this conflict cannot be resolved or mediated by relying on hierarchy, authority, or application of a meta-script.[103] Consider the following, fictional case:

> **CONFLICT**: Antigone is an experienced manager in the southern department of a big corporation. She has two bosses, Kreon (from HQ) and Polly (from Branch) that assign work for her. One day, Antigone finds on her desk two separate assignments that both her bosses independently want to be completed by the end of the week. However, until Friday, she can only complete one of the two tasks. Antigone decides to fulfill Polly's task instead of Kreon's. On Friday, she therefore gets in trouble with Kreon.

The second problem of NOVELTY arises in situations where: 1) taking up a role provides the bearer with tasks and functions (including rights *R* and duties *D*) regarding certain types of activities and 2) the role-occupant has to respond to an unprecedented situation *S* in which 3) the tasks and functions (including the rights *R* and duties *D* of her role) that formally define the role-occupancy are inapt to deal with this unprecedented situation in which the tasks and functions must be performed, where 4) this situation cannot be resolved by relying on hierarchy, authority, or application of a meta-script.

> **NOVELTY**: Jen is the spokeswomen for her government and is about to give her weekly press briefing. Jen has had this job for a long time and generally knows what to do. Minutes before

[102] This might ultimately be based on the fact that institutional groups face contradictory demands from the larger environment, in which they are embedded in. Kühl (2022, 32f.) argues that *owners, beneficiaries, members* and the *political environment* all place different, and possibly contradictory demands on institutional groups.

[103] A meta-script would amount to what Shapiro (2014) calls a *pre-designed cooperation plan*. See for a discussion of such "mesh-creating mechanisms" in the context of Batman's Shared Intentional Action: Shapiro 2014.

the briefing, Jen learns that there just has been an *alien invasion*[104] in her country. This has never happened before and Jen does not know the official stance of the government towards alien invasions. She does not even know whether her government in fact *does* have stance towards alien invasions. Jen proceeds to give the press briefing, trying her best to answer the questions of the alarmed journalists.

Finally, the problem of INDETERMINACY arises in situations where: 1) taking up a role provides the bearer with tasks and functions (including the rights *R* and duties *D* of her role) regarding certain types of activities and 2) in situation *S,* the mere possession of those tasks and functions (including rights *R* and duties *D*) only gives insufficient reasons to act in one way or the other, where 3) those multiple courses of action are equally possible and therefore only chosen due to private preferences or dispositions of the role-occupant.

INDETERMINACY: Anne just started as a lecturer for her local philosophy department. Anne has never been a lecturer before, although she has her own, private ideal of what (good and bad) teaching in university amounts to. Anne is told to teach a course in 20th-century philosophy of language, which is her field of expertise. She further knows about the formal requirements of this job, such as her duty to follow the curriculum, to take a test at the end of semester, etc. However, Anne does not know how she will go on about teaching: Whether she will be strict or laissez-faire, formal or informal with her students, which methods she will choose, etc. Seeking advice on what to do, Anne's Boss reassures her that she will do just fine. Eventually, Anne proceeds to teach the course the way she personally sees fit.

How can we make sense of these problems - or a mixture thereof - as well as the possible (fictionalized) ways the role-occupants respond to them?[105] The prevailing answer given by role-based theories of institutional action to such cases lies in the use of *discretion* of role-occupants to interpret and adapt their tasks and functions, as well as the corresponding deontic powers, in order to be applicable to novel situations and contingent circumstances *in situ*. How could we generally characterize the concept of discretion? Gilligan (1990) claims that the "central feature of discretion is a degree of autonomy, within a defined context, vested in the decision-maker" (Gilligan 1990, 6f):

[104] This is of course somewhat silly. But all I am presuming here is that this is a situation, that her government does not actually has a stance towards which would authorize her to respond within her capacity for representative agency. One may pick alternative situations to which this applies which do not involve extraterrestrials.

[105] This question relates to a small criticism regarding Lackey's theory of group knowledge. In talking about the capacity of groups to assert things, she posits that there is a special relationship that exists between a individual members and their groups that members act as proxy agents for (Lackey 2021, 181-189). A spokesperson, as Lackey argues, can thus make assertions *qua her role* that are not her own but that of her party (see Ch.4.2-4.3.). But similar to the other accounts that I will criticize below, Lackey ultimately doesn't really offer a sufficient description of the relation that holds between individuals and their roles which allow them to actually achieve this. I want to thank Katja Crone for pointing me towards this aspect.

> "Discretionary power is often characterized in terms of the authority to choose amongst alternative courses of action. So the paradigm of discretion is the power-holder faced with a choice between actions X, Y and Z; his discretion is said to be freedom of choice amongst those actions" (ibid).

Writing about agential discretion in the realm of law, Bennion (2009) characterizes discretion by a "looseness of outcome":

> "For an enactment to bestow a discretion on a person (D) involves a built-in looseness of outcome. In reaching a decision, D is not required to assume there is only one right answer. On the contrary D is given a choice dependent to a greater or lesser extent on personal inclination and preference" (Bennion 2009, 132).

Bennion further distinguishes between discretion being *open*, i.e., courses of action being "completely at large" (ibid) and discretion being *confined* to certain courses of action "within limits laid down expressly or by implication" (ibid).

In the context of institutional agency, Ludwig - briefly - points out that institutional roles allow individuals "do certain things in joint activities […] at their discretion" (Ludwig 2017b, 139; also: 2020a, 186; see for a discussion of *status role autonomy* in the case of proxy agency: Ludwig 2018a, 318ff.).

Miller (see especially: Miller 1998; also: Miller 2001; 2013; Miller & Blackler 2005) provides a more comprehensive explanation of what he calls "discretionary powers" of institutional roles. According to Miller, institutional roles are necessarily accompanied by so called *discretionary powers*, where, on a basic level, discretion can be defined as the authority to choose amongst alternative courses of action. He states that:

> "changing circumstances and unforeseeable problems make it desirable to vest individuals with discretionary powers to rethink and adjust old rules, norms, and ends, and sometimes elaborate new ones. Inevitably the individuals who occupy institutional roles are possessed of varying degrees of discretionary power in relation to their actions. These discretionary powers are of different kinds and operate at different levels. For example, senior- and middle-level public servants have discretion in the way they implement policies, in their allocations of priorities and resources, and in the methods and criteria of evaluation of programs. Indeed, senior public servants often exercise discretion in relation to the formulation of policies […] Lower-echelon public servants also have discretionary powers. Police officers have to interpret rules and regulations, customs officers have the discretionary power to stop and search one passenger rather than another, and so on. Traditionally, members of the so-called professions, such as doctors, lawyers, members of the clergy, engineers, and academics, have enjoyed a very high degree of individual autonomy, notwithstanding their membership in, and regulation by, professional associations. In recent times they have increasingly been housed in large, bureaucratic, hierarchical organizations in which their professional autonomy has evidently diminished somewhat […] Indeed, the working population more generally is increasingly

> employed by large, bureaucratic, hierarchical organizations, whether corporations or public-sector organizations - so much so, that arguably the central threat to individual autonomy in modern societies is no longer governments but rather corporations and nongovernment public-sector organizations […] Certain categories of individual institutional actors have discretionary powers and a reasonable degree of autonomy in the exercise of their institutional duties" (Miller 2010, 91f).

Discretionary powers, as Miller notes, come in varying degrees and can be exercised on different levels. To stay within the established taxonomy of member-to-action and member-to-member relations of power, discretionary powers regarding the member-to-*action* relations of power can include *altering*, *re-interpreting*, *changing*, *reformulating*, *permanently omitting* or *temporarily suspending* formally defined tasks and functions of an institutional role, laid down in its design-specification. Again, Miller gives some examples for this in the case of individuals occupying the institutional roles of policewomen:

> "There is a need for the exercise of discretion by police in the interpretation and application of the law. Notoriously, police can (lawfully) choose not to enforce some laws in some circumstances, e.g. they can issue a caution in relation to a minor infringement, if they judge the outcome of an arrest to be deleterious to the peace […] Second, the law does not, and cannot, exhaustively prescribe. Often it grants discretionary powers, or has recourse to open-ended notions such as that of the ‚reasonable man' or ‚reasonable suspicion'. Accordingly, a number of police responses might be possible in a given situation, and all of them might be consistent with the law. Third, upholding and enforcing the law is only one of the ends of policing, others include the maintaining of social calm and preservation of life. When these various ends come into conflict, there is a need for the exercise of police discretion" (Miller & Blackler 2005, 44).

In turn, discretionary powers regarding the *member-to-member* relations of power can be used to emphasize, or to de-emphasize certain hierarchical relations within formally established hierarchies (e.g, a boss choosing to work with one particular employee on a project because she holds her to be the most apt of her peers; yet choosing another employee for another project etc.), or to establish, suspend, arrange or re-arrange hierarchical relations between members altogether (e.g., by deciding to report something directly to another role-occupant instead of obeying one's primary duty to use the established lines of communication).

So generally, discretionary powers are vested in individual role-occupants in order for them to interpret and apply their institutional roles to concrete contexts. So - at least to some degree - individual role-occupants have to "make up their own mind" on how to perform the tasks and functions of their roles, where this can be explained by the use of discretionary powers that are vested in the institutional roles they occupy.

But if we accept the claim that, e.g., changing circumstances, unforeseen situations or ever-changing environments necessitate the use of discretionary powers, we come to face a predicament which the existing literature leaves unaddressed. One the one hand, I argued for institutional roles to be *agent-*

ambiguous, and that it does not really matter *who* is occupying an institutional role but rather *what the role-occupant does*. But on the other hand, I just argued that it is simply not the case that an individual completely "disappears" behind the role she is occupying, because institutional roles cannot exhaustively prescribe, or completely determine the behavior of the individuals who occupy them. So in some sense, the individual role-occupant really *does* matter.

In the next section, I will further analyze this tension, and argue that discretionary powers point to a challenge for role-based explanations of institutional agency. This is because both *the suspension* and *the use* of discretionary powers can lead to contributory actions of individual members that are inadequate to realize an overarching institutional group action. So the next section will present what I call the *two Problems of Discretion*.

The overall goal of posing these two problems is to motivate the need for a better understanding of just *how exactly* individuals are able to use their roles' discretionary powers. And this need to go beyond the established view of institutional roles stems out of the underdeveloped, or neglected relation between individual role-occupants and their assigned institutional roles. In the fifth chapter, I will thus try to argue that my own account of *Role Agency* offers a deepened understanding of this relation.

4.2. The Two Problems of Discretion

> **First Problem**: If a role-occupant does *not* make use of so called discretionary powers, this leads to the "work-to-rule" performance of her institutional role and work-to-rule performance can lead to a breakdown in an institutional group's capacity for action.
>
> **Second Problem**: If a role-occupant *does* make use of these discretionary powers, this may lead to a *performance-gap*, or to what I call the *divergence* of role-performance, and this performance gap can lead to a breakdown in an institutional group's capacity for action as well.

The first problem pertains to what I call the phenomenon of *institutional stupor*. To this end, I will show that a group's capacity for action can break down, i.e., that institutional groups can (partially or fully) lose their capacity for agency, although, or especially *because* all formal aspects of the institutional roles are being fulfilled to the letter. Empirical studies of so called *work to rule strikes* (see, e.g., Scott 1998; Bloch & Moorman 1993; Kühl 2022) show, seemingly paradoxically, that if individuals perform *exactly* what their roles require them to do, institutional groups stop to function. Here, institutional groups fail to function because individuals fail to use their discretionary powers, i.e. because they fail to adapt and interpret their assigned roles to situations where instructions are insufficient, circumstances change, or demands for spontaneous decision-making arise.

The second problem concerns what I call the phenomenon of *role-ambivalence*. Here, I will argue that an individual's functioning within a role can be partially impaired, or even made impossible, because the discretionary interpretation and adaption of her role turns *toxic.* Role-performances turn *toxic* if they actually thwart, circumvent, or impede the realization of the institutional group's goal to which the role-occupant is supposed to contribute to. Interestingly, this can happen in cases where - *technically speaking* -

the role-occupant fulfills the tasks and functions of her role and where she doesn't formally lack either the competence or commitment to do so. It then seems that the discretionary adaptation and interpretation of one's assigned role can be the cause of an institutional group's breakdown in agency too.

When dealing with discretionary powers, role-occupants may come to face a predicament: If there is too much "leeway", i.e. too strong divergence in the interpretation of the institutional role on behalf of the individual occupying it, then the agency of institutional groups can be threatened or jeopardized. However, if there *isn't* such a form of interpretation, the agency of institutional groups can be threatened or jeopardized too. The nature of this problem then is that with discretionary powers, there is the risk of institutional agency failing, but without it, this risk also seems to become more likely.[106]

First problem: institutional stupor

What's the reasoning behind the first *Problem of Discretion*? My main concern here is that institutional groups can come to exhibit *institutional stupor* if role-occupants don't make use of the discretionary powers of their assigned roles. By *institutional stupor* I mean the partial of complete breakdown of institutional agency due to the rigidity and inflexibility of an institutional group's structural set-up, or design.[107] As I argued above, institutional agency is based on the institutional group's members fulfilling the tasks and functions of their assigned roles. However, the design specifications of these institutional roles also run the risk of being agency-*impeding*, instead of agency-enabling or -enhancing. *Institutional stupor* occurs in situations where the design-specifications are inadequate to deal with specific circumstances that role-occupants may face. If, in such situations, role-occupants stick to their roles' design specifications, this may lead them to become "paralyzed" in fulfilling their tasks and functions.

An interesting case for the phenomenon of *institutional stupor* can be made by the phenomenon of "working to rule" or so called "work-to-rule" strikes (German: "Dienst nach Vorschrift"; French: "grève du zèle"). An institutional role-occupant can be said to be *working to rule,* if she adheres strictly to the prescribed functions and tasks of her assigned role (i.e., if she adheres strictly to the role's design specifications) and meticulously follows the entailed rules governing the performance of her tasks and functions (including the directions of her superiors) "to the letter". Nothing more but also nothing less is required for an individual to "work-to-rule".[108] Sociologist Stefan Kühl describes the phenomenon of working to rule as follows:

---

[106] One could say that discretionary powers lead to a *dilemma* concerning the use of discretionary powers. As this may not, however, match the formal definition of a dilemma, I reside with the less controversial option to call it a predicament. I want to thank Eva Schmidt for pointing this out.

[107] I am borrowing this term from the medical field, where stupor describes a *disorder of action*, which can be characterized both by "a reduction in or absence of relational functions (i.e. action and speech)" (Berrios 1981, 677f.) as well as an impaired reaction to external stimuli, i.e. an "excessively deep state of unresponsiveness" to the environment. (MSD Manual Consumer 2022).

[108] Work-to-rule *strikes* can be regarded as a collective and coordinated way of individuals *working to rule together*, often tactically orchestrated by labour unions as "on-site" strategies in order to accomplish or promote their goals.

> "It is not for nothing that ‚working to rule' is considered the most effective form of labor strike for paralyzing an organization. When all rules and instructions are followed to the letter, even a well-planned organization will grind to a halt. The organization will be broken by the rigidity of its formal structures and done in by its mania for order and ordinances, its frenzy for regulations, and its fetish for rules […]. Anyone who doubts this could conduct a kind of breaching experiment and spend several days doing only precisely what their organization requires. This would largely bring the work process to a halt. The person in question would be written off as a ‚bureaucratic virtuoso' who could never let a single rule go, an ‚insistent formalist' incapable of occasionally ‚letting things slide,' or a ‚rule-obsessed nitpicker' who doesn't know how organizations actually function. Co-workers and superiors would increasingly pressure the individual not to overdo their ‚bureaucratism,' or their strict adherence to the formal order" (Kühl 2022, 12f).

An empirical case study of a work-to-rule strike is given by Bloch & Moorman of a plant of the construction equipment producer *Caterpillar* in 1992. Back then, the labour union "United Auto Workers" (UAW) orchestrated its members' work-to-rule performances in order to promote their demands for higher wages:

> "The main thrust of the UAW's in-plant strategy, at least according to published reports, was working to rule (i.e., adhering strictly to job descriptions and work rules, and following supervisors' directions to the letter) - no less and no more. As UAW spokesman Jim O'Connor put it, the ‚effect of doing everything by rule is chaos' […] ‚[W]e're going by the book, and that just takes longer.' He proffered some other examples of work-to-rule tactics that employees might utilize, including ‚standing around rather than helping engineers, or calling in repair people instead of fixing small problems themselves'" (Bloch & Moorman 1993, 177f.).

It was estimated that the UAW's work to rule strike was highly effective and that the work to rule "slowdowns" reduced Caterpillar's production rate (i.e., its measurable output of production) by more than 40 percent (ibid). Another exemplary case of such a work-to-rule-strike is provided by Scott in the case of Parisian Taxi-Drivers:

> "When Parisian taxi drivers want to press a point on the municipal authorities about regulations of fees, they sometimes launch a work-to-rule strike. It consists merely in following meticulously all the regulations in the *Code routier* [the drivers' handbook specifying their tasks and functions, M.G.] and thereby bringing traffic throughout central Paris to a grinding halt" (Scott 1998, 256).

Cases of working to rule should strike us as odd. Didn't the third chapter argue for the merits of explaining institutional group agency through individuals fulfilling the tasks and functions of their roles? Why, all of the sudden, does individuals fulfilling the tasks and functions of their roles *impede* institutional agency? Aren't

these just straightforward *counterexamples* for the view that I am advocating for? Well, no. But to see why this is not the case, let me elaborate why *working to rule* may lead to institutional stupor in the first place.
Here, sociologists and organizational theorists (see, e.g., Alvesson 2015; Bosetzky 2019; Kühl 2007; 2022; Kubbe & Engelbert 2018; Ledeneva 2018; Luhmann 1964) point to the functional use of *informal rules and practices* in institutional groups. Kühl, using the vocabulary of "organizations" instead of "institutional groups", highlights the use informal rules and practices, which also encompass what he calls *pro-social rule-deviances*:

> "In practice, therefore, organization members permanently oscillate between basing their actions on formal rules and informal deviations. They debate whether to put critical information ‚on file,' as formally required, or instead go against the regulations and leave no written trace for the time being. They might reject a verbal request from another department and insist on using official channels instead, or they could cooperate and respond informally even though it violates the official rules of procedure. They can formally discuss a matter with their superior and risk an official refusal, or they can keep the conversation informal so that the issue can be brought up again at a better time […] Organizational prudence, therefore, does not consist of slavishly following externally imposed or internally defined rules, nor does it entail ignoring these rules altogether. Instead, it is the ability to occasionally deviate from the rules. *Ultimately, an organization can only preserve its rules by tolerating a good deal of deviance from them* […] *Rules have to be broken—at least from time to time—in order to remain true to the spirit of the rules* […] *and in order for these rules to continue to exist* […]" (Kühl 2022, 41f.) [own emphasis].

Kühl suggests that institutional groups rely on a double system of 1) their members operating by the formally established tasks and rules of their institutional roles' design specifications, and also 2) on *informal* rules and procedures in the application and interpretation of the formally established design specification. The latter dimension of role-performance sometimes even encompasses the breaking of, or deviance from official rules and protocols. Thus, institutional groups are prone to produce "gray zones" of compliance and non-compliance with the formally established tasks and functions, as well as corresponding deontic rights and duties:

> "There is a certain ‚murkiness' that must be accepted when dealing with the question of whether an activity complies with the rules or not […] Just think of how rules can be broadly interpreted, creatively stretched, skillfully undermined, discreetly ignored or tacitly flouted […] Or think of the practice of ‚following rules based on impermissible motives or impermissible purposes,' or the ‚right action at the wrong time,' or the ‚deferment of compliance' […] Despite all attempts to establish clarity, legal norms are not distinct ‚lines' that must not be crossed, they are ‚zones' in which organizations negotiate what they will and will not tolerate […] *In the factual dimension, one reason for the existence of gray areas is that even the most well-defined formal rule is open to interpretation. Formal expectations can be formulated extremely*

carefully, but the real meaning of actual expectations can never precisely be put into words" (Kühl 2022, 44f.) [own emphasis].

Approximating informal rules

If informal rules govern the actions of individual role-occupants and also effect an institutional group's overall capacity for action: How could we further characterize them?[109] Let me start with a list of examples of informal rules that may govern an individual's performance of her assigned tasks and functions:

- If the problem is minor, fix it yourself instead of calling the official clerk.
- In a process of deliberation, the superior gets to pitch her idea first.
- Don't snitch on your co-workers' failure in front of your boss.
- Whoever comes first into office in the morning makes coffee for everyone.
- Do not conduct meetings on a Friday after 3 p.m.
- When cooperating with an internal department, verbal agreements are sufficient.

As a first approximation, notice that such informal rules will tend to *regulate* role-based activities that existed prior to positing this rule. So for a start, we can characterize such rules to be *regulative*, rather than *constitutive* of an activity. To further assess the concept of informal rules, let us look at the first informal rule of the list. Let's assume that the first example provides an informal rule governing the work of an employee whose task is to operate certain machines in order to produce certain goods. Regarding this activity in general, a worker might be required to follow rules which are constitutive of her task (e.g., operating certain machines for producing these goods but not others; or operating machines in the first place). Without following these constitutive rules, this worker would cease to carry out the functions officially prescribed via the design-specifications of her institutional role. Now, regulative rules don't specify *what activity* the worker is supposed to fulfill, but *how* she should proceed in doing so. And there may be *formal*, or *official* regulative rules, i.e., the regulative rules which are officially established in the design-specifications of her role. These official rules may require the worker to carry out the functions in a specific way, e.g., to call in a repair clerk every time the machine malfunctions. We may, in those cases, say that the worker follows the *official protocol*. Following official protocol is following regulative rules governing an activity type, and the official protocol dictates which *formal* regulative rules are to be followed.
Now the *informal* rule governing her work states that she is to fix minor problems with the machine by herself, *instead of calling in the repair clerk*. This, too, is a case of the worker following a regulative rule

[109] I want to thank the participants of the workshop on *Group Minds and Collective Agency* at Leeds University for fruitful discussion of the nature of informal rules.

governing an activity type, but the regulative rule is not officially recognized by the protocol.[110] In fact, it consists of *breaking official protocol*. So the distinction between regulative and constitutive rules does not seem to sufficiently characterize the (in)formality of these rules because regulative rules can be both formal and informal in nature. Each of the examples above, e.g., could be a formal *or* informal rule. So the question remains open, what criteria could be used to demarcate the *formal* from the *informal* regulative rules governing the role's activity-type.

I think that a tentative way to draw out this criterium can be given in terms of authority- and power-relations. The content of an institutional role's design-specification can be regarded as "official" only if it is being determined through some collectively accepted procedure or policy-process, or if it is being determined by someone collectively authorized to do so, i.e., by someone having a certain status role which grants her the right member-to-member relation of power to issue a judgment regarding the design-specification of other roles. It is in virtue of this power, that such a judgment can come to have the status of an *official* fact about a role's design-specification. A further conjecture here is that the content must in some way be made epistemically transparent and accessible to the individual who occupies the so defined institutional role, i.e., by being conveyed to the role-occupant through some form of authorized means of communication like, e.g., the announcements or publications of a proxy-agent (e.g., through documents, oral or written contracts, or by laws or statues).

Let me advance on the idea that the criterium for distinguishing formal from informal rules resides in authority-according power-relations. Imagine a novice worker who has been assigned the task to produce some goods by operating a certain machines. Suppose that she works next a more experienced colleague, and that this colleague suggests to the worker to operate the machine in a certain way so that it runs more smoothly. In such a case, and because the member-to-member relations of power are equally distributed among the role-occupant and her colleague, the colleague does not seem to have the right kind of *authority to alter the design-specifications of her role*. And while she may show her a more convenient way to operate the machine, her suggestion does not have the status of a *formally* defined rule or command to do so. If, however, the novice role-occupant proceeds in the way in which her colleague suggests, she can be said to follow a regulative rule that governs a (sub)set of her tasks and functions. But she then starts to abide to a rule which may not be part of the *officially* established design-specifications of her institutional role. I would conclude that such a rule is *informal* in nature. Contrast this with a case where the novice worker has been assigned the task to operate a machine but her *boss* tells her how to do so. Now the boss may tell the novice the exact same thing as the experienced co-worker. However, I suggest that the latter case is not an instance of the novice following an *informal* rule. Why? Because the way in which her boss tells her to operate the machines becomes part of the official protocol, in virtue of the superior's authority-according capacity to alter the design-specifications of her role.

[110] The official protocol may even include some forms of sanctions for following informal rules, where these sanctions may be actually or potentially enforced. Institutional groups may face problems regarding the sanctioning of such informal rules. On the one hand, organizations need to uphold a system to track and punish rule-violations by their members. On the other hand, organizations depend on the establishment of a system of informal rules and practices which - in part - require breaking the official rules backed up by such a sanction-based system. See Kühl (2022) for a detailed discussion of this.

For another example, think about the above stated rule regarding making coffee in the morning. We could assume that the above stated rule is determined by tradition, custom, convention, habit, etc. but that the rule is never explicitly and officially endorsed by someone collectively authorized to do so. This rule may nevertheless be effective and cause every role-occupant within the rule's scope to accord her behavior to it. This rule can - in principle indefinitely - remain informal until someone with the power to alter or influence the design-specifications of said institutional roles states that the rule now has an official status. By altering (or "overhauling") the design specifications of the role-occupants by someone in the position to do so, this informal rule then may become formally established, i.e. it may become an institutional role's official *duty to make coffee in the morning*. So, if this rule is determined by someone in an authority-according position and those individuals whose role is thereby altered come -at least virtually- to know about their new duty, we could say that the rule has become *formally established*.[111] On this basis, we can then - *ex negativo* - define informal rules (viz. the rule-based practices and procedures) the following way:

> **INFORMAL RULES:** informal rules are those regulative rules governing a (sub)set of a role's tasks and functions whose content is *not* determined by the officially established and authority-according design-specifications of the institutional role.

[111] We can also see how the change from an informal to a formal rule also captures a change in mode of *enforcement* of this rule. A colleague who ignores an informal rule may be subject to interpersonal criticism, disapproval, bullying, or even outcasting. But a colleague who defies the *formal* rules may be subject to different modes of enforcement, which are integrated into the design-specification of her role, e.g., sanctions in the form of oral or written warnings, dissuasions, salary-cuts, etc. These modes of enforcement may differ, but the rough idea here is that *formal* rule-breaks may be *both* formally and informally sanctioned, but that formal sanctions are *not* available for breaking informal rules: While your colleagues and superiors may hold a grudge against your refusal to make coffee in the morning, their "hands are tied" to punish you for this *through official protocol*. This is part of the reason why work-to-rule strikes are an attractive method for labour unions.

I concede that such a minimal characterization of informal rules will be subject to counter-examples and borderline cases.[112] Ultimately, however, I think that the distinction between informal and formal regulative rules is not discrete, but rather gradual in nature. For one, because there may exist *formal* regulative rules that do not, in any shape or form, *actually* regulate the exercise of the tasks and functions of an institutional role. Such formal regulative rules may have simply been forgotten, or ceased to be relevant.[113] Second, this may be due to the ambiguous, vague and tacit ways in which authority and power govern the relations between role-occupants and their role-performances.[114]

Work-to-rule cases as use of discretionary powers

What's more important for our endeavor than to conceptualize the nature of informal rules, anyway, is to stress that informal rules and procedures play a *functional* role for institutional groups. If the formally established tasks and functions of an institutional role are inapt to successfully allow an individual to function within her institutional role, informal shortcuts, deviations from official protocol, or rule-breaks may help an individual *to uphold the otherwise impaired functionality* of her role. In a sense which I will

[112] This definition, e.g., might be too broad as it does include rules governing tasks and functions which are *otherwise unrelated* to the officially established design-specifications. Think, e.g., of religious rules that govern the conduct of individuals, such as the rules of Sabbath. These rules of Sabbath specify, among many other things, that on Sabbath, practicing jews may walk only 2,000 *cubits*, or ca. 1.2 kilometers per day. Thus, the rules of Sabbath may regulate the activities of a role-occupant *who also happens to be a practicing jew* in certain respects. Yet they are not what I have in mind when talking about informal rules. The reason for this is that the rules of Sabbath do not stand in an *explicit relation* regarding both the institutional role and the constitutive rules governing the tasks and functions of said institutional role. The rule that practicing jews may walk only 2,000 cubits on Sabbath *may regulate all sorts of behavior,* including operating machines, but they are not *explicitly* meant to regulate *this particular type* of activity. So we could amend the definition of informal rules by saying that the informal rules, practices and procedures governing a role-occupant's tasks and functions are 1) those regulative rules governing a (sub)set of the roles' tasks and functions, which 2) are *not* part of the officially established and authority-according design-specifications of the institutional role, but which 3) are related to these design-specifications in a certain, determinable context. Insofar as the design-specifications of an institutional role do not, e.g., make reference to the role-occupant's religious beliefs, we could then dismiss the rules of Sabbath from counting as informal rules governing the role-occupant's tasks and functions. It is, however, possible (think of an institutional group operating solely in an orthodox kibbutz) that the design-specifications of an institutional role actually *do* make reference to the role-occupants religious beliefs, which would allow us to render the rules of Sabbath to count as formal rules governing a role-occupant's tasks and functions.

[113] Think, e.g., about projectionists that set up movies at the cinema. Some decades ago, their tasks and functions may be heavily regulated by rules and procedures in order to prevent them from damaging the projector or the reels containing celluloid-films. Now, this may simply require them to push a button on a digital device.

[114] Imagine a boss calling in on a Saturday and asking her employee whether she can work overtime to finish a pressing project on her day off. Now both of their roles' design-specifications may explicitly state that, in such a situation, the boss has no authority over the employee whatsoever. But the employee may very well know that declining such an "informal" request would be detrimental to the *personal* relationship with her boss. Now the boss, being aware of the ways in which she can exploit this, might say that she did not "officially" command her employee to work on her day off, but that she was merely asking. This would render the employees compliance with the request as purely voluntary instead of authority-based. In this case, the *official* rule to comply with the orders of one's superior seems to be effective although - regarding both roles' formally established design specifications - this should not be the case.

further specify in Ch. 5, informal and formal rules can stand in a *symbiotic* (but also *antagonistic*) *relationship* to each other. Informal rules then can *shield* the agency of institutions from becoming paralyzed. Sociological and organizational theory has long since acknowledged this *functional* dimension of informal rules. Kühl describes the role-occupant's use of informal rules and procedures as the "grease" that "lubricates" formally established and rigid structures (cf. Kühl 2022, 41). Similarly, Osrecki talks about the informal "dirt of systems" which can be contrasted to an institutional group's formal "clinical cleanliness" (Osrecki 2014, 420) of officially established decision-structures and rules of procedure. Finally, Luhmann argues that the use of informal rules and procedures, including the occasional deviation from, or re-interpretation of formal rules, accounts for the "lightness" [Leichtigkeit, M.G.] (Luhmann 1964, 246f.) of institutional group agency.

To discuss an example, take the classical case of the *Prince Friedrich von Homburg*, depicted in a work of dramatic fiction by Heinrich von Kleist (1986). Said Prince von Homburg, a military general, is ordered onto a battle against the Swedes. As a general, Homburg's tasks and functions *explicitly* and *unambiguously* state that he is not to lead his man into battle unless explicitly instructed to do so by a superior. Yet, in a critical moment of the battle, von Homburg *defies* his explicit order and leads his brigade into combat, securing the brittle victory for the Prussians. The Prussian army's overall goal, of course, was to win this battle and this goal was to be achieved by means of a role-based and authority-according division of tasks. Now von Homburg played a necessary part in bringing about the group's overall goal of winning the battle, yet he did so not by *following*, but by *defying* official protocol and by *explicitly breaking the formally established rules governing the design-specification of his role as a military general*.[115]

Having gathered the basic concepts, let us now circle back to assess the work-to-rule phenomena, that supposedly function as counter-examples of a role-based explanation of institutional agency. The idea that I want to put forward is that the informal rules governing a (sub)set of the role's tasks and functions (i.e., the "grease" that "lubricates" an institutional group's functioning) necessarily depend on the use of discretionary powers on behalf of the individual. In order to contribute to an institutional group action, individuals may have to "make up their own mind" and use their role's discretionary powers to decide *how to carry out their tasks and functions*. These discretionary powers might be needed in order to effectively interpret and apply the formally established functions and tasks of institutional roles to the varying and contingent demands placed on them. And in cases of working to rule, individuals simply *stop making use of their discretionary powers*.

In the case of workers fixing small problems of the machines themselves (i.e., them making use of an informal practice), they make use of their discretionary powers by applying the informal rules governing a (sub)set of their roles' tasks and functions. Yet *officially*, they ought to call the repair-clerk. Given that calling the clerk is both time- and resource-consuming, it constitutes a less efficient practice than fixing the machine by themselves. And through the use of their discretionary powers in deciding *how to carry out their tasks and functions,* workers *circumvent official protocol* in order to perform their tasks and functions more efficiently. Conversely, in a work-to-rule scenario, a role-occupant can be said to willfully suspend her use of these discretionary powers. The result of following official protocol is that every time her machine

[115] See Kühl (2022) for further discussion of the normative and legal dimensions of this.

has even the slightest break-down, she calls the repair clerk, stands around and waits until the clerk fixed the minor problem. So I take it that by willfully suspending the use of her discretionary powers, i.e., by *working to rule* and sticking to official protocol on how to operate the machines, the worker contributes to the reduced, or slowed-down output of production. And it is in this way in which working to rule - understood as the willful suspension or deferment of discretionary powers - causes institutional stupor.

Similarly, one can speculate how the Parisian taxi drivers might have used their discretionary powers to weigh certain formal aspects of their role-description against each other, and on this basis might have developed *informal* procedures and practices. Scott here claims that the drivers take "tactical advantage of the fact that the circulation of traffic is possible *only* because drivers have mastered a set of practices that have evolved outside, and often in contravention, of the formal rules" (Scott 1998, 256) [own emphasis]. Let's stipulate that one such informal practice, leading to the most satisfactory service for both types of customers, is to take the most time-saving route for local Parisians, and to take slightly longer, but more scenic routes for tourists. In a work to rule strike, this informal practice might be suspended by the taxi drivers. Working to rule then would mean that a taxi driver (if it is a work to rule *strike,* then this happens in cooperation with others) suspends these informal practices, thus making both kinds of customers less satisfied in their experience of taking taxis and thereby reducing overall profit for her taxi company in the long run.

It is worth noticing, however, that the taxi drivers still obey to all the formally defined tasks and functions of their institutional roles. The point is that they do so in a sub-optimal, and less effective way.[116] Cases of working to rule can be thought of as forms of *malicious compliance* with one's tasks and functions.[117] Institutional role-occupants do not simply *fail* to perform the tasks and functions of their roles if they *work to rule.* In an important sense, their role-performance is *technically* correct. Yet, their work-to-rule performance is sub-optimal, or inadequate for the overall group's course of action, which they are supposed to contribute to.

While institutional roles are general placeholders for activities which different and multiple individuals can occupy, thereby not taking into account any particular idiosyncrasies (or personal traits) of the role-occupants, it also seems *not* to be the case that the individuals occupying institutional roles completely "disappear" behind their role-performance. The exercise of discretionary powers seem to be involved in an individual's role-performance, and role-occupants have to "make up their own mind" on how to interpret

---

[116] And although informal rules practices, procedures etc. might be formally established and officially recognized over time, therefore becoming visible in the formal structure of institutional groups, the diachronic persistence of institutional groups also encompasses the fact that institutional groups are subject to an *ongoing development* of informal practices in order to deal with ever-evolving environments and changing circumstances, as well as constantly changing demands posed on them by their stake-holders (i.e., owners, beneficiaries, members, etc.).

[117] Consequently, this has also been subject to exploitation. To this end, consider the historical anecdote that in 1944, the United States *Office of Strategic Services* [OSS] (the foreign intelligence agency which later became known as the CIA) developed a field manual for citizen-saboteurs in Nazi Germany. Part of the OSS's "Simple Sabotage Field Manual" deals with "General Interference with Organizations and Production". Paragraph 11(a) of the handbook, dealing with the sabotage of organizations and conferences, advises that, in order to sabotage the functioning of an organization, individual saboteurs should "never permit short-cuts to be taken in order to expedite decisions" and that they should "apply all regulations to the last letter" (United States Office of Strategic Services 1944, 28f.).

and carry out the formally defined tasks and functions of their roles. As such, individuals really do make a difference to the performance of their institutional roles. Hence, we should not simply think of role-occupants as input-output automata. Expecting their behavior to accord to such a sterile, formalized way seems to lead to institutional stupor, i.e., rigidity and inflexibility of an institutional group in light of its capacity for action *in situ*. So cases of institutional stupor lead us to acknowledge the necessity of discretionary powers involved in institutional roles. This establishes the first *Problem of Discretion*.

Second problem: role-ambivalence

Let me now turn to the second *Problem of Discretion*, the problem of *role-ambivalence.* The idea here is that discretionary powers may ultimately lead to the breakdown of institutional agency, because they open up the possibility for a *gap between actual and expected performance* of role-occupancy. In these cases, it is not the design-specification of an institutional role, but the discretionary powers of interpreting and applying the design-specification, which become agency-impeding. And this is true for cases where *technically*, individuals can be said to perform the tasks and functions of their roles, yet fail to do so in a way which contributes to their group's overall goal of performing an action. So what's the reasoning behind the second problem?

The rough upshot of my argument here is that if an individual is given *too much* leeway (by which I mean discretionary power in interpretation and application) on how to perform her role, her performance might not contribute to the shared plan or end of the institutional group that she is acting for, where this, in turn, will lead to a breakdown in an institutional group's capacity for action.

A first approximation of the problem of role-ambiguity would simply be to point to the possibility of failure to perform one's role because of a false interpretation and application of one's discretionary powers. This, however, isn't even the real problem. Rather, the more problematic, *epistemic* dimension of discretionary powers reveals itself in cases, where individuals have no way of evaluating themselves, if their action is based on a false interpretation and application of their discretionary powers. This is due to the fact that discretionary powers may not only encompass the choice amongst different pre-defined courses of action, but they may also encompass the *standards* for choosing amongst these different courses. Here, an individual might chose a course of action which "technically" counts as a performance of her tasks and functions, but where this performance is executed in a way which is detrimental to the individual's contribution of a group action.

So let me begin by recalling that the application of discretionary powers, i.e., making *judgment calls*, is part of acting as a role-occupant. This can, e.g., result in the role-occupant choosing action x over the actions y and z where all three courses of actions are equally viable and fall under what is required by the design-specifications of her institutional role. As discretionary judgment calls seem to be involved in an individual's role-performance, the resulting actions of role-occupants may *vary* according to what options an individual may choose on the basis of her discretionary powers. Such variation of output, however, may lead to an outright failure on behalf of the individual to perform the tasks and functions of her role. One can identify a number of situations, in which such failure might occur. In his theory of institutional agency, Kirk Ludwig

seems to acknowledge this, but he limits the potential sources of the above mentioned *gap of performance* to either a lack of *commitment* or a lack of *competence*:

> "The design functions of status roles specify how their possessors are to interact with others in virtue of their assigned status roles. *The possibility of a gap between assignment of role and performance provides scope for evaluating the role occupant for adequacy of performance. The gap arises from two sources. First, from the possibility of failures of competence or performance in the role. Second, from the possibility of possessing the role even in the absence of substantive acceptance of it*, that is, by virtue of not having commitment to fulfilling the functions associated with the status role (or even having commitments to subverting it). Given the possibility of a gap between function and performance, the design specification serves as a standard of evaluation" (Ludwig 2020a, 188f.) [own emphasis].

So a breakdown of institutional agency might occur because the discretionary call of a role-occupant was simply based on imprudence, a lack of judgment or competence, or it was made due to (willful or accidental) ignorance of certain facts and circumstances governing the tasks and functions. To alternate an example from the TV show "The Office" (Lieberstein 2007), a regional sales manager of a corporation might be assigned the task of filing for bankruptcy on behalf of her regional office. This might be part of a layered structure of institutional action, so that her contribution is a necessary part for the institutional group to perform a group-level action, e.g., restructuring its sales-division. If, due to a lack of competence or due to imprudence, the sales-manager decides to declare bankruptcy by yelling "I DECLARE BANKRUPTCY!" while sitting in her office, she not only fails to perform the tasks and functions of her role as regional sales manager, but she also fails to contribute to the overarching institutional group action of restructuring the sales-division.[118]

Alternatively, variation of output may lead to an outright failure on behalf of the (otherwise sufficiently competent and prudent) individual, because what at time t_1 seemed to her as a viable course of action leads to unintended consequences at t_2, consequences which are detrimental to the overarching plan of the institutional group. Here, imprudence or ignorance of certain facts and circumstances governing the tasks and functions within certain situations must not be involved. Yet, e.g., changes in the environment may lead to the occurrence of such detrimental consequences. Imagine, for example, a research assistant in the philosophy department who might be commissioned to plan the annual summer-party at the end of the semester. To this end, she might pick the last weekend of July instead of the first weekend of August for doing so. Planing the party in good faith and being optimistic about the usually good weather in this period of summer, the assistant might nevertheless see her plan fail because of a sudden, unpredicted storm. Although the choice of the particular date was at her discretion and the first weekend of August was equally viable for the party to happen, the variation of output due to the discretionary choice of the research assistant may lead to an outright failure of the philosophy department to celebrate its annual summer-party. But the (potential or actual) failure to perform actions should not strike us as controversial and such

[118] The scene I refer to can be found here: https://www.youtube.com/watch?v=C-m3RtoguAQ (Accessed: 22.11.2023).

cases should not bother us too much. One makes plans and god laughs. Notice, e.g., that the party might've failed even if the research assistant *hadn't* had the discretionary powers to choose amongst several dates possible, e.g., if the annual summer-party always takes place on Hegel's birthday, August 27th. So one's use of discretionary powers is not really the fundamental problem in such a case.

Instead, what I now want to focus on is a more problematic dimension of discretionary powers, which is *epistemic* in nature and concerns the *very standards of evaluation* upon which discretionary actions are based. Here, I aim to show a third source for a gap between expected and actual performance, which isn't considered by Ludwig and cannot be traced back to either incompetence, or a lack of commitment to fulfill the functions associated with one's status role.[119]

In the following, I will present cases in which (both competent and committed) individuals may encounter situations, where they have *no way of knowing themselves* if their actions are based on a false interpretation and application of their discretionary powers (i.e. an interpretation that is detrimental to the overall group action) and where they themselves have to determine a standard against which they could measure the interpretation and application of their discretionary powers. But what do I mean with the claim that discretionary powers do not only encompass the choice amongst different courses of actions, but can also encompass *determining the very standards of evaluation for choosing* amongst these different courses? Gilligan, differentiating between discretion to choose between options given a predefined standard, and the "central case" of discretion to settle these standards in the first place, writes:

> "Discretion in its more central sense relates to […] where judgements and assessments have to be made as to the standards themselves which explain and justify a decision. There is discretion in this sense, either because the standards leave room for variable interpretations, or because the official is left to create the standards for himself. […] Now while we can talk sensibly about discretion arising in the interpretation of standards, this still falls short of what may seem discretionary in the clearest and most central sense. This occurs in respect of *Z,* where the official is required to do, or to refrain from doing, some action, or where there are various ways of performing a task, and in deciding how or whether to act the official has to determine for himself the reasons and therefore the standards which are to guide his decision. Discretion in this sense occurs in an unlimited variety of situations - whether to grant bail or parole, whether to make a welfare payment or in deciding how much, in selecting the site for a new road, or in excluding a piece of evidence. The reasons for thinking of those situations as discretionary in its real, strong, or central sense, are twofold: first, the discretionary decision pertains to a final action (to grant bail), rather than being merely one step or element in the course of a decision (what does XY mean?); secondly, in these situations the office often *is given little if any guidance as to the standards to apply but is required to formulate them for himself*" (Gilligan 1990, 9f.) [own emphasis].

[119] Ludwig leaves open the possibility for provisions that secure both an individual's commitment to and competence in fulfilling the functions associated with the status role. Those provisions, which he considers to be a form of response "calibrated to the degree and dimension along which the performance is inadequate" (Ludwig 2020a, 188) consist of "either […] punishments for failure or rewards for fulfillment or a combination of both" (ibid).

To alternate an example borrowed from Gilligan (ibid), a fire-sergeant might be told to pick the five most experienced firefighters out of her brigade to extinguish a particular tricky forest fire. Picking the five most experienced firefighters out of the brigade is a basic use of the sergeant's discretionary power. It is, within a defined context, an autonomous choice amongst different courses of actions. However, the discretionary powers of the sergeant may also encompass deciding what the open-ended notion of "*being experienced*" means for this purpose. And according to the fire-sergeant's interpretation, the choice of who gets picked might vary, and it might vary *considerably.* The standard for deciding what "experienced" means might be interpreted in the way that, e.g., the *most senior* firefighters out of the brigade get picked, or that *those who extinguished the most forest fires* gets called into the squad. Or it may be based on who was *most recently trained for such events*, or *who knows the area of the fire best*, etc. Accordingly, the outcome of the sergeant's decision, depending on which standard she chooses, might *also* vary to a considerable degree. Now one can easily imagine scenarios where the sergeant's decision leads to a sub-optimal, misdirected if not totally unsuccessful attempt to achieve the envisioned outcome. If, e.g., the sergeant interprets "experience" to refer to the *principle of seniority* and chooses those firefighters who are part of the brigade for the longest amount of time, she might pick the oldest, physically least apt, and least motivated firewomen to put out the fire.[120] Alternatively, the sergeant might interpret "experience" to refer to knowledge of the area in question and pick those firefighters who live closest to the site. This might result in picking a panicking squad that has never extinguished a forest fire and who succumb to the emotional stress of seeing their neighbors' houses being burned to the ground.

Another example of a failure in an institutional group's capacity for action due to the misdirected use of discretionary powers might be the case of *racial profiling* by the police. Imagine, e.g., the case of a policewoman, who has been assigned the task to stop and frisk "suspicious" individuals. Concerning this task, her discretionary powers may also include determining what "suspiciousness" means. Now the policewoman may determine the standard of who's "suspicious" according to her own personal racial biases. Such use (i.e., mis- or abuse) of discretionary powers may not only lead to the policewoman's failure to fulfill the goal of lowering crime rates (because she's notoriously going after only one particular ethnic group), but her actions may also contribute to a diminished sense of trust in her police department and they may undermine the institution's overall goal of maintaining social calm and upholding social peace in the community. However, *technically*, the policewoman just fulfills the tasks of her role, only that she does so in a way which impedes or jeopardizes the overall goal she is supposed to contribute to.

Or imagine a teacher who grades her pupils according to their oral and written contributions in class, yet takes a quantitative instead of a qualitative interpretation of what "good oral contributions" are. Such a teacher might fail to grade her pupils according to their abilities and merits, because her standards for evaluation are inadequate to capture the pupils' performances. Again, the use of discretionary powers, especially regarding the interpretation of action-guiding standards, can be detrimental to the performance of the institutional role's tasks and functions. It is, however, *not* the case that the teacher *doesn't* fulfill the tasks of her role (which includes the task to grade students). *Technically speaking*, she does. The problem is

[120] Let me simply assume here that the physical condition and motivation of the firewomen are related to their age.

that she might do it in a way which is deleterious to her school's overall goal of evaluating pupils based on merit.

Let me recapitulate: In trying to perform the tasks and functions of her institutional role, an individual can find herself in situations where she might need some, or total interpretative guidance in the selection and application of discretionary standards. And at the same time, this guidance might be unavailable to her because she, herself, is supposed to set these standards. Reconsider, for example, the case of NOVELTY. Jen, the governmental spokeswoman does not know the official stance of her government towards the unprecedented situation of an alien invasion. But how should Jen determine how to respond to such a situation? How could she determine her options? And further, what should Jen choose as the very basis of determining her options of response? Should she interpret the situation as a threat to national or global security? Or as a promising surprise; a landmark in the existence of mankind?

In situations where there is little to no guidance on how to perform the functions of one's role and how to apply one's discretionary powers in a way that is anticipated to further the institutional group's goals, the problem arises for role-occupants on how to proceed in fulfilling their tasks and functions. And often, individuals will fail to come up with satisfying ways to contribute to their institutional group's action because of this. Discretionary powers, especially regarding the interpretation of action-guiding standards, then can turn *toxic*, i.e., they can be detrimental to realizing the institutional group's overall actions. In such situations, the use of discretionary powers can actually thwart, circumvent, or impede the realization of the institutional group's goal, to which the role-occupant is supposed to contribute to. And this is true even, or especially for such cases where, *technically speaking*, the role-occupant fulfills the tasks and functions of her role and doesn't formally lack the competence or commitment to do so. So in cases where role-occupants *do* make use of their discretionary powers, this may lead to a breakdown in an institutional group's capacity for action. This establishes the second *Problem of Discretion*.

4.3. Summary

The overall goal of this chapter was to show that an individual's performance of an assigned institutional role cannot *sufficiently* be explained by pointing to the individual acting on the tasks and functions of her institutional role. Let me quickly summarize the argumentative path which brought us to this point.

In the beginning of this chapter, I followed through on the idea that institutional roles are *action-specific* but *agent-ambiguous*. In Chapter 4.1., I tried to show that the agent-ambiguity of institutional roles does not only imply that one and the same role can be occupied by different individuals. It also implies that institutional roles can only provide any particular role-occupant with *generic instructions* or *generalized directives* on what to do. To explain how individuals are able to apply these generic instructions to the specific circumstances they find themselves in, I drew on the concept of *discretionary powers*. Discretionary powers are vested in individual role-occupants in order for them to interpret and apply their institutional roles to concrete contexts. I argued that - at least to some degree - individual role-occupants have to "make up their own mind" on how to perform the tasks and functions of the roles they occupy, and that this happens on the basis of the discretionary powers.

I then argued that if we accept this claim, i.e., that changing circumstances, unforeseen situations, ever-changing environments, etc. necessitate the use of discretionary powers, we come to face a dilemmatic predicament. One the one hand, I argued for institutional roles to be *agent-ambiguous*, and that it does not really matter *who* is occupying an institutional role but rather *what the role-occupant does*. On the other hand, I argued that it is simply not the case that an individual role-occupant's agency completely "disappears" behind the role which she is occupying. I followed up on this tension in Ch. 4.2. Here, I raised the *two Problems of Discretion.* On the one hand, *not making use* of one's discretionary power can lead an individual to contribute to an institutional group's breakdown in its capacity for action. On the other hand, the very use of one's discretionary power can lead to such a result as well.

But where does this place my argument? In the next chapter, I will argue that role-based explanations of institutional agency need to be supplemented. To this end, I will develop the concept of "Role Agency" which aims to account for a deepened understanding of an individual's relation to her institutional role.

Before moving on, let me make clear the goal of my account of Role Agency. My account of Role Agency aims to provide a theoretical framework to understand the relation between individuals and their institutional roles. My goal here is to provide a theoretical explanation for how individuals relate to their institutional roles. As such, I do not aim to provide a practical *solution* to the two *Problems of Discretion*, or to *prescribe* a way how individuals *ought* to relate to their roles in order to avoid these problems. My account of Role Agency is not a guideline that individuals might follow in order to successfully overcome uncertain, indeterminate or novel situations they might find themselves in. Instead, the more modest goal of my account of Role Agency is to give us the theoretical means necessary to make these problems *intelligible* in the first place. So I merely hold that, in virtue of my account of RA, both *Problems of Discretion*, as well as possible solutions that individuals may employ to resolve them, can be better understood and addressed more specifically.

As a provisional approximation, Role Agency describes an individual's ability to reflexively engage with, and act on the role that she is assigned in a group context. To this end, Role Agency captures aspects of an individual's role-performance which go beyond those individuals merely exercising their deontic powers or fulfilling certain tasks or functions. It captures the ways in which individuals come to *internalize,* and on this basis understand, interpret and alter their assigned roles and the corresponding tasks. Exploring the concept of Role Agency will also allow us to further understand how individuals may solve problems concerning their discretionary powers, especially concerning the problem of role-ambivalence. Role Agency provides us with a way to explain how individuals develop a reflexive self-understanding regarding their role-occupancy. In turn, such a reflexive self-understanding can explain how individuals shield the agency of their assigned roles against the backdrop of institutional rigidity and inflexibility. So with Role Agency, we come to understand what might be going on in cases where role-occupants actually encounter situations in which they have to make use of their discretionary powers. And my account of Role Agency allows us to see how individuals may develop strategies for employing or suspending their discretionary powers.

5. Role Agency

As one recruiter put it ‚I had to advise a lot of people who were looking for jobs […] And I'd tell them the secret to getting a job is to imagine the kind of person the company wants to hire and then become that person during the interview. The hell with your theories of what you believe in, and what your integrity is, and all that other stuff.'

Arlie R. Hochschild: The Managed Heart.

Man is essentially the role-taking animal.

George Herbert Mead: Mind, Self, and Society

In this chapter, I present my account of "Role Agency" (henceforth: RA). As a tentative approximation, Role Agency describes a form of agency that individuals engage in when acting in institutional roles which includes both the *internalization* and *idealization* of institutional roles. To this end, RA describes more than merely exercising one's deontic powers or fulfilling certain tasks or functions (which we could call role-*taking)*. What I aim to capture with the concept are the ways in which individuals come to understand, interpret, and alter their assigned roles and the corresponding tasks (which we could call role-*making*). To be a bit more precise, this chapter will argue that RA can be best understood as an exercise of *Role Perspective Taking* (R_{PT}), which is based on the interrelated dual-mechanism of Role-*Internalization (*R_{IN}) and *Role-Idealization* (R_{ID}).

I hold my concept of Role Agency to be explanatory valuable, but not because it proofs the above established role-based explanations of institutional agency wrong. Rather, it allows us to make some substantial ground on the under-theorized relation between individual role-occupants and their institutional roles (see Box 2 in Ch. 4.). RA merely aims to *supplement* the existing approaches. I will try to show how this supplementation, in turn, can help to explain how role-occupants may come to deal with problematic situations regarding their use of discretionary powers, i.e., situations like those we encountered in the last chapter. In any way, RA is not to be read as an independent account of institutional agency; and it does not attempt to replace existing role-based explanations of institutional action. I will begin this chapter by presenting three questions, which this chapter aims to provide answers for. I then quickly explain the account's scope and limitations. After this, I will go through the building blocks of my account in a more detailed fashion.

Three questions regarding Role Agency

I claimed that the established accounts of institutional agency under-theorize the concept of institutional roles. But what exactly is left unexplained by the above mentioned accounts? As I argued above, the

standard description of institutional roles runs into certain problems of explaining how individual role-occupants come to make sense of situations of ambiguity, uncertainty, as well as their use of discretionary powers when acting in their institutional roles. It is simply not the case that individuals "disappear" behind their role-performances. Occupying institutional roles seems to require deliberative and creative processes of interpretation and application, rather than passive compliance with the assigned tasks and functions. I will therefore argue that occupying an institutional role requires a special form of engagement within one's institutional role. My account of Role Agency aims to capture this form of engagement. On this basis, this chapter aims to answer three questions concerning different, but interconnected dimensions of Role Agency: An *epistemic*, a *reflective* and a *relational* dimension.

In its *epistemic* dimension, RA should be able to answer the question how role-occupants may come to understand and have control over their institutional roles in the first place. How, and by which means, do individuals make sense of their assigned institutional roles? Does the occupation of institutional roles require only the acquisition of formally defined knowledge (i.e., the official facts) related to one's institutional role and its design specifications? Or do individuals have to know *more* than what their tasks and functions are? If so, what?

Concerning the *reflective* dimension of role-occupancy, my account should be able explain how role-occupants come to *evaluate* their expected role-performances. This is especially important for the above mentioned cases, where an individual has no means to know herself whether her actions are contributing to an overall course of action. How, e.g., can role-occupants weigh the importance of multiple tasks and functions assigned to them if those tasks and functions cannot be fulfilled at the same time, or if they stand in conflict with one another? And how come individuals to grasp such evaluative standards regarding their role-performances? Are they *internal* rather than external standards of role-performance? If so, how do individuals arrive at such internal (or external) standards?

Lastly, my account should answer the question of how to explain the *relational* dimension of role-occupancy, i.e., how individuals mutually influence each other's role-performances, or how they are influenced in their own role-performance by others, including their broader environment. The *iterability* of institutional roles seems to suggest that role-performances will diverge from one another if one and the same role(-type) is occupied by different individuals. But are there ways in which individuals mutually influence each other in order to ensure role-*coherence*, and mitigate such divergence of role-performance? My aim will be to show that an individual's reflective understanding of her role is something which is not achieved purely by herself. Instead, intersubjective, social and collective processes play into the management and monitoring of institutional roles. This may either happen in virtue of other members of the institutional groups, in virtue of the individual's broader environment, or via mechanisms that institutional groups *themselves* deploy. These questions regarding the *epistemic*, *reflective*, and *relational* dimension of role-performance will consequently guide my explanation of the concept of *Role Agency*, and I will argue that RA can answer each of them.

Scope and limitations of Role Agency

Before I move on, let me make clear the scope of my account, and thereby also its limitations. It should be noted from the start that my account does not aim to give a *general theory of social roles*. Instead, it is restricted to the concept of *institutional roles*. The difference between these two is due to particular features of *institutional* roles, which *social* roles in general do not seem to exhibit. To stay within the established vocabulary, institutional roles are what Ludwig calls *agent status roles*, i.e., they are roles that involve role-occupants to exercise their agency to contribute to a group's overall goal. A look at the examples of *agent status roles*, e.g., policewomen, judges, lawyers, employees, professors, students, senators, governors, etc. (cf. Ludwig 2018b, 61) is illuminating to draw out this difference. Within the general category of social roles, we find social roles such as *women, spouse, parent, friend, neighbor, elder.* These roles do not, in and of themselves, give rise to the form of organized, and structured groups that my project wants to explain the agency of.[121]

Second, and connected, is that RA should be understood as being restricted to those institutional roles that actually *prime for action*. Above, I made the distinction between *activity* and *passivity* that characterize institutional groups. Depending on their nature, institutional groups can exhibit a divide between *active* members, whose roles *prime the individual for action*, and *passive* members whose roles do not require the individual to act in certain ways. Within institutional groups, passive institutional roles do not require the individual to fulfill tasks and functions, or they may do so in only a limited sense. One may very well join and become a member of an institutional group, without ever doing something related to one's passive membership-role (i.e. one can become a so called "Karteileiche"). Such a "simple" and passive member might nevertheless have certain rights or duties, e.g., the right to partake in assemblies, the right to vote in a referendum or election, etc., the duty to pay one's fees, etc. But one's membership might not depend on whether one *actually* exercises these rights and duties.

On the other hand, there are those roles within institutional groups which do not simply consists of *having membership-status*, but which require the active, operational exercise of tasks and functions. Consider e.g., being a political party's *general treasurer*, occupying the role of a *trade-union secretary*, or being *member of the editorial board of ISOS*. Now membership in *some* institutional groups, like business corporations, or police departments, *necessarily* involves that the institutional role one occupies *primes for action*. One cannot become a member of, e.g., a profit-driven corporation just by signing a form and paying certain fees. Within such groups, membership is conditional on one's role being connected to tasks and functions designed to contributing to a collective outcome, or to realize a collective goal. And it's *these* roles, i.e. roles

[121] Such feature-based groups, of course, may, but also may not, provide the basis for collective action. We could here talk of *organized, institutional* groups that have as membership-conditions to occupy a *social role* (e.g., *American Mothers, Inc.*; or the *National Neighborhood Watch*) that are distinct from the unorganized, feature-based social groups (e.g., the group of American mothers; or the group of every individual who has a neighbor) they depend on. The latter kind of groups, while being constituted by individuals occupying certain *social* roles, usually does not require some form of acceptance, or recognition of membership, including the delegation of tasks and functions among their members, on behalf of the individuals who come to occupy them. However, being a member of *National Neighborhood Watch* means that an individual comes to occupy an institutional role, i.e., joining an institutional group and on this basis accepting the structural arrangements of such a group. In contrast, and by definition, one becomes a neighbor by having someone moving in next to them, whether one is aware of this or not.

which are substantially connected to tasks and functions functionally designed to contribute to an institutional group's goal, that my account of RA is targeting.

Limiting my account this way, RA can and does not aim not explain every aspect of what it means to occupy any given institutional role. For example, I remain silent on the question whether individuals could possess the capacity for Role Agency regarding such "passive" institutional roles. Also, Role Agency does, e.g., not answer to questions regarding the occupation of *multiple* social and institutional roles by an individual. Insofar as individuals can occupy different social and institutional roles at the same time, questions regarding the possibility of such roles mutually influencing each other, or individuals experiencing *inter-role conflicts* (e.g., being the head of the United Nations Climate Change Conference *and* being the CEO of one's country's National Oil Company) remain unanswered. Role Agency also does not explain possible ways in which individuals may come to resolve such inter-role conflicts. Finally, RA does not answer questions as to how occupying a given *institutional role* may be influenced by an individual's given *social* role or roles, e.g., whether and/or how *being a mother* influences the conduct of individuals within the institutional role of a *judge*, etc.

A special "role mode" of agency?

Before going through the building blocks of my account, let me qualify my claim that Role Agency is a "form" of individual agency. What do I mean by saying that individual agency may take on different "forms"? In short, my claim is that RA is a *subtype of individual* (*human*) *agency* that is based on the individual's capacity to adapt externally defined beliefs, desires, and goals. I will clear up this notion below, and explain how RA should be conceptualized as a form of *Perspective Taking*.

But first, I want to demarcate my theory from a proposal of Michael Schmitz (2017; 2023). Schmitz argues that individuals occupying institutional roles engage in what he calls the "*role-mode*". When individuals engage in the "role-mode", Schmitz states, they *identify as role-occupants*, and on this basis adopt the "vantage points" of the institutional roles they occupy. According to Schmitz, the "[r]ole-mode is a mode of identification with the role, with the group in the context of which a role is defined, with its powers and obligations, and with the practical and theoretical skills and the domains of knowledge it requires" (Schmitz 2023, 190). When individuals adopt the role-mode, this entails a "modification of consciousness" (Schmitz 2023, 191) that allows the individual to adapt "the point of view of the role and take positions in light of the theoretical and practical knowledge, the values, tasks, obligations, powers and so on, that the role affords" (Schmitz 2023, 194). My own account of Role Agency is in many respects similar (and sympathetic) to this approach. Yet, it also differs in important aspects. The main difference of my account to that of Schmitz's "role-mode" pertains to the theoretical underpinnings of such forms of individual identification with institutional roles.

For Schmitz, the role-mode is a variety of what he calls the *subject-mode.* Schmitz describes his "subject mode" account of collective intentionality as an "attempt to marry a mode account in the tradition of Tuomela and John Searle with a subject account as it has been defended by Margaret Gilbert and Bernhard Schmid" (Schmitz 2023, 183). Regarding the *mode*-aspect of this wedlock, Schmitz project aims to convey "layers" of collective intentionality, which build upon each other. Here, higher forms of collective

intentionality, such as the "documental level of shared contracts, constitutions, legal frameworks, cultural heritage etc.“ (Schmitz 2023, 186), are built upon a "conceptual and propositional layer of belief, intention, theoretical and practical knowledge" (ibid). These, in turn, rest on a "preconceptual and nonpropositional layer of perceptual, actional and emotional experience" (ibid). Regarding the subject-aspect of his account, the key claim of his theory is that individuals - regardless of the level of intentionality - always experience and represent their position towards the world in an irreducible mode of *group mindedness*, which "consists in experiencing and representing others as co-subjects of positions towards the world" (Schmitz 2023, 189). Accordingly, adapting the role-mode, i.e., the "point of view" of an institutional role, always entails that individuals not only represent themselves as occupying such roles (and on this basis, represent certain positions that their role entails) but also that they represent other group-members as *co-subjects* of these positions. Identifying with one's role necessarily entails "experiencing and understanding [oneself] in relation to co-subjects and by being in turn experienced and understood by these co-subjects" (Schmitz 2023, 192).

Now such group mindedness, in turn, enables individuals to jointly constitute institutional groups, which Schmitz holds to be "higher-level plural subjects" (Schmitz 2023, 182). Note that it follows directly from Schmitz's account that institutional groups always depend on their individual members realizing the necessary *group mindedness* of engaging with each other as co-subjects. This includes acting in one's role within an institutional group which, again, gives an individual a role-specific vantage point.

Institutional roles give the individual access to "positions in light of the theoretical and practical knowledge, the values, tasks, obligations, powers and so on, that the role affords" (Schmitz 2023, 194). But by occupying these roles, individuals also represent their group-members as *co-subjects* of these positions. It follows that on Schmitz's account,

> "a group can only take a position through at least one individual taking it as a group member and / or as the occupant of a role within the group. At the same time, that one individual takes such a position is *never sufficient* for the group taking it *because representing the group's positions is always essentially a task shared by all group members*" (Schmitz 2023, 198) [own emphasis].

Now my main worry with Schmitz's proposal is not his description of the way in which individuals relate to their institutional roles, i.e., them adopting the "vantage points" of the roles occupy, or them employing role-specific "forms of reasoning". Rather, my worry is this: The fact that Schmitz's role-mode is convoluted with his *subject-mode*, makes the explanation of individuals acting in the role-mode inapplicable to *institutional* groups in the first place. More precisely, I hold Schmitz's theory to have a problem in explaining how a subject-mode could underlie anonymous and compartmentalized cooperation within large, and complexly structured institutional groups, which may undergo a steady change of membership. This especially concerns the possibility of proxy-agency within institutional contexts.

Consider institutional roles that require individuals to fulfill the tasks and functions of their institutional roles without either knowing what the group's overall goal is that they are contributing to, nor having an understanding of who comprises the group that they are acting for. Imagine, as an example, some individual

being hired as a spokesperson for a large institutional group, where this role entails the one-off task to assert the group's position towards a particular state of affairs, e.g., which candidate the group will support in an upcoming election. We can imagine that this individual is given access to the necessary information, gets to know all the relevant facts and that she understands what her goal of acting *as a spokesperson* is. Now, in asserting the group's agreed upon position on the candidate, she is authorized to act as a *proxy*, i.e., to represent the group's position towards a certain state of affairs.

The above described role-based theories of institutional group agency (Ch. 3) fare particularly well in explaining such cases of proxy-agency. Acting as a proxy-agent of an institutional group (just as acting in one's institutional role in general) does neither have to entail full knowledge of the institutional arrangement that one is tacitly accepting, nor that one has to understand the overall group's goal, or how one's role figures in bringing it about, when performing a contributory action as a role-occupant. Because institutional groups divide their labour, role-occupancy entails specialized sub-tasks that can be fulfilled individually, without making reference to a "we" or group. Because these sub-actions are functionally integrated, they still contribute to a group's overall goal of performing a complex action.

However, the idea that adopting a role-perspective ultimately requires a form of *group-mindedness*, i.e., that it requires representing *others as co-subjects*, seems to be inadequate, and ultimately redundant in such cases. Although a group may assert its position through its proxy, it is unclear why this is a task which necessarily has to be shared by all group members, or that requires a form of *group-mindedness* from the proxy. We can, e.g., assume that some members of the institutional group may have no idea that the spokesperson was hired to assert the group's position towards the candidate. Likewise, it does not seem neither necessary, nor plausible to assume that the spokesperson acts in Schmitz's mode of *group mindedness*, i.e., by representing other group-members as *co-subjects* of the positions she, in her role as the spokesperson, asserts. After all, she may not even know who else comprises the group she asserted this position for, but simply did as her role required her to do.

What we need, then, is a way to describe how the adaptation of a role-mode, or the identification with a role can come about without such a form of group-mindedness. I hold my account to be capable to do so, because RA is a form of individual agency that does not require a *special mode of group mindedness* or to represent other group-members as *co-subjects* of one's role-positions. Instead, RA is based on an individual's capacity to adapt externally defined beliefs, desires, and goals. Contrary to Schmitz's claim that occupying roles and taking up their perspectives is something that individuals achieve by switching to a special *mode* of intentionality, I will argue that such perspective taking ultimately is to be analyzed as a matter of the *content* of the individuals' intentional states.

Now I still claimed above that RA is a form of agency, that consists in a *modification of individual agency*. To mitigate the potential misapprehension of my claim, let me re-formulate this within Schmitz's vocabulary: RA is held in the I-mode and does not make reference to a group, some further unspecified "we", and it also does not involve the individual displaying a form of "group mindedness". We could say, for the lack of a better word, that RA is a *modification,* but not a *Schmitzean mode* of individual agency. This *modification* consists of individuals adopting an externally defined framework of beliefs, desires and goals (what I call a *perspective*) that functions in their individual, or I-mode agency. But for clarities sake, I will refrain from

using the term "mode" to describe this modification of individual agency that my account of RA wants to describe.

Role Agency as a subtype of individual agency

I now want to address my claim that Role Agency is a "modification" of individual agency. What do I mean by saying that individual agency may be "modified"? In short, my argument sees RA to constitute a *subtype of individual agency* that based on the individual's capacity to adapt a framework of externally defined beliefs, desires, and goals. Role Agency, as a capacity of an individual to adapt a framework of externally defined beliefs, desires, and goals, etc. does not directly require the individual to realize, or hold collective intentional states; to have "we"-thoughts; or to engage in an irreducible (we-)mode. RA is simply a form (or subtype) of individual agency. Still, RA has some interesting features in relation to individual agency. This is because institutional roles shape the agency of individuals in distinct ways, *enlarging* it in some dimension, while *decreasing* it in others. So let me say a bit more about the relation between individual agency and RA, i.e. the way in which individuals may act when occupying their institutional roles.

First, RA is *contextual* in a way in which individual agency is not. By this, I mean that RA is form of individual agency that may be exercised by an individual during a certain period of time and only within certain contexts, or situations obtaining. Consider e.g., the institutional role of the *Chair of the German Federal Assembly*, which has as its function to elect the President of Germany. The Federal Assembly is an constitutive organ of the German state which comes together every five years, and which dissolves immediately after electing the head of state. Occupying the role of the *Chair of the Federal Assembly* is defined through a very specific set tasks and functions, but it is *temporally restricted for a short amount of time within a highly specified context*. Hence, the institutional role of being the *chair of the* German Federal Assembly provides for a temporally and contextually limited form of RA. Other institutional roles require from their occupants a steadier form exercise, e.g., occupational roles that are exercised "9 to 5" on workdays, maybe throughout several decades. But even such steadier institutional roles are contextual. Institutional role-occupants may leave their tasks and function "at the door" when exiting their workplace, or during their time off. RA then is temporally and situationally *bound* and an individual can *cease to exercise RA* without ceasing to be an individual agent. RA is also temporally and situationally bound in the sense that individuals may cease to exercise RA for a while and, at some later point, may *resume* their exercise of it. Thus, the dependency of RA and individual agency is *asymmetric*: An individual's agency does not exhaust itself in one's capacity for RA. One can cease to exercise one's RA without ceasing to be an agent, but one can exercise RA only if insofar as one already is capable of exercising one's individual agency.

The relation between RA and individual agency turns out to be *asymmetrical*, or incongruent in another way: not regarding one's general capacity for action, but regarding one's *role-specific agentive capacities.* This asymmetry of Role Agency runs in two directions. RA is a form of agency that is both "*bigger*" and "*smaller*" than individual agency. Let me explain. RA is form of agency that is "bigger" than individual agency, because it *enhances* (or enlarges) the agentive capacities of the individual role-occupant. *As role-occupants*, we can do things we otherwise could not do. That's really one of the neat aspects about them. One way to understand this claim is to point to the fact that some institutional roles require the role-

occupant to develop certain practical skills, or theoretical knowledge necessary for performing certain actions, that they otherwise would not have developed. Occupying the institutional role of a corporate bankruptcy lawyer, e.g., requires the individual to gain knowledge related to filing legal papers for a corporate bankruptcy. This may be a task that consist of several, complicated administrative and legal steps that someone without the role-specific knowledge of such a lawyer would not be able to do. Likewise, developing the practical skills necessary for performing heart transplantations require from a surgeon years of exercise and training. A friend of mine, a stonemason, can carve perfect cubes from rocks of marble, a task at which I would most definitely fail.

But note that to understand the enhancement of agentive capacities of institutional roles this way makes them contingent on what capacities an individual agent has in the first place: It is at least logically possible that an individual develops the practical skills of performing heart transplantations without ever occupying the institutional role of a surgeon. Likewise, it is at least conceivable that I, myself, could carve perfect cubes from rocks of marble without being, or thereby becoming a stonemason. Also, if an individual acquires certain skills or abilities in relation to her role, she also can cease to occupy the role without ceasing to have these abilities, e.g., by retiring. If, one day, my friend quits working as a stonemason, he can still shape rocks this way.

What is then more fundamental to the enhancement of the agentive capacities of role-occupants that some actions require for their performance the existence of an *institutional context*, or *background* against which they can be performed. Herein lies the true sense of enhancement of agentive capacities that roles provide. Institutional roles allow individuals to do things, that they could - *in principle* - otherwise not do: *Judges* are able to *sentence* individuals to prison, lawyers are able to *represent defendants* in courts, *university professors* are able to *give doctorates*, *politicians* can *vote* on a piece of legislature, *policewomen* are allowed to *stop and frisk* citizens, etc. Importantly, individuals are able to do these things only within an existing institutional (and societal) context, e.g. the law system, the educational system, the political system, a state-run executive, etc. This form of agency-enhancement is connected to the fact that institutional roles come with *positive* interpersonal, or *deontic powers*, such as rights, permissions, authorizations, or entitlements attached to them. It is only on this basis, within an existing institutional context securing the recognition of individuals to have these powers, that their roles enhance their agency. My claim that roles enhance the agentive capacities of individuals should be primarily understood in *this* contextual sense.

But the relation between RA and individual agency is asymmetrical in another way: RA is a from agency that is not only "bigger" but also "smaller" than individual agency as well. By this, I mean that it can also *restrict one's agentive capacities.* Like the enhancement, the restriction of agentive capacities through role-occupation depends on the existence of an *institutional context.* Here, *negative* interpersonal, or deontic powers attached to an institutional role, e.g., duties, obligations, or restrictions, may prohibit individuals to perform actions they could (or would) otherwise do. Consider, e.g., the fact that in Germany, paramedics are legally *prohibited* from injecting syringes, or give certain types of medicine to victims of an accident, at least unless a medical doctor *authorizes* them to do so. An individual occupying the role of a paramedic may otherwise be very well practically capable of injecting a syringe, yet *acting as a paramedic* restricts her capacities to do so. So this should not be understood as the claim that roles restrict an individual's capacity

for action *in general,* but rather, that an individual's agentive capacities can be restricted *qua acting as a role-occupant*.

Or consider, e.g., the capacity to reason, or to deliberate about evidence and to reach certain conclusions based on such evidence. Here, *judges* may base their verdict on deliberating about evidence only insofar as the evidence introduced is *legally admissible*. Basing her judgment on both the *admissible and inadmissible* evidence, an individual (who is also a judge) might hold a defendant to be guilty. Yet, when reasoning *as a judge*, that same individual is restricted to deliberate only on basis of the *admissible* evidence, and, due to this restriction, she might consequently rule in favor of the defendant. Hence, the individual's capacity of *reasoning as a judge* is restricted by the institutional design and context of the role.[122]

So Role Agency is a form of individual agency that is contextual and asymmetrical. The important upshot here is that the relation between RA and individual agency implies an asymmetric dependency of RA on individual agency. An individual's agency does not exhaust itself in one's capacity for RA. One can cease to exercise one's RA without ceasing to be an agent. However, one can exercise RA only if insofar as one already is capable of exercising individual agency. RA then is a *modification*, or for the lack of a better word, a *form*, or *subtype* of an individual's overall agency. However, the *institutional contextuality*, or *institutional embeddedness* of RA allows individuals to exhibit forms of agentive capacities they would otherwise not be able to exercise. Thus, RA is more similar to what C.T. Nguyen (2019, 2020) calls a "layered agency"[123], than to Schmitz's role-mode.

In the next section, I want to address the question of *how* individuals are able to take up this form of RA. What enables individuals to appropriate their institutional roles' agentive capacities? How do they come to adapt these role-perspectives?

---

[122] There are, of course, cases in which individuals do not follow through on the restrictions that their institutional roles provide for them viz. their actions, e.g., a paramedic injecting a syringe although she is legally *prohibited* to do so. Such actions, which fall outside the agentive capacities of an individual *acting qua role-occupancy,* cannot be considered to be the *authorized*, or *officially licensed* actions of role-occupants. Rather, we should think of such cases as unauthorized, or *rogue* actions of individuals, which are *parasitic on* actions that are officially related to institutional roles. Institutional Role Agency in this sense is a form of *authority-bound* agency. But as discussed above, this is not a categorical distinction between official and unofficial actions of role-occupants, because roles do not exhaustively determine all possible actions of the individuals who occupy them. There will then, within any institutional context, I conjecture, be "grey zones" of what counts as official, or unofficial action. This is not to say that such overreach of power can never be clearly identified and institutional groups may exhibit mechanisms to regulate such behavior. One mechanism, especially employed by large, and complex institutional groups, is to set up *compliance policies* that deal with such violations of authority.

[123] Nguyen develops the concept of *layered agencies* primarily in the context of games and so called "striving play". But he also gives in to the idea that institutional roles provide temporal, or layered "agencies". He suspects, "in fact, that, once games have helped us to get a good picture of how layered agencies work, we will find them elsewhere in life" and assumes "that many professional roles —the professor role, the lawyer role, the judge role—involve taking on a layered agency. For example, one might think that a person interested in arriving at a balanced and considered set of beliefs might best achieve their goals by temporarily submerging themselves in a role in which they advocated wholeheartedly for one position, inside an adversarial system in which somebody else took on an opposing role" (Nguyen 2019, 21).

5.1. Role Perspective Taking

Next up is the concept of *role perspective taking* (henceforth: R_{PT}) and the way in which it figures in my account of RA. I will argue that R_{PT} is key to understand how individuals come to appropriate their institutional roles' agentive capacities. As already mentioned, I do not think that occupying roles and taking up their perspectives is something that individuals achieve by switching to a special *mode* of intentionality. Rather, such perspective taking ultimately is to be analyzed as a matter of the *content* of the individuals intentional states. I will now put flesh on the bones of this claim by showing how *role perspectives* provide for such content. To reason, intend and act *as a role-occupant* requires the individual to take the "perspective" of this role. In turn, a perspective is (just) a role-specific set (or framework) of beliefs, desires, goals etc. that is adopted by the individual and constitutive of said individual *acting qua role-occupancy*.
Now the term "perspective" is somewhat ambiguous. On the most basic level, I take it to refer to a relation that an individual entertains to an object or state of affairs (cf. Crone 2021, 11817). In the following I will defend a minimalistic use of the term perspective, and stick with Kay Mathiesen's (2006) use of the term. Mathiesen takes the term "perspective" to denote a "framework of beliefs, desires, and goals that orient an agent in the world" (Mathiesen 2006, 243f.). According to Mathiesen, a perspective provides a "point of view from which an agent can respond to her environment - to reason, make choices, and form intentions" (ibid). If we adopt this minimal definition of perspective, we can derive from it the following claim, which I will now try to defend:

> **Role Perspective Taking (R_{PT})**: When an individual adopts (or takes on) the role-perspective, she reasons, intends, and acts from the perspective of the institutional role that she occupies, i.e., she reasons, intends, and acts based on the framework of beliefs, desires, and goals of her institutional role.

Is this claim really feasible? One major way to be skeptical is this: Perspective taking is mostly discussed in the debate about *mind reading* and in the *Theory of Mind* more generally. Here, perspective taking is taken to be based on the capacity to mentally model other persons' beliefs, desires, feelings or goals.[124] For example, Maibom (2020, Ch.2) describes perspective taking in general to consist of three consecutive steps. When we take another *person's* perspective, we "1) move our first-person perspective to the other person, 2) we imagine the other person's situation, the result of which is that we 3) acquire information about what it is like to be in that situation" (Maibom 2020, 36). Such forms of "agent perspective taking" (ibid), which Maibom contrasts with the process of "visual perspective taking", may then be one way - out of several - in which individuals simulate the *mental states of other persons* in order to understand and predict their behavior, or to develop *cognitive empathy* with the targeted individual. Now if we were to adopt this definition, we should be hesitant to hold the concept of perspective taking to be applicable to institutional roles in any shape or form. After all, Maibom's *agent perspective taking* is about simulating *mental states*

[124] I wish to remain neutral on whether such a capacity for mind-reading should be best understood in terms of theory-theory, simulation-theory, or whether pluralistic approaches can best account for this (see for discussion: Spaulding 2018.).

and concerns the perspective of *persons*. On what basis could we assume that individuals can take an *institutional role's* perspective? After all, institutional roles are not persons; and they don't have mental states, such as beliefs, desires, and goals to begin with. And what else (if anything) should provide the basis for modeling a *role's* perspective? If perspective taking is a process which allows me to *feel what you feel*, or *to understand the world from your point of view*, then adopting the perspective of an *institutional role* seems mysterious. Institutional roles don't feel anything, nor are they subjects with experiences, or *points of view*. The concept of perspective taking would then proof to be inadequate when applied to institutional roles.

My claim could be misunderstood in another way, namely that I do claim that institutional roles indeed *are* subjects, whose perspectives are taken by the individual occupant. But this would be metaphysically obscure, as it would require me to postulate the existence of some sort of free-floating, supra-individual subject, whose mind and mental states exists independently of any individual agent. Perspective taking, then, would be some sort of *reverse exorcism,* i.e., a process through which an supra-individual entity's mind is *implanted* into the individual (instead of evicted). Fortunately, we do not have to call the priest to work around this objection.

Role Perspective Taking as based on role-specific reasons for action

While institutional roles themselves don't have mental states which could be simulated by the individual role-occupant, *acting as* a role-occupant indeed requires the individual to realize certain role-specific mental states, i.e., to hold beliefs, desires, and goals which are either associated with, or constitutive of acting in institutional roles. Why is that?

The answer is that occupying a role provides an individual with *role-specific reasons for action*. These *role-specific* (or role-*dependent*) *reasons for action*, in turn, can be cashed out in terms of beliefs, desires, and goals, some of which are constitutive of occupying an institutional role and carrying out its tasks and functions. To take the perspective of an institutional role then bottoms out in the individual's capacity to adopt, and consequently act on these role-specific reasons for action. R_{PT}, understood in this sense, does not describe processes through which individuals mentally simulate other *persons*' mental states. Nor does it postulate the existence of some sort of free-floating, supra-individual subject, whose perspective is taken.

Let me make some qualifying remarks about these *role-specific reasons for action*, which provide the basis for role-based perspective taking. To avoid confusion, such role-specific reasons are *external*, *desire-independent* reasons for action, rather than *internal reasons* of individuals connected to their private motivational states (see for a discussion of this desire-independency of such reasons: Searle 2010). But how can institutional role's provide reasons for action? Here, I reside with the above mentioned account of role-based reason to perform an action (Ch. 3.2.) by Kirk Ludwig:

> "For any *x, x* has a role-based reason to *A* in *C* at *t* iff: a) *x* has a status role at *t* which obligates *x* to *A* in *C; and iff* b) it is part of the design specification of a status role *x* has at *t* that *x A* in *C*" (Ludwig 2017b, 149ff.).

The central idea is that roles are (at least partially) defined through tasks and functions an individual is obligated to fulfill when she occupies a given institutional role. If x' task, e.g., as a clerk of the factory (C), is to unlock the gates at 6 a.m. (A), this provides x with a *role-dependent* reason *r to A in C*. The reason *r to A in C* stems out of the obligation that is connected to occupying this particular institutional role. And the obligation, in turn, consists in fulfilling the tasks and functions through which the institutional role is (partially) defined. Thus, the reason *r to A* requires x to realize certain actions that need to be performed in order to fulfill these tasks and functions. In turn, these actions require the individual to realize certain mental states in order to perform them, e.g., plan-states about x's performance of such an action, beliefs about the nature of this task, or certain motivational states. Thus, *role-dependent* reasons *to A* are connected to corresponding sets of beliefs*,* desires, intentions etc. insofar as the realization of these mental states is necessary for bringing about the actions specified by the *role-dependent* reasons *to A.*

For another description of such reasons, recall one more time Schmitz's description of the role-specific reason that an individual might have while occupying the institutional role of a policewoman. One of the constitutive tasks of policemen and -women is to sanction illegal behavior when observed. So, at some time (t) and in a situation (S), a role-occupant (x) might observe illegal behavior, e.g., illegal drug use. In turn, this provides x with a role-specific reason (r) to sanction individuals (A) who exhibit such behavior:

> "The policeman may reason deductively from her belief *as a policewoman* that a certain man has smoked a joint and his (let us assume) general obligation *as a policeman* to arrest people who do such things, to the particular obligation to arrest this man. It is necessary that this belief be one that the man holds as a policeman because if, for example, his personal belief was based on inadmissible evidence – say, obtained through illegal wiretapping – it could not provide a legally valid reason to arrest the man even if it was true" (Schmitz 2017, 62).

Crucially, this reason *r to A* also involves the realization of specific, role-dependent mental states, that are constitutive of acting for a role-based reason: When a policewoman reasons deductively from her belief *as a policewoman* that someone has used illegal drugs and is ought to be arrested, this requires her to form relevant *motivational states* to act on the obligations she has *as a policewoman* to arrest people who do such things, and it may also require to develop role-specific plan-states, or intentions to organize and follow through on such action.

Now role-specific reasons for action may require the individual to realize certain mental attitudes, which can either be *associated* with, or *constitutive* of these roles. To say that they are constitutive is to say that these beliefs, goals, intentions etc. must be necessarily realized in order to perform the actions which are constitutive of the role that an individual occupies, i.e. the tasks and functions through which one's role is (partially) defined. We can further characterize these role-dependent reasons to be *temporal* and *contextual* in that i) x would cease to have them if x at some point was no longer to occupy the institutional role; and ii) they do not guide the action of x outside of the *context* of her role-occupancy. The *temporal* and *contextual* character of role-based reasons may fall together, which can be demonstrated by the fact that some individual x may have a reason *r to A* while she actively occupies the institutional role (e.g.,

during work-hours), but where this reason does not travel outside of the context C of her occupation (e.g., that she has no role-based reason to A *on the weekend,* or *on her day off*).

The definitions of role-based reasons for action of Ludwig and Schmitz omit the question whether these reasons should be understood to merely be *motivating* reasons, i.e., reasons for which agents act, or whether they should be understood as *normative* reasons for someone to act. A normative reason here can be understood as a reason an agent has "because it favors […], supports, or makes a case for, or helps justify, that course of action" (Alvarez & Way 2024). Both Ludwig's and Schmitz's definition do, however, point towards an interpretation that sees role-based reasons to be normative, at least *prima facie*, insofar as the *obligation* of a role-occupant implies that the action has a *deontic* status, i.e., "that all things considered, she [the role-occupant M.G.] ought, must or may do that thing" (ibid). The deontic status of a role-based reason, in turn, can be traced back to the definition of institutional roles as *agent status* roles, which entail (both positive and negative) deontic powers that are attached to a given role (see Sec. 3.1. above).[125]

At this point, and to avoid further confusion, let me say something about the relation between R_{PT} and the features of institutional roles of being *iterable* and *interchangeable.* Here, a type- and token-distinction should be kept apart: When I claim that reasons for action are *role*-specific, I claim that they are not the reasons for actions of *any particular* individual who occupies a role, but that they are the reasons that govern the conduct of individuals who occupy a given *type* of institutional role *in general*. Two individuals occupying the same institutional role (either diachronically, or - in case of the role being multiply realized - synchronically) will have the same *type* of role-specific reasons for action. But if each individual acts on this *type* of role-specific reasons for action, this will lead to *token-realizations* of role-specific reasons for action by the individual (but of course, there might be types of institutional roles that are instantiated only once, or by one individual token realization).

Take the example of two individuals who each occupy the same institutional role-type of a policewoman. According to my claim, both individuals will have the same *type* of role-specific reason for action and that each individual will hold *token*-reasons for their *particular* actions that their institutional roles require them to carry out. But at different times throughout both of the individual's role-occupancy, the individuals may

[125] For an extended discussion of how and whether social roles provide for normative reasons see Blackman 2023. In short, Blackman argues for using Mark Schroeder's *Humean account of normative reasons* as a structural basis for what he calls "Role Pluralism" concerning the nature of social roles and the existence conditions for normative reasons. He defends a view of role-based reasons according to which role-based normative reasons are *externalist* and *non-hierarchical*. By stating that role-based normative reasons are external (rather than internal reasons), he posits that they do not depend on the particular motivational constitution of role occupants (i.e. that they do not ultimately depend on a role-occupants *desires*) but rather that for R to be a role-based reason for X to do A is for there to be some end (p) such that "X occupies a role with the end of p, and the truth of R is part of what explains why X's doing A promotes (or realizes) p" (Blackman 2023, 166). These role-based normative reasons are *non-hierarchical* in the sense that they do not depend on the individual occupying a "master role" (e.g., the role of *being a human*) from which all the normativity of occupying other roles is derived (Blackman 2023, 161-163). It is, however, a different question whether the role-based normative reasons that Blackman argues for hold across all sorts of *social* roles, or whether the normative reasons of *institutional* roles (e.g., being a judge) differ from the normative reasons of social roles (e.g., being a parent). Blackman does not distinguish between these two forms of social roles. Another question is whether such role-based normative reasons should be understood as *moral* reasons. I want to thank Eva Schmidt for helpful comments on the nature of such role-based reasons.

hold different token-reasons for their particular actions. If, e.g., individual x_1 is conducting traffic control, this requires her to carry out certain specialized sub-tasks and functions, which in turn require her to have specific token-reasons for these particular actions. At the same time, individual x_2 might be, e.g., interviewing a suspected criminal. Accordingly, x_2 will have different token-reasons for *this* particular action. The point is that generally, both individuals will have the same *type* of role-specific reasons for action insofar as they occupy the same role-*type* that requires them to carry out the same tasks and functions. This implies that there are *types* of role-specific reasons for action that may actually never be realized in token-instances, e.g., when throughout the time of her occupation, a policewoman never has to conduct traffic control although it is part of what the role *potentially* would require her to do. And while the *types of role-specific reason for action* can remain intact through the ongoing change of occupation of any given institutional role, the realization of a token *role-specific reason for action* is something that always requires a particular individual. Thus, R_{PT} concerns the adoption of a *type*-perspective, which is realized by *token-instances*.

So when an individual adopts (or takes on) the role-perspective, she reasons, intends, and acts based on the framework of beliefs, desires, and goals of her institutional role. What warrants my claim that institutional roles come with such a framework is that a) institutional roles are (primarily) defined through role-specific tasks and functions that must be fulfilled by the individual role-occupant and b) that these role-specific tasks and functions provide the individual role-occupant with *role-specific reasons for action*. Identifying the framework of beliefs, desires, and goals with these *role-specific reasons for action* illuminates how the process of perspective taking can indeed be applied to institutional roles.

It follows, however, that my notion of *role* perspective taking must strictly be separated from the notion of perspective taking employed in the debates in the Philosophy of Mind, including *ToM*, or *empathic understanding*. If having a perspective means to "to perceive the world from a particular spatiotemporal location" and thus to be the "the origin of a perceptual field" (Baker 2012, 20f.), then institutional roles will not provide any appropriate perspective which could be taken by an individual. The capacity to mentally model, or simulate other *persons*' beliefs, desires, goals etc., is not applicable to institutional roles because, again, institutional roles are neither persons nor do they have mental states.[126]

Let me re-emphasize that R_{PT} is to be understood in a minimalistic way. My notion of R_{PT} should be strictly understood as a technical term and I do not wish to claim that such modification of individual agency consists in a transformation of the visual, or sensual experiences of those who take role perspectives. Nor do I wish to claim that it requires a special, irreducible mode, or that such perspective taking is

[126] Of course, one might take the perspective of a *person* (in the sense specified by Baker) who *also* occupies an institutional role. But this is not the *role* perspective taking that I am concerned with.

accompanied by qualitative states, or a change in phenomenal consciousness.[127] There might be a *what-it-is-like-ness* encompassing the adaption of institutional role perspectives, but I do not hold my notion of R_{PT} to provide arguments neither for, nor against such a claim. Role perspective taking is primarily a cognitive achievement based on an agent's capacity to form particular intentional states.[128]

How does Role Perspective Taking come about?

So in order to take up the perspective of an institutional role, individuals have to have some form of epistemic access to the framework of beliefs, desires and goals that their roles provide. They also have to know what the *role-specific reasons for action* of their institutional roles are. How could we further understand this? Let me start to approximate the process by which perspective taking comes about. We may begin with Schmitz's description of individuals appropriating, "*trying on"*, or "*growing into"* institutional roles they come to occupy. For Schmitz, the ability for taking the perspective of institutional roles rests on an individual's capacity to "switch" between personal, I-mode, and role-mode. In turn, this is fundamentally based on the capacity for deliberation. Schmitz notes that the

> "kind of deliberation of whether to join a group or to adopt a role is essentially different from reasoning and acting from the point of view of a group or role. When engaging in the former kind of deliberation, I may try on, as it were, the attitudes and responsibilities of the new role to see if they fit me; *but when I really adopt them, when I really grow into this role and make it mine, which of course will typically take time—even after I have already legally adopted it—I see the world from the vantage point of the role*" (Schmitz 2023, 196f.) [own emphasis].

This, I believe, is at least a good approximation of the process through which individuals learn to engage in RA, and them coming to adopt their role-perspectives. Adopting the externally defined framework of beliefs, desires, goals etc. that institutional roles provide will usually take time and effort, and cannot be

[127] However, regarding the visual metaphor of perspective taking, it has been argued that occupying certain institutional roles really does transform the visual experiences of individuals. Styhre gives a description of the so called "professional gaze". He argues that medical doctors acquire a professional vision in analyzing functional magnetic resonance imaging (fMRI). Professional vision here describes "a specific and contingent 'way of seeing' that is embedded in professional identities, ideologies, formal training, and everyday work experience" (Styhre 2010, 366), However, the capacity to gain knowledge about a patient through the use of professional vision is not a mere visual process: "[T]here is no 'seeing per se' detached from other embodied practices and procedures. Practices of seeing are always based already on the capacity of saying and gesturing and, in many cases, also on the capacity to make use of other forms of sense perception, e.g. olfactory or audible capacities" (Styhre 2010, 370). See also: Goodwin (1994).

[128] Also, when I claim that an individual is able to adopt a role perspective, I do not wish to make claims about developmental landmarks that underlie an individual human agent's capacity to do so. Role perspective taking may necessarily require the development of certain skills, or cognitive abilities, and very well could coincide with having a robust theory of mind, the capacity for counterfactual reasoning, or joint attention. If I were to be pressed on this, I would say that role perspective taking, similar to the psychological concept of *agent perspective taking*, is something that can be achieved by a neurotypical human adult, but not by infants or chimpanzees. However, I do not wish to answer these questions, nor to speak in favor of any theory of cognitive development here.

achieved instantaneously, as these frameworks can be immensely complex. But the question, which Schmitz's description leaves unanswered, is *how* role-occupants may come to understand and have control over their institutional role perspective in the first place. How could the psychological process of "switching" between role- and I-mode be further analyzed? What psychological mechanisms and processes allow for such appropriation of perspectives that Schmitz refers to? And why is this appropriation *processual*, so that it typically "takes time" to "grow into" an institutional role? Regarding this epistemic dimension of Role Agency, I'll try to provide an answer to these questions by pointing to the mechanism of *Role-Internalization* (R_{IN}). The concept of Role-Internalization aims to provide a more fine-grained description of the "kind of deliberation" that allows individuals to "try on, as it were, the attitudes and responsibilities of the new role" (ibid). In analyzing these processes of appropriating role-perspectives, I'll argue - *pace Schmitz* - that these mechanisms are not solely theoretical, or *deliberative*. Rather, appropriating role-perspectives encompasses both a theoretical dimension of "knowing that" *and* a practical form of knowledge, which - with Ryle (1949) - we might call a form of *"knowing how*". We further have to distinguish a *formal* and *informal* dimension of appropriating role-perspectives. I'll argue for the usefulness of keeping these dimensions apart.

5.2. Role-Internalization

In sociology and social psychology, the term *internalization* is usually used to describe a number of processes through which external things like rules, norms, beliefs, etc. are transformed into internal representations of these things (cf. Lizardo 2021; see also, Dahrendorf 2010; Merton 1957).[129] Likewise, my concept of *Role*-Internalization aims to explain the processes by which individuals come to transform the externally defined framework of rules, norms, tasks, functions, etc. of the institutional roles into the above mentioned internal role-perspective. We may get an initial idea of what Role-Internalization (henceforth: R_{IN}) is about by looking at the difference between the ways in which *experienced* role-occupants perform their tasks, and the ways in which *novices* (try to) do the same.[130]

Consider, again, the example of a fire-fighter. From a novice role-occupant's perspective, an expert fire-fighter's actions may seem routined, easy, effortless, almost happening without thought. Expert fire-fighters might provide instant solutions for what novices fail to recognize as a problem in the first place. Where novices panic, expert firefighters may remain calm and collected. In turn, an expert might look at the novice's actions and hold them to laborious, clumsy, and hesitant. We might simply say that the difference between experts and novices consists of the former having *experience*. This surely is correct. But experience of *what*? And what exactly happened during the process of a novice *becoming* an expert? I hold my concept of R_{IN} to provide a more fine-grained explanation for this difference between experts and novices than simply referring to experience.

[129] Having a long history in sociology and social psychology, especially influenced by classical authors like Durkheim, Freud and Parsons, or Mead, the term "internalization" has been a *contested concept*, being used to describe different things at different times (see: Lizardo 2021).

[130] Of course, novice-expert-relations are not confined to institutional roles, as one, e.g., can be a novice or expert in water-skiing or rope-jumping, where this sort of activity is not connected to any particular institutional role. But institutional roles nevertheless do seem to provide a clear example for this distinction.

What I aim to capture with R_{IN} is an explanation for how the framework of rules, norms, etc. that governs an institutional role is appropriated by an individual taking up the role-perspective; and how, and in which ways, such a framework becomes efficacious within the individual. Thus, my concept of R_{IN} seeks to illuminate the ways in which individuals acquire the knowledge, skills and capacities necessary to perform the tasks and functions of institutional roles.

Before we get started, notice that R_{IN} should not be understood as a single, universal process that equally applies to every individual who comes to occupy any institutional role. Instead, I want to describe with it a dynamic (rather than static) phenomenon that comes in different forms and shapes, depending on which institutional role is to be internalized. So I want to invite the reader to think of it as a set of *gradual* and *non-linear* processes. As to non-linearity, R_{IN} will typically begin at the time when an individual joins an institutional group, but it may continue throughout the individual's membership in the group, or even occur *prior* to an individual joining. But because institutional roles can change over the course of time, R_{IN} can also require constant monitoring and updating from the individual.

The process of R_{IN} helps us to answer questions regarding the epistemic dimension of RA i.e., how role-occupants may come to understand and have control over their institutional roles. It useful to first analyze the concept of R_{IN} along two broad dichotomies: along a *theoretical (*or *knowledge-based)* and *practical (*or *application-based)* dimension; and along a *formal* and *informal* one. We can derive from this four different aspects of Role-Internalization: 1) formal-*theoretical* R_{IN}, 2) formal-*practical* R_{IN}, 3) *informal*-theoretical R_{IN} and 4) *informal*-practical R_{IN} (see Figure 1 below). Let me briefly sketch each of these aspects before continuing with a more detailed description of R_{IN}.

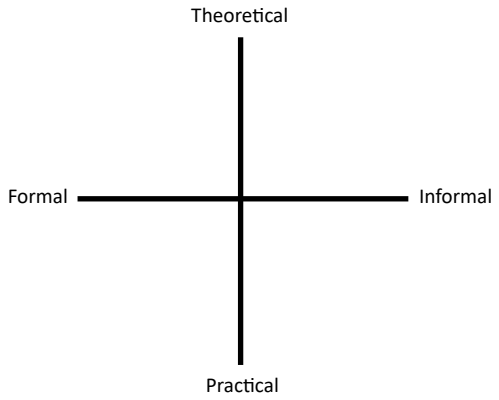

Figure 1: The four dimensions of Role-Internalization

1) formal-theoretical Role-Internalization

Regarding its formal theoretical dimension, R_{IN} may include the acquisition of formally defined knowledge (i.e., the "official facts") related to one's institutional role. The primary way in which individuals internalize their roles in this formal-theoretical dimension is through the acquisition of the formally defined knowledge regarding their role-specific tasks and functions as established by the role's *design-specification*.

I hold this to be an essential aspect of R_{IN}. When individuals internalize their institutional roles, they first come to know their tasks and functions, and their corresponding deontic powers such as the rights, duties, responsibilities etc. of their roles. Formal-theoretical R_{IN} provides the individuals with access to the *role-specific reasons for action* that their roles provide*,* at least insofar as their institutional role is defined through such tasks and functions. This is a necessary pre-requisite for R_{PT}. To reason, intend, and act based on the framework of beliefs, desires, and goals of one's institutional role presupposes that this framework has to be -in some way or the other- epistemically accessible to the individual.

However, the formal-theoretical dimension of R_{IN} does not exhaust itself in the acquisition of knowledge about the tasks and functions of institutional roles. The theoretical-formal dimension may also include acquisition of knowledge about things like the group's formal *structure,* the *power relations* and *hierarchies* (i.e., knowledge about the formal relation between one's own institutional role and others, about the connection and interrelation between one's tasks and functions and those tasks and functions other role-occupants fulfill), the group's communication-channels and decision-making procedures, its legal framework and possible legal ramifications, the group's history, its embeddedness in larger contexts, etc.

Institutional groups may exhibit certain, distinct mechanisms through which such theoretical (or educational) R_{IN} is provided for their members. But because different institutional groups may employ different mechanisms, I do not wish to give a definite answer to the question as to which *exact* underlying mechanisms enable this form of R_{IN}. This, I hold, cannot be answered from the philosophical arm-chair. Rather, I want to identify some structural commonalities that I hold to cover a range of such mechanisms. Notice that the *formality* of this theoretical R_{IN} usually indicates a level of *documentality*, on the basis of which institutional groups can disseminate such formal-theoretical facts to their (prospective) members. This can, e.g., take the form of written legal contracts, occupational handbooks, bulletins, constitutions, manuals, organizational flow-charts, mission statements, or codes of conduct, etc. Such *documented* facts about the group's structure and the role's design-specification are particularly apt to function as mechanisms for theoretical R_{IN}, because they allow for the dissemination of knowledge in a *one-to-many* relation of communication. Once a design-specification of an institutional role is formally documented in such a way, it can be provided for - and consequently internalized by - different individuals occupying this role. This allows for the design-specifications of institutional roles to remain diachronically robust through the potential change of membership.[131] Institutional groups, however, may also exhibit mechanisms that allow the individual to acquire such knowledge in ways that do not rely on written language or documents, e.g., through standardized onboarding-processes, educational programs, instructional or compliance courses, ceremonies, and rites of passage, etc.

2) formal-practical Role-Internalization

As to the formal practical dimension of R_{IN}, such processes do not include the acquisition of facts or knowledge etc. about one's role but rather the acquisition of role-specific practical skills, techniques, a

[131] It is, however, worth emphasizing the non-linear character of such Role-Internalization, as an institutional role's tasks and functions may change throughout time, possibly as a result of adapting to new environmental changes. I will say more about this below.

savior faire, or technē. Recall that individuals need to be able to both interpret *and* apply their formally defined institutional role in order to carry out the task and functions. The practical dimension of R_{IN} concerns the application-side, i.e., it regards an individual's capacity not only to theoretically understand and interpret her role, but rather to apply, control and *act* on it. This practical dimension of R_{IN} can be captured by Ryle's (1949) distinction between two models of knowledge, i.e., it's a form of practical knowledge (*knowing how to perform a task)*, and not a form of theoretical knowledge (*knowing that such and such is the case*). The distinction, albeit contested,[132] is helpful, partly because it isn't intuitively plausible to assume that individuals are able to apply their role's tasks and functions on the onset simply by coming to theoretically understand what this requires them to do. Instead, institutional roles (e.g., surgeons, stonemasons, professors, etc.) may require extensive amounts of practical training and exercise to perform, let alone to "master" the tasks and functions.

By saying that this form of practical internalization is *formal* in nature, again, I claim that it presupposes the acquisition of such practical knowledge to be part of the *institutional role's design-specification.* Keep in mind that the design-specifications of institutional roles do not only determine *what* tasks and functions are to be performed by the individual occupant, but that they also determine - at least to some degree - *how* these tasks and functions are to be performed. Again, the exact mechanisms of institutional groups to provide the individual with the means necessary to acquire such practical skills and, on this basis, to achieve such practical R_{IN}, may ultimately vary. Thus, institutional groups may exhibit a variety of such mechanisms. Some may, e.g., consist of "on the job" training-processes, internships and mandatory work probes, mentoring- and coaching-programs, or so called *job-shadowing*, i.e., observational learning from the actions of experienced group-members by a novice.[133]

The *formality* of such practical R_{IN} also pertains to the fact that such mechanisms may be *standardized* in some way, and on this basis *equally applied* to different individuals occupying the same institutional role. The standardization of mechanisms by which individuals come to practically internalize their institutional roles, too, may vary. But usually, to perform *standardized* tasks implies that the content, sequence and

[132] Like anything else in philosophy, Ryle's distinction between these two forms of knowledge is contested. So let me simply assume that there *is* such a distinction. For a brief overview of why intellectualists and anti-intellectualists can't agree whether there really are these distinct forms of knowledge see: Tanney 2021.

[133] This is not the only function of such mechanisms, as I will argue in the next section on Role-Idealization

outcome of those actions are identifiable and on this basis can be somehow *measured* along a pre-defined metric.[134]

3) informal-theoretical Role-Internalization

When we try to describe how individuals come to understand and apply their institutional roles, we have to consider not only the ways in which the formally documented design-specifications of institutional roles play into the exercise of the tasks and functions of role-occupants. We also have to consider how *informal* rules, norms or practices of institutional roles govern the conduct of individuals. Reflecting on this *informal* dimension of R_{IN} allows us to understand how individuals perform the tasks and functions of their roles in ways that we otherwise could make little sense of. For further clarity, I'll distinguish an informal *theoretical* dimension of Role-Internalization from an informal *practical* one.

The *informal* theoretical aspect of R_{IN}, just like the formal-theoretical dimension, concerns the acquisition of knowledge and facts related to one's institutional role. It, too, is a form of theoretical knowledge *that such and such is the case*. However, what needs to be internalized is a form of theoretical knowledge which cannot be accessed from the institutional role's *formally defined*, or officially documented features, e.g., through contracts, organizational charts, handbooks, or job-descriptions.

But just because such a form of knowledge is "nowhere to be found" within an institutional group's official documents, contracts, mission statements, etc., that does not mean that it *does not* pervade in institutions.[135] For our purpose, we can broadly distinguish between *informal means of disseminating knowledge* (through coffee-breaks or hallway meetings) and the dissemination of *informal* knowledge (e.g., knowledge about a group's c*overt* decision-making procedures; networks of camaraderie, alliances and enmities; its covert marketing strategies, etc.). Both should be considered part of informal theoretical R_{IN}. The informal aspect of R_{IN} may also require the individual to gain knowledge about the above mentioned

---

[134] For contemporary theories of standardization see especially Lampland & Star 2009. Also: Ungan 2006; Higgins & Larner 2010. For wider implications of such standardization see George Ritzer's "The ‚McDonaldization' of Society" (1983). For an original theory of how to standardize the practical dimension of institutional roles in the context of industrial work, see F.W. Taylor's (1911) "The Principles of Scientific Management". Taylor's way of describing standardized work-performance (which came to be know as *Taylorism*) is particularly interesting here. It can be roughly summarized by the so called "task idea", which is somewhat connected to my own idea that the standardized application of one's tasks and functions requires its practical internalization. He writes: "Perhaps the most prominent single element in modern scientific management is the task idea. The work of every workman is fully planned out by the management at least one day in advance, and *each man receives in most cases complete written instructions, describing in detail the task which he is to accomplish, as well as the means to be used in doing the work.* […] *This task specifies not only what is to be done but how it is to be done and the exact time allowed for doing it.* […] These tasks are carefully planned, so that both good and careful work are called for in their performance, but it should be distinctly understood that in no case is the workman called upon to work at a pace which would be injurious to his health. The task is always so regulated that the man who is well suited to his job will thrive while working at this rate during a long term of years and grow happier and more prosperous, instead of being overworked. Scientific management consists very largely in preparing for and carrying out these tasks" (Taylor 1911, 25) [own emphasis].

[135] The empirical literature of *organizational learning* (e.g., Stohl & Redding 1987; Akgün et al. 2003; Loon & McShane 2010) widely acknowledges the fact that organized groups rely on both formal *and* informal processes through which employees of organizations acquire and share knowledge, e.g., through so called "water-cooler-discussions" (Akgün et al. 2003, 849).

gray areas of performing one's rule-governed tasks and functions, including understanding the actually *prioritized* or *trivialized* aspects of one's institutional role. Or it may require the individual to learn about which kinds of pro-social rule-breaks and informal deviancies from formal rules are *tolerated*, and which ones are *sanctioned* by other group members.

Again, the exact content and nature of such informal knowledge may not be determinable from the philosophical armchair. Yet, what is more interesting for our purpose is to look at the *functional* relevance of the *informal* theoretical dimension of R_{IN}. As argued above, such informal knowledge proves to be necessary for individuals in using their discretionary powers to apply the tasks and functions of their roles *in situ*. Thus, the informal theoretical dimension R_{IN} figures prominently in explaining how individuals are able to navigate the complex and often contingent institutional group's actual operational practices. When, e.g., different tasks and functions of one's role may not be realizable at the same time (recall the scenario of CONFLICT in Ch. 4.1.), the internalization of informal aspects of one's role may help an individual to weigh conflicting tasks against each other.

Consider a police-woman, who learns about her department's *informal* policy to let minor demeanors of individuals "slide" as to not disturb a brittle public peace in the community. Although her role's *official* design-specification would require her to investigate *every* misdemeanor, such a police-woman, on having acquired knowledge about such informal decision-making procedures, may prioritize the long-term goal of upholding social peace over the short-term goal of, e.g., catching petty thieves and shop-lifters. Or consider the difference between knowing only about an institutional group's *official communication-channel* and knowing about both the *official* and *informal* ways of communicating within an institutional group. Knowing that one's colleague is a struggling alcoholic is not something that can be learned from the onboarding-manual. But such informal theoretical knowledge (*that* such and such is the case) may be necessary to be able to effectively cooperate with her. Or think, as another example, of an individual whose task it is to organize her department's official distribution of tasks. The individual might try to convince her superior to change the official distribution of tasks within her department, because she holds the current distribution to be ineffective. Said individual may find out that it is more productive to communicate this inquiry through informal ways (e.g., over lunch), than to "go on the record" during one of the department's official meetings, thereby risking to both expose her superior's lack of organizational prudence *and* undermine her authority on front of others. In turn, the eventual *failure* to convince her superior to change the official distribution of tasks by going *on the record* in an official meeting may become intelligible if we know that within the department, such matters are to be discussed through informal channels of communication. In such a scenario, the *lack of informal-theoretical* R_{IN} (that such and such is the case regarding decision-making procedures) figures in explaining how failure to carry out one's role-specific tasks came about.

4) informal-practical Role-Internalization

Lastly, the informal practical dimension of R_{IN} concerns the "shortcuts", bypasses, creative workarounds, temporary arrangements and provisional makeshifts of *applying* one's tasks and functions. Recall again the above mentioned example of workers at the Caterpillar plant (Ch. 4.2.), which are - by the formal design specifications of their institutional role - officially required to call in the repair-clerk when certain machines

break down, but where the workers, in order to save time, have internalized the *informal practice* of fixing smaller problems of the machines by themselves. Alternatively, think about the way in which employees in fast food restaurants evade and undermine their Fordistic work-regimes by "cutting corners", e.g. by telling customers that the ice-cream machine is broken when in fact, the costs of working extra due to the long-lasting cleaning-procedure outweigh the benefits of selling ice-cream to customers.

Again, such *informal* practical applications of one's role proves to be important for individuals in using their discretionary powers to navigate the complex and contingent institutional group's *actual* operational practices, as opposed to the formally established practices specified by the role's official design-specification. And it is on this basis, that they can shield their institutional role from becoming paralyzed by the formally established rules and procedures which don't match the contingent circumstances under which the role's tasks must be applied *in situ*. The informal-practical dimension of R_{IN} allows us to see how the workers at the Caterpillar plant don't simply fail to perform their tasks and functions when they break *official protocol*. Instead, it can explain how such violation of officially established rules is actually an *informal part* of their institutional roles.

Keeping the theoretical-practical distinction apart

What is the distinction between these different dimension of R_{IN} useful for anyway? Keep in mind that ultimately, the concept of R_{IN} wants to answer how role-occupants may come to understand and have control over their institutional role perspective in the first place. The following two sections highlight some reasons to keep apart the different dimensions in the analysis of R_{IN}, and on this basis, develop a more fine-grained description of how role-occupants may achieve this form of perspective taking.

The first reason to keep the dimensions apart is that, depending on the nature of a given institutional role, either practical or theoretical aspects of carrying out a role's tasks and functions may be more (or less) important for engaging in RA. It is therefore worth distinguishing these two aspects from one another, as they can, and often do come apart. A good example demonstrating such a disconnection between theoretical and practical internalization can be found within the military. Consider the task of U.S. soldiers (during WWII) to *march in formation*. This task is an explicit part of the design-specification of a soldier, i.e., it's *officially documented* in the so called "*basic field manual*" (or "Soldier's handbook"). Now the seventh chapter of this basic field manual, called "*The School of the Soldier without Arms*" exhaustively specifies the ways in which soldiers have to conduct themselves when participating in military marches. Take the example of the marching style of "Quicktime":

> "QUICKTIME.—Being at the halt the commands to move forward in quick time are: 1. FORWARD, 2. MARCH. At the command FORWARD, you shift the weight of your body to the right leg without making any noticeable movement. Do not start to move forward. At the command MARCH, step off smartly with your left foot and continue to march with 30-inch steps straight to the front, at the rate of 120 steps per minute. You do this without stiffness and without exaggerating any of the movements. Swing your arms easily and in their natural arcs, 6

> inches to the front and 3 inches to the rear of your body" (United States Department of War 1941, 85).

I hold the example of *marching in Quicktime* to intuitively show that the formal-theoretical internalization (i.e., to read and memorize the manual) alone does not suffice to explain how role-occupants achieve carrying out their tasks and functions. Having the theoretical *knowledge that* marching in Quicktime consists out of the above describes steps by no means guarantees to provide the soldier with the means necessary to perform this particular task, because *marching in Quicktime*, or to "swing one's arms easily" and "without stiffness" is primarily a *practical* skill. Thus, individuals need to *practically* internalize (and not just theoretically understand) *how to do this*. Thus, in order to perform the task of marching in *Quicktime*, practical internalization of this task, i.e., extensive practical training (the so called *drill*) is required. Although she might have all the relevant theoretical knowledge, a soldier's failure to perform the task of marching in *Quicktime* thus can be explained by pointing to a lack of practical internalization on her part (in turn, we might ask why military units would see extensive practical training processes, i.e. *Boot Camp* as necessary to begin with, if it would suffice to simply tell the novice soldiers to read a book).

Keeping the formal-informal distinction apart

It is worth keeping the formal and informal dimension of Role-Internalization apart too. One way this distinction between formal and informal aspects of institutional roles has been dealt with in the literature, is to discuss it in the context of "institutional culture". To this end, Seumas Miller, states that:

> "Aside from the formal and usually explicitly stated, or defined, tasks and rules, there is an important implicit and informal dimension of an institution roughly describable as institutional *culture*. This notion comprises the informal attitudes, values, norms, and the ethos or ‚spirit' which pervades an institution. […] *Culture in [this] sense influences much of the activity of the members of that institution, or at least the manner in which that activity is undertaken*. So while the explicitly determined rules and tasks might say nothing about being secretive or ‚sticking by one's mates come what may' or having a hostile or negative attitude to particular social groups, these attitudes and practices might in fact be pervasive; they might be part of the culture" (Miller 2019, Sec. 1) [own emphasis].

I generally agree with Miller that one has to acknowledge the importance of such informal attitudes, values, norms, etc. However, claiming that such informal attitude etc. simply "pervade" institutions does not answer the question as to how, and why these informal norms etc. actually influence "much of the activity of the members", or "the manner in which that activity is undertaken" (ibid). Here, I hold it to be useful to look at the ways in which the *internalization* of one's role may be influenced by such informal attitudes, values, norms, etc, and also how individuals *mutually influence* each other's activities on the basis of such internalization.

Again, it's worth emphasizing that it might depend on the very nature of a role-occupant's tasks and functions, whether the formal or informal dimensions of internalizing one's role may be more (or less) important. On this basis, we may distinguish different (idealized) kinds of institutional roles which require different emphases of either formal or informal Role-Internalization by the individual.

One the one hand, one can think of institutional roles that are defined through heavily formalized tasks and functions, but which require for their exercise little or few informal rules, or unofficial ways of application. We might call these heavily formalized institutional roles (e.g., accountants, analysts, data entry operators or insurance workers) *bureaucratic roles*. To exercise such *bureaucratic roles*, individuals primarily have to internalize the formal (both theoretical and practical) aspects of their institutional role. On the other hand, there are institutional roles, which we might call *troubleshooter roles*, for which the exact opposite holds: To carry out tasks and functions that are little formalized, but which require for their exercise many different, informal, *ad hoc* ways of application. Think, for an example, of Repair Service Technicians, whose jobs consist not in the steady exercise of pre-defined tasks, but rather in providing help for spontaneous, and contingent problems, i.e. so called *MacGyvering.*[136] To exercise such *troubleshooter roles*, individuals primarily have to *circumvent* official protocol, rather than following it, i.e., they may be primarily required to work out informal "shortcuts" and provisional makeshifts of application.

Another reason to keep the informal and formal dimension of R_{IN} apart, is that the informal and formal aspects can both be *symbiotic* and *antagonistic*. The discussion of institutional stupor already gave us a good understanding of such a symbiosis between formal and informal R_{IN}. Recall the cases where role-occupants actively utilize the informal dimensions of their institutional roles in order to shield the institutional role from becoming paralyzed by formally established rules and procedures. I call this sort of relation between the formal and informal dimensions of institutional roles *symbiotic*, because it is only through the *combination* of both dimensions, that role-occupants achieve the envisioned outcome. If it was not for the informal practices and procedures, the formal tasks and functions would not be performed, or at least the performance would be impeded. In turn, the use of such informal practices becomes intelligible only in light of the formally defined tasks and functions that they operate on to begin with.

However, the formal and informal dimensions of institutional roles do not always have to complement each other this way. A good example of the antagonism between the formal and informal dimension of institutional roles might the distinction between formally established and officially endorsed "Police-Culture" and an informal, interpersonal "Canteen-" or "Cop-Culture", which develops out of the local occupational conditions of police-men and -women (see, e.g., Westmarland 2008). The latter "Cop-Culture" may see for informal practices (e.g., pledges of secrecy regarding the misconduct of colleagues, racial profiling, turning off body-cams) that are incompatible with, and undermine the formally established and officially endorsed "Police-Culture" (to be law-abiding, to live up to democratic ideals of equality, to be a civil servant and therefore accountable to the public). In practice, individuals occupying the role of policemen and -women then may experience an *intra-role* conflict as to which expectations to live up to (see for analysis of this form of role-conflict: Dahrendorf 2010). The antagonistic nature of institutional roles

[136] The Oxford English Dictionary defines the verb "MacGyver" the following way: "To construct, fix, or modify (something) in an improvised or inventive way, typically by making use of whatever items are at hand; to adapt expediently or ingeniously" (Oxford English Dictionary 2023)

may also lead to *inter-role* conflicts, i.e., to conflicts between institutional role-occupants. An individual police-women who, e.g., refuses to engage in the informal practices of her colleagues (e.g., racial profiling, witness-tampering etc.) sooner or later might find herself working the graveyard shift, and being avoided by her colleagues. It is by recognizing both the informal and formal aspects of her institutional role by which such conflicts become intelligible. But such an antagonism between informal and formal rules may run in the other direction as well. Formally established ways to perform the tasks of a role may be actively installed and endorsed (instead of undermined) in order to counter-measure informal practices and procedures. Take the example of institutional "boys clubs", i.e., the informal, yet pervasive practice of promoting and giving patronage only to male members of an institution, to the disadvantage of all other group members (See: van Dijk 2023). While such practices may be part of the institutional group's *informal culture*, their effects may be remedied (or at least mitigated) by establishing, or re-emphasizing officially endorsed and formally defined gender-neutral guidelines and procedures of promotion, which role-occupants must adhere to.

Role-internalization as a dynamic and non-linear process

Let me now further characterize the process of Role-Internalization. The following reasons speak in favor of conceptualizing R_{IN} as a dynamic and non-linear process: First, R_{IN} is a dynamic process because institutional roles, too, are dynamic. Recall that the member-to-member relations of power between roles allow for the design-specification of institutional roles to be modified by other role-occupants. One's boss, e.g., may decide to re-allocate human resources, and on this basis re-design the division of tasks among her subordinates. Consequently, role-occupants may have to hand over certain tasks to other co-workers and, in turn, gain new ones. The fact that institutional roles are dynamic in this sense (i.e. they may gain or lose certain tasks and functions) requires to model the process of R_{IN} itself as dynamic too. And R_{IN} may require not only the renewal and ongoing updating of theoretical knowledge of one's *own* role, but it may also require the renewal and ongoing updating of one's theoretical understanding of how one's role *relates to both the other institutional roles,* as well as *other roles' tasks and functions*. This, of course, applies to both the theoretical and practical dimension of R_{IN}. If, e.g., changes in the division of labour within a factory require a worker to operate new machines, she may have to acquire the practical knowledge of how to operate them; if new technologies emerge, role-occupants may have to adapt and update their technical know-how, etc. Further, this potential change of the design-specifications of one's role travels to the *informal* ways in which roles are to be understood and applied. New tasks and functions of her role may require the individual to gain knowledge about the *gray areas* within these tasks can be performed; and new rules require an understanding for the informal ways in which breaking them may be tolerated. If one's task is to operate new machines, this may require the individual to learn new informal workarounds concerning their safety-protocol.

Second, not all institutional roles will require the individual to internalize each of the four dimensions equally (see Figure 2 below). As I mentioned, some institutional roles may require less theoretical knowledge about one's position, tasks or their legal ramifications, etc. but more practical knowledge about the application of these tasks and functions. Take the example of working in a *brigade de cuisine*. While the

formally defined tasks and functions of a *Pâtissier* might be theoretically relatively clear-cut and comprehensible (making deserts), their application can depend on a multitude of contextual and situational factors. So an individual might have already theoretically internalized her role in the kitchen brigade relatively well, but still needs to *practically internalize* her role, i.e., she must yet acquire the practical know-how to apply the tasks to different contexts and instances (i.e., accommodating different demands from customers, such as allergies or religious dietary restrictions; maintaining consistency of the dishes throughout the varying quality of the raw materials delivered; managing the varying quantity of dishes that need to be served; resolving issues like equipment-failure or shortages of ingredients, etc.). In such cases, R_{IN} is dynamic in the sense that one dimension of one's role might be more thoroughly (or deeper) internalized than the other.

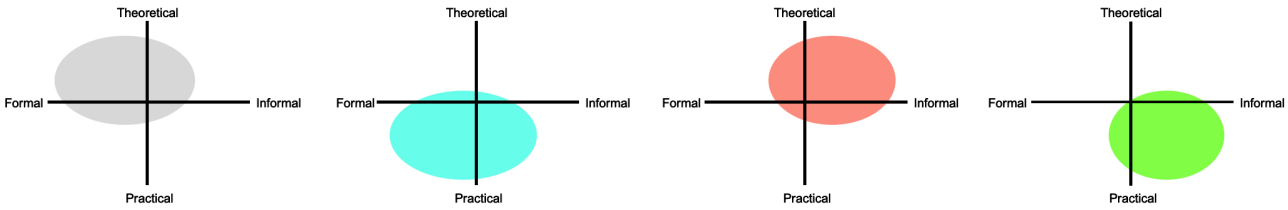

Figure 2: Four possible profiles of Role-Internalization

A third reason to think of R_{IN} as dynamic, *non-linear* process is that R_{IN} can be temporarily disrupted and kept on different levels. To claim the non-linearity of R_{IN} immediately evokes an important question: Is there such a thing as an upper limit of R_{IN}? And is there a *lower* one? Regarding this, I hold that it is best to think of Role-Internalization to exist on a scale that is *upper-open* but *lower-closed* (cf. Fassio & Logins 2023, 2055). The claim here is that there exists a necessary minimal threshold of R_{IN}, but not necessarily an upper limit.[137] One reason *not* to think about R_{IN} as a steady, and linear progress of "growing into one's role", but rather as a non-linear process with a minimum threshold, is that Role-Internalization might be temporarily disrupted, or even impeded. Also, R_{IN} may be non-linear in this sense because it might be *intentionally kept* so. Some individuals, due to the institutional group's action being compartmentalized and anonymized, might be intentionally kept at a level of "minimal" internalization of their tasks and functions. This minimal level may consist in knowing only enough to exercise one's tasks, without, e.g., knowing about the ways in which one's tasks contribute to the overall group's goal. Think again of the Calutron Girls, who were intentionally kept "in the dark" by their superiors about the group's structure they were part of and about

[137] Role-internalization then has a *minimal threshold* and every individual who internalizes the different aspects of her role past this minimum threshold can, *ceteris paribus*, be said to have internalized her role. When we consider the category of individuals to which "has internalized her role" applies, there are necessary conditions for belonging in this category. I think the most plausible candidates for such necessary conditions consist in a) that individual knowing that (or: being aware of the fact that) she occupies an institutional role and b) her knowing that (or: being aware of the fact that) the institutional role she occupies encompasses certain tasks and functions that need to be performed. Once this minimal threshold is crossed, individuals can be said to have internalized their roles, although the individual level of internalization beyond this threshold will vary, e.g., from novices to experts.

the group's goal they were contributing to. For the Calutron Girls, R_{IN} was purposely kept at a minimum through clearance levels and mandatory secrecy.

This view of R_{IN} also implies that the process of R_{IN} can be based on false assumptions, and that individuals can *misrepresent* what they *actually* contributes to a group action, or how their tasks and functions *actually* figure in bringing about an institutional group's goal. This may be due to one's own false beliefs about the functional integration of one's tasks and functions. But it might even be because one's role is *intentionally misrepresented* by one's superiors. To this latter end, David Graeber's book on *Bullshit Jobs* recites a real-life case of someone being hired as a secretary for a publishing company. The official tasks and functions of this institutional role were to 1) answer phone calls, which happened "maybe once a day", 2) to keep a candy dish full of mints, and 3) to wind up a grand-father clock once a week (cf. Graeber 2018, 30f.). The secretary, Graeber notes, was surprised to later find out that she suffered from a completely false understanding of what her work *actually* contributed to, or what functions her tasks and functions *actually* played in bringing about her institutional group's goal:

> "Why did the Dutch publishing outfit need a receptionist? Because a company has to have three levels of command in order to be considered a ‚real' company. At the very least, there must be a boss, and editors, and those editors have to have some sort of underlings or assistants - at the very minimum, the one receptionist who is a kind of collective underling to all of them. Otherwise you wouldn't be a corporation but just some kind of hippie collective. Once the unnecessary flunky is hired, whether or not that flunky ends up being given anything to do is an entirely secondary consideration" (Graeber 2018, 35).

For Graeber, the secretary occupied what he calls a "flunky" role, which is a role that exists "only or primarily to make someone else look or feel important" (Graeber 2018, 28f.). But notice that the fact that the individual was hired exclusively so that her superiors could appear important was intentionally hidden from her. Thus, she was contributing to a - dubious - goal, that she was actually unaware of when she was fulfilling her officially prescribed tasks and functions. Yet, we can assume that she internalized her role as secretary (i.e., we do *not* have to assume that she *didn't* undergo the process of R_{IN}) despite her misrepresenting the ways, in which her tasks *actually* contributed to an overall group's goal, which, I think, is not a necessary condition for R_{IN}.

What is the scope of Role-Internalization?

At this point in the analysis, I want to address a problem concerning the notion of the "externally defined beliefs, desires, and goals" and the way in which this notion functions in my own account of R_{PT} and R_{IN}. Above, I argued that in order to take up the perspective of an institutional role, individuals have to have some form of epistemic access to the *role-specific reasons for action* that their institutional roles provide. I argued that these role-specific reasons for action, in turn, can be cashed out in terms of a framework of beliefs, desires and goals that individuals need to adopt in order to act *qua* their institutional role.

But if we accept this, then the question must be answered as to which extent individuals will need to be able to have access to these frameworks, and what such knowledge about one's role-specific reasons for action actually entails. One key question here is whether we should think of the Role-Perspective, i.e., the framework of externally defined beliefs, desires and goals to be confined only to the tasks and functions that are specific to *one's own* institutional role, or if we should think of them to also relate to the tasks and functions of *other* role-occupants. Both options, at least initially, seem to entail certain problems.

If we, e.g., think of the framework of externally defined beliefs, desires and goals to be *not* only locally confined to one's institutional role, but to also relate to the frameworks of other role-occupants, then how could this be harmonized with cases of isolated, or compartmentalized roles? Above, the discussion of the Calutron Girls seemingly displayed a case where role-occupants were ignorant not only about the institutional group's structure, plans and goals, but also about how their own roles viz. tasks and functions were related to the other group members' roles, tasks and functions. If, e.g., the goals of an institutional role that have to be internalized must relate to *other* role-occupants, then how could the Calutron Girls be said to have internalized their roles and thus adopted their roles' perspectives in the first place?

If, on the other hand, this was not to be the case and the internalization of one's role would require only the adaptation of the externally-defined goals connected to one's *own* role, then my account faces the risk of being inconsistent: Why, one could ask, should we think of institutional groups to consists of an embodied (or realized) structure of *interrelated* roles, if the individuals internalizing such roles and adopting their roles' perspectives do not actually need to grasps the relations that hold between them?[138]

To answer this question, we should first remind us of the idea that roles come in a variety of shapes and forms. The idea that I want to endorse here is that within institutional groups, it might both be the case, all at the same time, that some individuals do have to know about the way in which their roles relate to others in order to be said to have internalized their roles *and* that some individuals will be in partial or complete ignorance about this. In this sense, R_{PT} and R_{IN} should not be thought to be one-size-fits-all processes that equally apply to every institutional role whatsoever. Instead, I will now argue that, depending on the very type of role that one is to give an explanation for, it can be both the case that, for one type of role, the framework of externally defined beliefs, desires and goals can be confined only to the tasks and functions which are specific to a given institutional role, whereas for other roles, this framework might have to relate to the tasks and functions of *other* role-occupants too.

I already gave ample reasons to think that roles can clearly be interconnected in the sense that adopting one's role-perspective will require some form of knowledge about the way in which the role relates to other roles within the group. For such roles, internalizing the goals relevant to exercise one's own tasks and functions will necessarily involve the individual knowing about the ways in which one's own tasks and functions relate to those of other group members. Take, as an example, roles which are hierarchically connected in the sense that such a role's task consists of *the management of other*, subordinated institutional roles. Managing roles which require the individual to alter, change, or re-design the design-specifications (specifying the tasks and functions) of *other* institutional roles necessarily require knowledge by the individual about these roles that one is supposed to manage in the first place. It would therefore be

[138] I want to thank Katja Crone for bringing this problem to my attention as well as for fruitful discussions about how to solve it.

strange to say that one can internalize the role-perspective of a managing role, e.g., a football coach without knowing how and to which extent this role relates to one's subordinates, e.g., the football team's players. But within institutional groups, there may very well exist roles which are relatively isolated and can be exercised without knowing much (or anything) about the relationships that hold to other role-occupants. Regarding such *stand-alone*, *solitary* roles, it may indeed suffice to exercise only the tasks and functions which are specific to a given institutional role, and these tasks and functions may not require knowledge about the ways in which one's role relates to other roles within the group in order to successfully execute them. The Calutron Girls could be said to occupy a type of role, which allows individuals to operate mostly independently for significant portions of their tasks and functions. Accordingly, in order to take up the role-perspective of a Calutron Girl, knowledge about the ways in which their framework of externally defined beliefs, desires and goals relates to other roles might not have been necessary.[139]

Notice that the claim that there are isolated roles is not inconsistent with the claim that institutional groups consist of an embodied (or realized) structure of *interrelated* roles. Regarding this objection, my answer would be that the roles of the Calutron Girls were *designed by someone else* to encompass as little knowledge about the group's goals as possible. In an important sense, the Calutron Girls did occupy institutional roles that were relational in nature. It is only that the relation between their roles and those roles who specified their design-specification were *asymmetrical* and *hierarchical*. As argued above, role-based accounts of institutional agency fare particularly well in explaining how such hierarchical and asymmetric relations may hold between different roles. Both the member-to-action and member-to-member-relations of power that exist within institutional groups can account for the ways in which the tasks and functions of institutional roles (and thus the role-based reasons for action and the framework of beliefs, desires and goals they give rise to) can be determined, planned, changed or altered by other group members, who reside in the upper (task- or domain-specific) echelons. And once we acknowledge this asymmetrical und hierarchical form of authority within institutional groups, we can explain how such *isolated roles can be designed by others* (and thus be interrelated after all).[140] This explanatory avenue allows us to capture the case of the Calutron Girls as a case of institutional action performed by interrelated role-occupants although the individual Calutron Girls, at that time, did not know from their individual perspective how the tasks and functions of their roles related to those of others.

At this point, one might still be hesitant to say that such roles can be *internalized* in a substantial way, i.e., one might say that the framework of beliefs, desires and goals of their institutional role is not substantial enough to allow for R_{PT} to occur. But notice that my account of RA argued only for a minimal threshold of R_{IN} and that the Calutron Girls can be said to have surpassed this threshold, and thus, to have adopted their

[139] In the case of the Calutron Girls, their ignorance could be described as *not knowing* that such and such is the case, i.e., that their roles are contributing towards building the atomic bomb. For a lack of a better expression, we could say that the Calutron Girls *lack* justified and true beliefs about their roles. But this form of ignorance about one's role is just one way in which an agent could be said to be ignorant of the externally defined framework of beliefs, desires and goals of her institutional role. For the above discussed *flunk-secretary* could be said to be ignorant of her role, too. The ignorance of the secretary, being deceived and misled about the nature of her role, is different than in the case of the Calutron Girls, because the flunk-secretary has *a justified, but false* belief about the nature of her work.

role-perspective. Above, I argued that it is best to think of Role-Internalization to exist on a scale that is *upper-open* but *lower-closed* (cf. Fassio & Logins 2023, 2055). My claim here was that there exists a necessary minimal threshold of R_{IN}, which consists of a) an individual knowing that (or: being aware of the fact that) she occupies an institutional role and b) her knowing that (or: being aware of the fact that) the institutional role she occupies encompasses certain tasks and functions that need to be performed. I argued that once this minimal threshold is crossed, individuals can be said to have minimally internalized their roles, although the individual level of internalization beyond this threshold will vary. If that was to be the correct analysis, then it's plausible to assume that the Calutron Girls could indeed have minimally internalized their roles and the framework of beliefs, desires and goals that their role-specific reasons for action provide for them.

How does deontic power fit in?

Another thing that needs clarification is the relation between Role-Internalization, role-specific reasons for action and the deontic powers, that individuals possess *qua role-occupancy*. The straightforward answer to this question is to differentiate between the *content* of these tasks and functions and the *means* through which an institutional role's tasks and functions are exercised. Deontic powers, as I argued, are *interpersonal powers* that hold between role-occupants which, given a certain context, both *enable* and/or *restrict* the individual in *acting qua their institutional role*. These two dimensions can be captured via the claim that deontic powers can be held in two *modes*, a positive (e.g., in the form of rights or entitlements) and negative (e.g., in the form of duties or obligations) (see Ludwig 2020a, 186). In the literature on role-based accounts of institutions agency, the concept of deontic power primarily figures as an explanans for the ways in which institutional roles relate to one another: If role r_1 has the positive deontic power p over role r_2, then r_2 can, in turn, be said to stand in a negative relation of power to r_1. If, e.g., your boss has the right to order you to carry out certain tasks, then this can be described as a *negative duty* of you to fulfill these tasks. Because roles are *agent-ambiguous* in the sense of being *interchangeable* (or *transferable*) and *iterable,* this form of power is *impersonal* and does not depend on the specific individual who occupies either a *command* or *compliance* role. The deontic relations of power between your role and the role of your boss might remain intact despite a change of membership of behalf of either of you two.

But notice that while deontic powers may regulate the *ways* in which role-occupants may act and relate to one another, they are, in and of themselves, not informative about *what* an agent is to do because of her institutional role. Within your institutional role, you may, e.g., have the duty (i.e., the negative deontic power) to Φ. But the action (Φ) in question is not specified by the *deontic power itself*, but rather relates to the tasks und functions of the role that you occupy. In turn, one and the same action (Φ) that one is to perform may stand in different relations to deontic powers, depending on the role one occupies (i.e., one might have the *right to take part* in an institute's reading-circle or one might have the *duty* to do so). To this end, the *content* of an institutional role's tasks and functions should be understood to be determined by the role-specific reasons for action laid down in a role's design specifications. The deontic powers of an

institutional role then are to be modeled as a specification of the *way* in which these tasks and functions ought to be carried out.[141]

Why Role-Internalization is not the whole story

The process of Role-Internalization provides the basis on which individuals come to understand and have control over their institutional roles. As argued, both theoretical and practical, as well as formal and informal ways of R_{IN} help to explain how individuals come to adapt the perspectives of their roles. And connected to this is an explanation for how role-occupants are able to overcome the first problem of discretionary powers, i.e. institutional stupor. R_{IN} is not *just* about coming to understand the formally defined, and officially documented aspects of performing one's tasks and functions, but it also involves the acquisition of (both theoretical *and* practical) *informal* knowledge. It is on this basis that individuals can shield their institutional role from becoming paralyzed against the backdrop of institutional rigidity and inflexibility, i.e., formally established rules and procedures which don't match the contingent circumstances under which the role's tasks must be applied *in situ*.

But, as with the problem of role-ambiguity, we also saw that discretionary powers can turn *toxic*, especially if an individual herself has no means to assess and evaluate her interpretation of how to contribute to a group's overall goal. So, the consequent question is how individuals come to *reflect* and *evaluate* their expected role-performances, especially in situations where an individual has no means to assess *herself* whether her actions are contributing to a group's overall goal or not. Here, I think that the process of *Role-Idealization* offers useful insights of how individuals arrive at action-guiding principles for preventing their

[141] The relation between role-based reasons for action und deontic powers can be further clarified by stressing that, within large and complexly-structured institutional groups like corporations or governmental agencies, the tasks and functions of active, action-related institutional roles are usually *mandatory*, i.e., that they are held within negative modes of deontic powers in the form of obligations or duties. It is therefore usually the case that the primary way in which job-descriptions, or design-specifications of institutional roles are formulated relies on negative deontic powers. For one, this can simply be explained by pointing at the *hierarchical* structure that such groups usually have. Now for some, it might seem counterintuitive that such hierarchically-structured groups with negative modes of enforcement are the primary ways in which individuals tend to organize themselves. But as Scott Shapiro notes, such a form of authoritative organization is fundamentally a rational way to respond to complexity. For Shapiro, whether submission of an individual to authority is rational, in turn, depends on whether it is rational for an individual to delegate one's planning authority to another (cf. Shapiro 2014, 18). At first glance, such a delegation of power might seem irrational because "we are normally the best and cheapest judges of what we should do" (ibid). But within complexly-structured institutional groups, the submission to authority is rational insofar as it "can conserve precious cognitive resources by deferring to others without risking too much error" so that "we should plan for others to plan for us" (ibid). Related, institutional groups exhibit different mechanisms by which they monitor, sanction or compensate role-occupants for the exercise of these negative deontic powers, e.g., through payment of money or other forms of reward. There are different explanations for why this might be the usual way in which institutional roles operate. One important explanation is that institutional groups, in order to function, are incentivized to mitigate the ways in which their goal-oriented action depend on the intrinsic, personal motivation of their members. Niklas Luhmann, e.g., points out that such internal forms of motivation might not be feasible for all types of institutional groups. A company which produces four-fruit jam, Luhmann argues, will have difficulties to rely on the personal motivation of its members to contribute to such an outcome and therefore might rely on external forms of motivation like, e.g., paying them a salary to do so (Luhmann 1973, 142f.).

interpretation of role-performance turning toxic. The main point I want to make in the next section is that the process of *Role-Idealization* may provide the individual with a *regulative ideal*, or *ideal standard* against which her non-ideal performance can be judged, in order to determine whether an actual course of action of her is contributing to the institutional group's goal. The process of *Role-Idealization*, I will argue, provides the individual with the means necessary to *reason counterfactually* within the perspective of her institutional role. They function as a rail guard for applying one's role and they provide the individual with a quick and easy heuristic to mitigate their doubts in situations of uncertainty.

5.3. Role-Idealization

I will now outline my concept of *Role-Idealization* (henceforth: R_{ID}). This section will proceed the following way: First, I will provide some further insights into how R_{ID} aims to capture another aspect of the relation between individuals and the institutional roles they occupy, which hasn't been discussed yet. Then, I will present my account of R_{ID}. In a next step, I will demarcate my own approach from Kirk Ludwig's (2017a) concept of *prototypical* or *ideal plan-conceptions,* which, in my opinion, does not sufficiently explain how role-idealizations are fixed to an *external* standard. Pace Ludwig, I'll try to show that R_{ID} must encompass such an external, or independent standard in order to remain *regulative* and *action-guiding*. By saying that such a standard is "external to" or "independent from" the individual role-occupant, I simply mean that it can not be determined by the individual alone, i.e. it can not be derived solely on the basis of individual introspection. Finally, I will identify three processes of interpersonal or collective management and monitoring of prototypical role-idealizations that provide individuals with such an action guiding, externally-fixed standard. In this last section, I will also give reasons as to why such forms of R_{ID}, as well as mechanisms that guide and secure it, are important especially regarding the agency for large and complexly structured institutional groups. I'll end this section by discussing a real-life, illustrative example of the collective management and monitoring of role-idealizations.

Role-Distance vs. Role-Idealizations

Let me first address the question how the process of R_{ID} captures an aspect of the relation between individuals and their institutional roles, which has not been discussed yet. Now from what has been established so far, the process of R_{IN} describes the ways in which individuals come to understand and adapt their institutional roles' perspectives. R_{IN}, so to speak, is about the way in which individuals can *identify with their roles*. But the relation between individuals and their institutional roles turns out to be more intricate in nature and does not exhaust itself in such forms of identification. Thus, the process of R_{ID} aims to reveal another aspect of the relation between individuals and their institutional roles, which is not captured by R_{IN}. This aspect rests on an individual's capacity to critically *reflect* and *evaluate* her expected role-performances. When describing how individuals perform the tasks and functions of their institutional roles, it is important to keep in mind those ways in which individuals "step back", and critically reflect on the institutional roles they occupy. Individuals - including their capacity for action - are not being completely "absorbed" by their roles, and they maintain (varying) degrees of personal autonomy in relation to them.

To this end, the concept of *Role-Distance* has been invoked as an initial attempt to illuminate this relation.[142] For example, Hans Bernhard Schmid recently argued that individuals may keep a *distance* to their institutional roles, and that this role-distance constitutes a functional aspect of *acting qua role-occupancy:*

> "As sociological role theory has emphasized, role play involves not only role identification (which seems questionable given the differences between self-identity and role identity), but role distance, too. Competent role players know the difference between themselves and their role. They do not over-identify with their role, and they certainly do not ‚play at being their role' in the way criticized as ‚inauthenticity' (or ‚bad faith,' a synonymous term) by existentialist philosophers. *Role distance is required for at least two reasons. First, one needs to know when one should play by the role script (the normative expectations that constitute one's role) and when to break or deviate from them.* A good sales-person knows when to stick to the sales pitch and when to talk to the customer in a more ‚personal' way—and when to drop the act entirely and just be a person. Second, role distance is also required when it comes to balancing (and navigating between) different roles (for it is rare for anyone to have just one role), especially in the case of role conflicts" (Schmid 2023, 235) [own emphasis].

I agree with Schmid for the most part here. It follows directly from my description of *acting qua role-occupant* (see: Ch. 3.2.) that the difference between acting for individual, or private reasons and acting for role-based reasons for action can encompass such forms of conflict, and that the concept of *Role-Distance* can explain how such conflict may arise. Recall Gilbert's example of someone turning down a potential candidate, speaking "conscious of his role as a representative of the department" (Gilbert 1987, 196) while *personally* thinking that she is the best candidate available. This sort of conflict between one's private and one's role-based attitudes can encompass a great variety of affective and evaluative judgments, where such contrast may be evaluated negatively, or positively by the individual. In a similar vein Schmitz (2023), within the framework of his role-mode, argued that role-occupants

[142] Sociologists use the term "Role-Distance" mainly to describe observable behavior of individuals within their roles. However, the sociological concept of "Role-Distance" is broad and includes more types of social roles than just institutional roles, e.g., social roles like "mother", or "husband", etc. Maybe the most prominent sociological account of Role-Distance is provided by Goffman (1959; 1961), who defines Role-Distance not directly as a psychological relation holding between individuals and their roles, but rather as *actions* aimed to "effectively *convey* some disdainful detachment of the performer from a role he is performing" (Goffman 1961, 98). Although Goffman concentrates on the observable behavior of individuals, his concept of role-distance still aims to capture the *psychological relation* between an individual and her role. Role-distance, according to Goffman, is defined as the "'effectively' expressed pointed *separateness* between the individual and his putative role" (Goffman 1961, 95). Role-Distance can be contrasted with so called Role-*Embracement*. For an individual to "embrace" her role is to: "disappear completely into the virtual self available in the situation, to be fully seen in terms of the image, and to confirm expressively one's acceptance of it" (Goffman 1961, 106). In contrast, the term role-*distance* refers to only those behaviors which are "relevant to assessing the actor's [i.e., the role-occupants] attachment to his particular role and relevant in such a way as to suggest that the [role-occupant] possibly has some measure of disaffection from, and resistance against, the role" (Goffman 1961, 108).

> "may just learn to live with the fact that [they] represent values and rules that are not always the ones [they] personally favor. [They] may even strongly identify with doing this because [they] believe it to be essential to the functioning of a diverse and liberal society. But [they] could also decide that this conflict is unbearable and that the only right thing for [them] to do is to resign [their] role, or to try to start a revolution" (Schmitz 2023, 197).

The sort of distance individuals may hold towards their roles, as well as to the rights, duties and obligations, is *one* way in which individuals may come to have a reflective, and evaluative attitude towards their roles. But notice that the focus of such an analysis lies on the relation between individuals *as private persons* and individuals *as role-occupants*. To sloganize this sort of role-distance, Schmid's above mentioned sales-person could ask herself: "Is it just *me* who wishes to act like this? Or is this really how *a sales-person* would act?". Now my analysis of R_{ID} has a slightly different focus. What I want to capture with the concept of Role-Idealization is something other than individuals knowing the difference between *themselves as private persons* and their role; or knowing about the difference between their private reasons for action and their role-based reasons for action.

Rather, I want to examine an individual's capacity to critically reflect and evaluate their *own*, *non-ideal* role-performances in light of a *prototypical,* or *ideal* role-performance. To stay within the example of Schmid, I do not wish to analyze the ways in which a sales-person experiences herself *qua sales-person* as different from herself *qua private person;* how an individual may feel alienated, or estranged from the role she occupies; or how she may reduce the psychological dissonance by becoming ironically detached from the role; how she may become cynical about her role because of her private resentments against it, etc.

Rather, I want to analyze how such a sales-person may be able to realize that some aspects of her performance are deficient, inadequate or *non-ideal* when being compared to how a *prototypical*, or *ideal sales-person* would perform the same tasks. This, I think, provides a basis on which such a sales-person is able to subtract, or suspend, the personal, idiosyncratic, and contingent aspects of her role-performance. To sloganize *this* sort of role-distance, a sales-person might ask herself: "What aspects of my performance *as a sales-person* are based on *my interpretation* of this role? And is my *interpretation* really congruent (or compatible) with how to *ideally* act as a sales-person?".

Notice, that these two mentioned forms of distance from one's role can come apart: knowing the difference between, e.g., oneself as a private individual and one's role when acting *as a role-occupant* (e.g., regarding one's private and role-related reasons for action) does not necessarily entail that one knows the difference between what aspects of one's role-performance are based on *one's idiosyncratic interpretation* of this role. In turn, I hold this latter capacity of orienting oneself towards an idealized, prototypical role-performance to make intelligible how individuals prevent their role-performances from *turning toxic*.

Let me further explain the difference by recalling the case of a policewoman (Ch. 4.2), who has been assigned the task to stop and frisk "suspicious" individuals, where the policewoman determines the standard of who's "suspicious" (or not) according to her personal racial biases. I argued that her (mis-)use of discretionary powers leads to her role-performance turning *toxic*, i.e., that her actions fail to actually contribute towards the institutional groups' overall goal (of both maintaining social peace in the community and upholding the law, etc.). What I want to capture with the concept of *Role-Idealization* is not the way in

which the police-woman may keep a distance towards her institutional role, e.g., by asking herself: "Is it just *me* who wishes to stop and frisk this individual? Or is this what *a policewoman* would do?" Intuitively, the police-woman could, on *this* basis, not realize how her discretionary interpretation of her actions *as a police-woman* turn toxic. This sort of distance towards her role would simply allow her to understand that stopping and frisking individuals is part of her performance *as a policewoman*, and not something she would herself, as a private individual, do.

Rather, the sort of critical distance I have in mind would consist of the police-woman asking herself: "What aspects of my performance *as a policewoman* are based on *my interpretation* of this role? And is my interpretation really congruent with how to *ideally* act as a police-woman?" Regarding her motives for determining the standard of "suspiciousness", she might ask herself "is stopping and frisking only individuals of one particular ethnic group *really* congruent with how to *ideally* perform this task as a police-woman?" I hold *this* second, interpretative level of critical reflection and evaluation of her role-performance to be able to provide the basis which enables the police-woman to subtract or suspend the idiosyncratic, and contingent aspects of her role-performance, i.e., her personal racial biases.

So again, critical reflection of one's role can happen on two different levels: First, one can, as a *private* (i.e., non-role-occupying) individual critically reflect one's role. And second, one can *as a role-occupant* critically reflect one's own role-performance in light of an *ideal* or *prototype*. Role-Idealization is concerned with the latter level of reflection.

Role-idealization as a heuristic strategy

At the most fundamental level, R_{ID} is the process through which an individual is provided with a *regulative ideal* or *ideal standard* against which her *non-ideal* performance can be judged. I hold R_{ID} to be best understood as a two-step mechanism, where the first step is to *determine* an *idealized*, or *prototypical* role-performance, and then, the *second* step is the attempt to imitate this imagined, idealized behavior. When engaging in the process of R_{ID}, a role-occupant might ask herself how to *ideally* act out her role, thereby employing what we could call an *imitate-the-ideal* heuristic.

The idea of conceptualizing this process as a *heuristic* is borrowed from Hertwig & Hoffrage (but also see: Gigerenzer et al. 1999). At the most basic level, a heuristic is just a strategy by which individuals come to make decisions. According to the authors, a heuristic strategy consists in simplifying informational input, "with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods" (Hertwig & Hoffrage 2012, 6). Interestingly, the authors claim that such heuristics are able to provide individuals with the means necessary to operate within the complexity of the social world (and thus within social roles) *without* demanding an equivalent of cognitive complexity from the individual employing such heuristics (cf. Hertwig & Hoffrage 2012, 15-22). Primarily, heuristics are a simple (i.e., cognitively cheap) way for individuals to navigate the complexity of the social world around them. The authors conclude here, that

> "contrary to a suspicion still harbored by many social and cognitive psychologists, simplicity in cognitive mechanisms does not open the floodgates to irrationality […] or to other horrors named […] and unnamed. Nor do heuristics capitulate in the face of complexity, uncertainty,

> scarcity of information, or time pressure. They are the indispensable tools that the mind - that parts dealer and crafty backwoods mechanic - can recruit to find solutions to intractable problems in a complex and uncertain world" (Hertwig and Hoffrage 2012, 26).

One particularly interesting heuristic that the authors pick out is the so called "Imitate-the-successful heuristic" which is summarized this way:

> **IMITATE-THE-SUCCESSFUL HEURISTIC:** Determine the most successful agent and imitate his or her behavior (e.g., action, judgment, choice, decision, preference, or opinion) (Hertwig & Hoffrage 2012, 8).

It's important to note, that the authors do not necessarily hold such a heuristic to require *actually* observing the model's behavior. Instead such a heuristic can encompass *counterfactual* forms of reasoning, where "it may be sufficient merely to ask oneself *what the model would have* done" (ibid) [own emphasis]. When direct behavioral observation is not possible, "individuals as well as institutions can keep records of the behavior of efficacious predecessors, allowing others to benefit from their wisdom and success—and their failures" (ibid.). Let me now adapt and modify this "Imitate-the-successful heuristic", and on its basis derive my definition of the "IMITATE-THE-IDEAL HEURISTIC", which I hold to convey the fundamental process by which R_{ID} comes about. It runs the following way:

> **IMITATE-THE-IDEAL HEURISTIC**: Determine an idealized (or prototypical) role-occupant and imitate his or her behavior (e.g., action, judgment, choice, decision, preference, or opinion).

When employing this heuristic, a role-occupant, in a first step, comes to identify the idealized forms of behavior (action, judgement etc.) associated with the prototypical role-occupant. The next step would see her trying to imitate such behavior.[143] But crucially, notice that this process also implies that a role-occupant comes to identify the *discrepancies* between the idealized standard and her own, non-ideal performance (the *second*-level critical reflection mentioned above). On this basis, she may *regulate* her own, non-ideal behavior in accordance with the so derived standard (or narrative) of the prototype.

The reader might've become nervous, because ideally, I would already have answered a crucial question: How is the *content* of such an ideal, or prototypical role-performance determined by the individual who engages in such Role-Idealization? What agent and whose behavior is determined (and consequently imitated) by someone employing the heuristic of Role-Idealization? Who is this *ideal* role-occupant and what's so ideal about her? Does she live in the platonic heavens, amongst the ideal forms?

Unfortunately, I will provide an answer to this only after discussing a proposed answer to this question provided by Kirk Ludwig (2017a; 2017b). Pace Ludwig, I will argue that mere individual introspection alone doesn't do the job, and that an idealized role-perspective must in some way be fixed an *external* standard. I will argue in the next section, that one can identify processes of interpersonal or collective management

[143] You may think of it as a *snowclone* of the Christian *"What would Jesus do?"* but for institutional roles (What would <ideal occupant of role x> do).

and monitoring of prototypical role-perspectives within institutional groups. I hold these processes to provide an adequate basis for such an external standard. But let us first turn to Ludwig.

Ludwig's "margin of error" and "idealized conception of role performance"

In his theory of institutional agency (Ch. 3.1) Kirk Ludwig, too, uses the concept of *idealized role performances* to explain how individuals are able to cooperate in non-ideal circumstances by relying on an *prototypical or canonical plan description*. On Ludwig's account of collective action, the individuals involved in collective action have to have a shared plan concept, which encompasses individual actions to contribute to said shared plan. As explained above, the individual intention to contribute to such a shared plan only needs to specify a *prototypical or canonical plan description*. So on Ludwig's account, there is a certain leeway that individuals have in interpreting their individual contributory performances, because plans come with what he calls a *margin for error* (cf. Ludwig 2017a, 214f). This is the primary way in which Ludwig's theory could be said to be able to respond to the *second Problem of Discretion*. Pointing to such a *canonical plan conception* explains how individuals come to have interpretative guidance in the selection and application of discretionary standards. In turn, these standards inform the individual how to perform the functions of one's role in a way that is anticipated to further the institutional group's goals. What Ludwig here appeals to is the intuition, that a plan does not have to be "perfectly" executed, and not everything has to go *exactly* according to plan, for individuals to perform a joint action:

> "We noted in the case of individual action that plans come with what can be called a *margin for error*. When I am trying to kill someone, I may aim at his head but shoot him in the torso. I did not intentionally shoot him in the torso, but I did intentionally shoot him, and I did intentionally kill him, even if not every detail of the plan to kill him by shooting occurred in accordance with the way I envisioned it. In saying that the plan to kill my victim had a margin for error, I have in mind that the way I specify the plan if asked specifies a *kind of prototype or ideal conception of what occurs*, while in fact I envision a range of variation on the prototype, ways that the end can be achieved that are close enough, for it to count as coming about in accordance with my general plan for doing it. We can put it by saying that the plan concept is a prototype concept, and anything close enough to the prototype counts as falling under the plan concept in question" (Ludwig 2017a, 214) [own emphasis].

Two things should be noted here. First a shared plan can be executed under a certain variation of output on behalf of the individuals. Plan's having such a *margin of error* allows them to be more or less ideally executed. Such room for variation, as we saw above, might ultimately be necessary to stop institutional groups from exhibiting institutional stupor. Crucially, Ludwig's explanation of how to determine the *margin of error* involves a "vague penumbra" (ibid) of whether an action falls under the canonical plan description or not. Ludwig gives another example of Pallbearers, who slightly deviate from the pre-planned way the ought to carry a casket to a grave. Such pallbearers

> "may operate with slightly different conceptions of their ordering and, without noticing, carry the casket to the grave anyway, and in doing so they surely do so together intentionally. But this is not a problem. Various ordering are within the margin of error of the specified plan—*what shows the appropriate margin of error is what they see as deviating too far from the idealized conception* of how it was to go if it is to count as close enough. *There is no independent standard to which to appeal*" (Ludwig 2017b, 28) [own emphasis].

Second, Ludwig states that an idealized, or prototypical role performance can serve as a *regulative ideal or standard* against which a non-ideal performance can be judged, in order to determine whether it falls in- or outside of the *vague penumbra* (cf. Ludwig 2017a, 215). Accordingly, a role-occupant might ask herself how to ideally act out her role, thereby employing what I've called the *imitate-the-ideal* heuristic, where she first determines an ideal role-occupant's plan description (the "idealized conception of how it was to go") and then tries to imitate this imagined behavior (e.g., action, judgment, choice, decision, preference, or opinion), i.e., she might try to regulate her own behavior in accordance with the so determined ideal.

Now in Ludwig's case of the pallbearers, a "that'll do"- or "that should do the trick"-mentality, so to speak, seems to suffice. For the pallbearers all seem to have sufficiently overlapping understandings of the *ideal plan* to carry the casket to its grave; and they also seem to have a way of determining whether variations in their performances fall in- or outside of a *margin of error* concerning this *ideal plan*. Also note that in the case of the pallbearers, the shared plan to collectively ϕ (to carry the basket) is already developed, or not subject to an *ongoing* development. So far so good. But once we connect this explanation of Ludwig with the above mentioned *Problems of Discretion*, things start to get shaky. This becomes apparent not only when we consider that actions based on discretionary powers can be vastly more complex than the range and scope of Ludwig's example of carrying a casket suggests, but also when we take a closer look at the way in which such a "margin of error" could be determined by the individual in cases, where the individual has to "make up her own mind", i.e., where use of discretionary powers is necessary.

The crucial problem with Ludwig's theory is this: A "that'll do"- or "that should do the trick"-mentality, so to speak, does by no means guarantee that individuals succeed in performing the tasks and functions of their roles, if (i) they themselves have to settle the standards of *what* will "do the trick" and if (ii) they base their standards of *what* will "do the trick" on false, or misguided interpretations. But if there *is no independent standard to which to appeal to*, the individuals themselves may be unable to determine whether (ii) is the case, i.e., whether their interpretation is false or misguided in the first place. Simply pointing to an "idealized" role-performance or an *imitate-the-ideal* heuristic does not - in and of itself - shield these standards from being based on a false, or misguided interpretation. However, such interpretation might ultimately lead to an individual's performance to fall outside of the "vague penumbra" of the canonical plan description.

To see this, let us assume, with Ludwig, that what *I* see as deviating too far from the idealized conception is what actually determines the appropriate margin of error. Let us also assume, again with Ludwig, that at the same time, there is no *independent* standard to which I could appeal for this ideal conception. But if this is both correct, then my interpretation of an idealized plan-conception, including what I believe is deviating too far from the ideal conception, becomes void of regulative guidance. To see why, notice that an *idealized*

performance must -*pace Ludwig*- in some way be fixed, or subject to an *external* standard if it is to be a *regulative* ideal.

For one, because if it was entirely up to the individual to determine such an ideal conception, the result may *vary* consistently from individual to individual. Two individuals occupying the same institutional role would not necessarily agree on how to ideally perform their tasks, if they have two different, and potentially incompatible conceptions of what such an ideal performance would amount to. Now if that was the case, the guidance for interpretation and action that the ideal should provide would be redundant.

Second, if an idealized role performance was *not* fixed by an external (or non-independent) standard, it seems to run the risk of becoming a *self-fulfilling prophecy*, in which case it also seems to lose its regulative guidance. If *what counts as an idealized performance* is based on *whatever I take such an idealized performance to be*, then every interpretation of my actual performance, which is measured against this ideal, could also be said to fall under my interpretation. The difference between one's non-ideal performance and the ideal against which one interprets this performance would, so to speak, collapse into itself. So in a situation where an individual is in need of interpretive guidance on how to perform (all or certain) functions and tasks of her role (because she cannot foresee whether certain courses of action fall within the *vague penumbra* of a canonical plan description) she might ask herself "what would an *ideal role-performer* do?". But said individual, as a result of her, and *her alone* determining what an ideal role-performer would do, then might be faced with merely self-affirming (and therefore non-guiding) answers on how to perform the functions and tasks of her role.

This seems to be a contradictory result. Think, e.g., of a particularly unmotivated individual who wants to do only the bare minimum of what is required from her as a role-occupant. Let's imagine that such an individual comes to face a situation where she cannot foresee whether certain courses of action fall within the *vague penumbra* of the canonical plan description. Now this individual might ask herself "what would an *ideal role-performer* do?" in order to determine whether a particular course of action of her contributes to her groups plan. But if it is entirely up to this individual to determine such an idealized role-performance, she might (maybe out of malignancy or laziness) conclude that the ideal role-performance consists out of doing only the bare minimum of what is required from her as a role-occupant. Clearly, it sounds strange to say that this individual is thus trying to imitate an *ideal* role-occupant *just because she decided that she is actually doing so*. Rather, I think it's intuitive to say that her interpretation of an ideal role-occupant misses the mark, i.e., that it is false or misguided.

Alternatively, the lack of an independent standard may lead to an interpretive regress, where an individual, lacking interpretive guidance, tries to answer the question of what *an ideal role-occupant would to* by asking herself how *an ideal role-occupant would answer the question of what an ideal role-occupant would do.* This individual would then also need a second, or *meta*-ideal of a role-occupant against which her adaption of the first-level ideal of a role-occupant could be measured, where the second-level ideal would need a third-level ideal, etc. Either way, Ludwig's explanation for how a canonical or prototypical role performance can be the standard against which individuals measure their own performance seems to be off. The underlying problem here is that Ludwig's theory does not encompass an independent, external standard to which to appeal to (cf. Ludwig 2017b, 28). Yet such a standard turns out to be necessary.

But what could provide such an external, i.e. independent standard? In the following sections, I will try to identify processes of interpersonal or collective management and monitoring of prototypical role-perspectives that provide individuals with such an action guiding, externally-fixed standard. I will here try to show, that while we can describe the *orientation* of one's performance in light of an idealized role-perspective as individual process; the *determination o*f such an idealized role-perspective must be thought of as a collective, or social process.

The creation, management, and monitoring of R_{ID}

Introspection about idealized role-performances alone will not suffice as an explanation for how R_{ID} can guide the actions of individuals, especially in situations where such idealized role-performances constitute the standard on which individuals base their prospective actions on. But what will? In this section, I will explicate some processes which I hold to provide such a standard. This sections aims to answer the questions as to where, in her outer world, an individual should look out for idealized role-performances. In order to systematically analyze the social and collective creation, management and monitoring of R_{ID}, I will distinguish three levels on which such R_{ID} can occur and identify certain processes, which correspond to each of these levels.

A first place to look for idealizations to reside in processes occurring on an interpersonal level. I will argue that the process of *role-modeling* corresponds to this interpersonal level. A second place is that of an individual's broader environment, within her social and cultural surroundings. Corresponding to this level is the process of *self-stereotyping*. Third, institutional groups themselves seem to exhibit internal mechanisms that aim to provide a conception of an idealized role performance. Corresponding to this level, are what I will call *processes of institutional ideal formation*. These group-level *processes of institutional ideal formation* are particularly interesting, because they can be understood as an attempt of particularly large and complexly structured institutional groups to mitigate the potential heterogeneity of their individual role-occupants' attitudes. This heterogeneity, I argue, poses a threat to institutional agency, and *processes of institutional ideal formation* can thus be described as an attempt to mitigate these threats.[144]

Interpersonal-level R_{ID}

So let us begin with the interpersonal level. Here, Role-Idealization might come about in a straight-forward way by an individual actively trying to imitate, or *model* an ideal role-occupant through the use of already existing role-occupants within her institutional group.

One way to grasp the interpersonal level of R_{ID} is by describing it as a *novice-expert-relation* that holds between an individual (the novice) who tries to identify and imitate the behavior of a role-occupant who is already part of the group (the expert). The novice-expert-relation then consists of an individual observing and imitating the behavior of a pre-existing group-member. Such observational learning may include processes like *job-shadowing*, or *"on-the-job" training processes*, where novices are able to actively observe

[144] I do not claim that these processes constitute an exhaustive list of how idealized role-performances can be collectively managed and maintained.

an experienced role-occupant in the performance of her tasks and functions. But it may also work the other way around, in the form of *mentorship programs* or *performance reviews* by experts. Here, an expert may provide an assessment of the ways in which a novice performs a task and function of her role, where the novice is supposed to incorporate the feedback into her role-performance. But how could this type of mimicry be said to provide a conception of an *idealized* role-performance? Why does observational learning enable an individual to imagine what an *ideal role-occupant* would do, instead of just enabling the individual to imitate the observed behavior of others (and maybe repeat the non-ideal behavior of the alleged expert)? To explain how interpersonal learning process can actually provide a novice with role-*idealizations*, note that the novice-expert-relation is not necessarily restricted to a one-to-one relation. Rather, it should be modeled as a *one-to-many relation*. By this, I mean that a novice, observing and on this basis comparing *several* expert-models, can assemble an *ideal* role-performance by means of comparison and abstraction: One feature (f_1) of role model M_1 might be modeled by the novice, but not f_2 and f_3. In tun, the features of an idealized role model may consist of an *assemblage of several, different experts* M_1-M_n, whose (partial) features are evaluated and combined (see Figure 3 below).

Conceptualizing the interpersonal R_{ID} as a one-to-many relation seems to be a more empirically accurate description of the ways in which individuals use *role models.*[145] Filstad examines the ways in which novices in institutional groups use *role models* in what she calls *organizational socialization*.[146] Her research suggests that instead of imitating the behavior of just one individual within her group, novices tend to *construct* an idealized role-performance by observing, comparing, and evaluating the actual performances of *several experts*. So instead of having *total role models*, individuals actually tend to actively utilize "*multiple contingent role models*". Filstad's study, which was conducted by interviewing newly appointed employees in a real estate agency, set out to explain how novices use established members as role models in organizational socialization. A first finding is that newcomers indeed tend to actually rely on role models in order to understand both the institutional group's arrangements, as well as their own institutional roles within these arrangements.

But the initial result of her study seemed somewhat contradictory. On the one hand, the study showed that newcomers *do* actually use "supervisors, co-workers and even secretarial staff more or less actively through observation, interaction and communication" (Filstad 2004, 399) in order to internalize the organization's demands and expectations. On the other hand, when being asked about this, the newcomers tended to straightforwardly *deny* that they identified particular individuals as a role model, or that they actively tried to model their behavior:

---

[145] The term "Role Model" was originally coined by sociologist Robert K. Merton (1957) who used it to explain how medical students adopt and internalize their roles as medical practitioners. Since then, the term "Role Model" has become part of the common vernacular, but it also received great attention in academic discussions. Morgenroth et al. claim that the term is used over 400.000 times in scholarly papers (Morgenroth et al. 2015, 467). They also provide a useful summary on the ways in which social scientists have tried to capture the phenomenon. See for studies that use the term role model to describe the internalization of particular institutional roles in e.g., medical education (Paice, Heard, & Moss, 2002), nursing schools (Perry 2009), or in the academic or educational sector (Almquist & Angrist 1971). For a literature review of the impact of role models on entrepreneurial behavior see: Abbasianchavari & Moritz 2021.

[146] See Chao (2008) for further discussion of the theory of *multiple contingent role models*.

> "When it comes to ‚identification', ‚idols' and ‚role models', the newcomers are quite unconscious of these terms. They admit that the terms are difficult to recognize in their organizational socialization, and furthermore many of them claim not to use role models. On the contrary, they claim not to pay much attention to what other established real estate agents do, but want to create their own style. Accordingly, they explain that they do not identify with anyone in particular" (ibid).

Filstad concluded that "there is quite an inconsistency between their explanations and their behavior. Most of them are quite convinced that they do not have role models, or co-workers that they identify with. But observations suggest that they do" (ibid). Now in oder to explain this sort of dissonance, Filstad concluded that organizational socialization happens not in terms of "*total* role models", i.e., one individual being recognized by the novice as incorporating *every* characteristic feature of an ideal role-occupant. Rather, individuals tend to observe and imitate *multiple* individuals; and subtract their individual idiosyncratic behavior. Novices, Filstad concludes,

> "do not search for or recognize total role models. So when it comes to how newcomers use role models in organizational socialization, the answer is that they use several role models. They select different qualifications from several role models in order to create their own personal style and role behavior" (Filstad 2004, 400).

So observational learning on the interpersonal-level does actually seem to provide the basis for *idealized role-performances*. It should, however, not be thought of as a one-to-one relation of observing and imitating the behavior of just *one* pre-existing group-member. The more promising way in which individuals identify and utilize role models is to think of it as a more complex, and creative process of assembling an *ideal type* of role model by observing, comparing, and evaluating the actual performances of *several* experts.

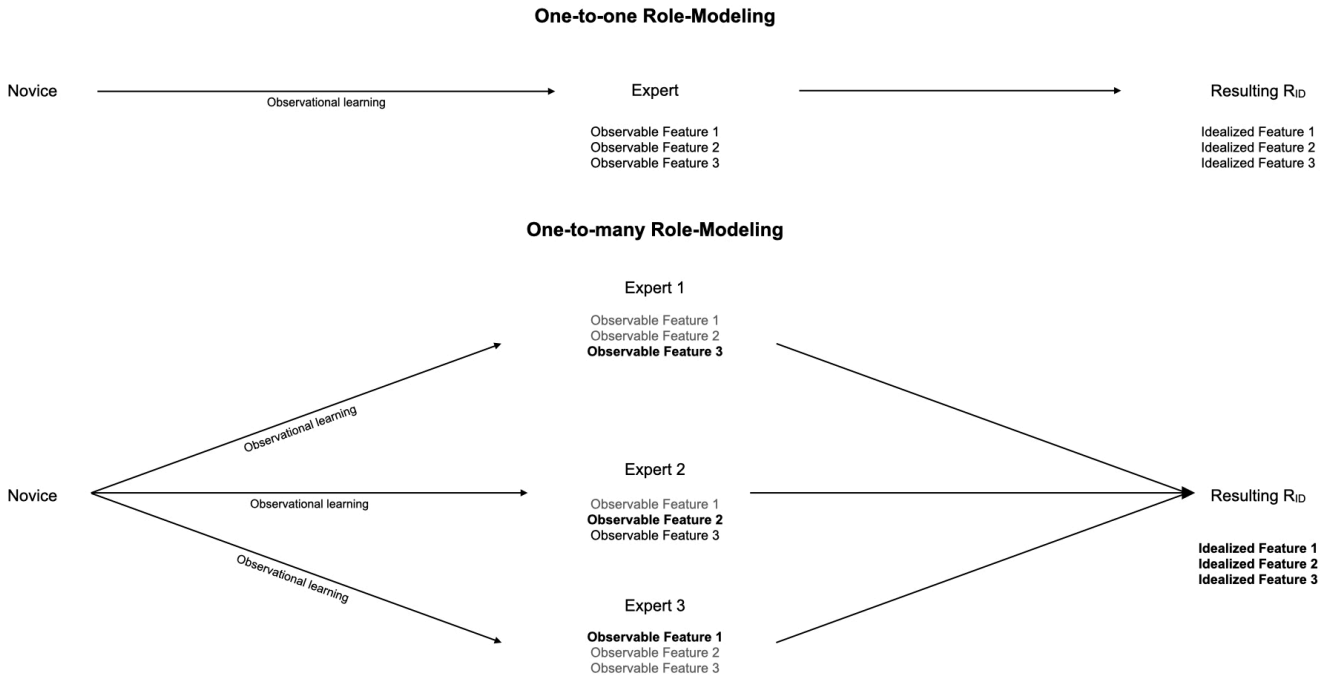

Figure 3: Two processes of role-modeling

Environmental-level R_{ID}

Let's turn to the second level of R_{ID} and see, how Role-Idealizations may be fixed to an external standard there. My conjecture is that an individual, in trying to find an idealized standard against which she can measure her non-ideal performance, can be influenced by cues in her broader environment, especially by existing stereotypes about institutional roles.[147] I hold the concept of stereotypes to offer a direct explanation for how the individual-level process of o*rienting* oneself in light of an idealized role-perspective is connected to collective, or social-level process of *determining* such an idealized role-perspective. Whereas social stereotypes can be conceptualized as *social*, or *cultural entities,* they give way to the individual-level process of *stereotyping* (cf. Blum 2004, 252). Stereotypes facilitate the adaptation of an idealized role-perspective, because they *shape* the minds of those who engage with them. Let me elaborate.

For brevities sake, I will use the term *stereotype* to refer to "a collection of beliefs about the traits characterizing typical group members" (Lavelle 2022, 29f. see also: Dovidio et al. 2010, 7-9).[148] Let me

[147] By stating that the cues can be found within her *broader environment*, I primarily mean that they are not confined to the context of the institutional group that the individual is part of, i.e., that they can exist *outside* of the context of the particular institutional group.

[148] Alternatively, Beeghly defines them as "universal generalizations about a social group" (Beeghly 2015, 676); and Spaulding defines them as "conceptually rich systems of belief about social groups" (Spaulding 2018, 26). These definitions are, however, too broad for our purposes, as they apply to beliefs about the traits characterizing members of all sorts of social groups, and do not differentiate between *organized* and *structured* groups.

narrow the definition down by taking the term "*institutional group stereotypes"* to refer to a *collection of beliefs about the traits characterizing typical members of institutional groups.*[149] From here, we can narrow it down even further by defining *role-stereotypes* as *a collection of beliefs about the traits characterizing typical occupants of institutional roles.*[150]

My claim that stereotypes are part of the individual's broader environment (and in this sense "social" or "cultural") rests on the assumption that the transmission of stereotypes does not happen only through interpersonal means within the context of the institutional group. Rather, stereotypes can reside in what Fricker (2007) calls the *shared hermeneutical resources* of populations, which are "the shared meanings […] members [of a given population] use to understand their experience, and communicate this understanding to others" (Romdenh-Romluc 2017, 1). These shared meanings can be stored in, and transmitted via a variety of means, including social artifacts, e.g., books, movies, music, visual arts, advertisements, print media, stories, fables, propaganda, etc. Now, of course, all sorts of stereotypes (e.g., about ethnic groups, or gender-roles) may be part of the collective hermeneutical resources, and be transmitted via such artifacts (see: Balkaran 1999 for the way in which mass media harbors racist stereotypes against African-Americans; for a general analysis see: Ross 2019). But the stereotypes transmitted in the hermeneutical resources may also encompass *institutional role stereotypes*.

Consider, just as one example, the depiction of professional and institutional stereotypes in television shows. Here, McLeod et al. conducted a study about the portrayal of individuals working in *IT* and examined whether they were depicted in stereotypical ways. The authors examined five television shows and tried to find out whether IT experts were portrayed according to common stereotypes, including categories like *Appearance* in the form of a "poor dress sense", *Personality* in the form of "Nerdiness or geekiness" and *Employment Role*, where "the stereotypical IT job is one requiring someone stuck in a dark office all day in front of a computer with paraphernalia such as Sci Fi posters or junk food littered around the room" (McLeod et al. 2013, 4). The study found that while many existing stereotypes were challenged by such TV shows, "all IT expert characters displayed some stereotypical characteristics, with IT technicians portrayed in the most stereotypical way" (McLeod et al. 2013, 1).[151]

Now one study of five television-shows alone, of course, amounts to a rather anecdotal evidence about the way role-stereotypes may reside in the collective hermeneutical resources of a given population.[152] But first,

---

[149] Now this definition can be applied both to *particular* institutional groups and institutional group-types: An individual might think that people who work for, e.g., *Boston Consulting (the token-group)* are sordid and sleazy; and they may derive this stereotype from some institutional group stereotype they have regarding the group-*type*, i.e., *consulting firms*.

[150] This can be understood in two ways as well. First, that particular role-*types*, independent from the institutional group that they belong to, exhibit such traits, e.g., that *accountants in general* are boring and dry individuals, that *firefighters* are brave and risk-seeking individuals, or that *people working in IT* are socially-inapt basement-dwellers. Second, role-stereotypes can encompass beliefs about the traits characterizing individuals occupying institutional roles in *particular* institutional groups, e.g., when people hold stereotypical beliefs about *priests of the Catholic Church* to be conservative, backward-looking and frumpy etc.

[151] See García-Crespo et al. (2008) for similar results; or see Lieberthal (1976) for an analysis of TV and movie images reinforcing stereotypes about workers as being vulgar, boorish and ignorant.

[152] But let me remind the reader that there exist entire *genres* of television programs, i.e., *Workplace TV Shows* dedicated to certain professions, e.g., lawyers, doctors, politicians, policemen and -women, the military, corporate jobs etc. Let me also remind the reader that the average American, as of 2022, watches about three hours of television *per day* (Stoll 2023).

I hold that the same kind of argument could be made with other means of transmission, e.g., stereotypes in literature, cinema, or advertisements. Second, and more importantly, all I want to claim here is that 1) stereotypes about institutional roles exist and that 2) an individual might be confronted with stereotypical depictions of certain institutional roles in her broader environment.

But what follows from this? To see what makes the concept of stereotypes particularly attractive for my endeavor, let me, in a next step, illuminate the ways in which stereotypes play a function in social *cognition*. Notice, that the way in which stereotypes have been described up until now is being *other-directed*, i.e., they are beliefs about the characteristics of *other* individuals. Naturally, one particularly interesting way of looking at stereotypes has been advanced in the field of *social cognition* and within the *Theory of Mind*. Here, *pluralistic* accounts of folk psychology argue for stereotypes to be one of several possible *strategies for mental state inference* (Spaulding 2018; Ames 2004; Fiske & Taylor 2013, Ch.11; Westra 2019). Spaulding, e.g., states that stereotypes, as a strategy for mental state interference, rely on the process of *social categorization.* This, in turn, describes a process of "sorting people, behaviors, and events into social categories" (Spaulding 2018, 25), which "is essential for successful navigation of the social world" (ibid) and "helps make the social world more comprehensible and predictable, and thereby allows us to manipulate the social world for our purposes" (ibid). So the process of *stereotyping* is a strategy that individuals deploy when trying to interpret, understand and predict the behavior of *others* by means of categorizing them according to a stereotype.[153] As such, stereotyping is closely connected to the above established concept of social heuristic, as it, too, is constitutes a cognitively cheap and quick way of processing information about an individual's social environment.[154]

This view of stereotyping as a mindreading-tool for interpreting and predicting behavior of *others*, however, is only one half of the story. Interestingly, individuals are also able to apply such stereotypes to *themselves*. Hence, scholars have identified the process of *self-stereotyping* (or "self-directed stereotyping") as "the perception of the self as a prototypical group member" (Spears et al. 1997, 538). If we adapt this definition of *self-directed* stereotyping (or self-stereotyping) and plug it into the definition of institutional role-stereotypes, then *self-directed role-stereotyping* can be defined as *the perception of the self as a*

[153] Ames (2004) argues that stereotyping is but one of the available strategies in "the mindreader's toolkit" and that people rely on different strategies at different times. According to Ames, stereotyping is the *primary* way through which we make sense of others when we perceive them to be *unlike* ourselves. On the other hand, Ames discussed studies that suggest that the process of *projection* is the primary way through which we make sense of others when we perceive them to be *like* ourselves.

[154] Fiske and Taylor (2013) develop their theory of stereotypes under the psychological paradigm that individuals are "cognitive misers", i.e., that "people are limited in their capacity to process information, so they take shortcuts whenever they can. People adapt strategies that simplify complex problems; the strategies may not be correct of produce correct answers, but they emphasize efficiency" (Fiske & Taylor 2013, 15). Stereotyping is one of these strategies by which individuals try to make sense of their social environment in such quick and simplified ways.

prototypical role-occupant.[155] I hold this *self-directed role-stereotyping* to be the fundamental process in play when individuals engage in the above described form of R_{ID}.

Let us try to see the forest, instead of just the trees: In trying to find an idealized standard against which an individual role-occupant can measure her non-ideal performance, said individual might be influenced by stereotypical depictions of role-performances in her broader environment. On this basis, she might engage in *self-directed role-stereotyping* and start to perceive of herself as a prototypical role-occupant. And if she, on the basis of such self-stereotyping, comes to think of herself as a stereotypical role-occupant, then she might adapt certain attitudes, traits or patterns of behavior, which she associates with occupying such an institutional role. So if an individual tries to find an idealized standard against which she can measure her non-ideal performance of, e.g., an *accountant*, said individual might be influenced by stereotypical depictions of accountants in her broader environment. On this basis, she's enabled to engage in *self-directed role-stereotyping* and to perceive of herself as a prototypical accountant. And if she, on the basis of such self-stereotyping, comes to think of herself as a stereotypical accountant, then she might adapt certain attitudes, traits or patterns of behavior, which she associates with occupying such an institutional role. She might think, e.g., that she ought to behave in a more rigid, prudent and conscientious manner when performing her institutional role *viz*. the corresponding tasks and functions, because her stereotypical understanding of accountants, being informed by cues in her environment, sees them as boring, solemn and prudent individuals. Likewise, a fire-fighter, due to the process of self-stereotyping, might be more likely to run into a burning building, if she thinks that fire-fighters in *general* are ought to behave as brave, risk-seeking and heroic individuals (understood here as traits of the role-stereotype).

Another cogent way to theoretically frame this process is provided by Tadeusz Zawidzki's theory of *mindshaping* (2013; 2016). According to Zawidzki, social cognition should not solely be understood as the capacity for *mindreading*, i.e. the capacity to interpret and understand other individuals' mental states.[156] Instead, Zawidzki aims to highlight the *regulative dimension* of our socio-cognitive capacities: "our complex, diverse, and flexible capacities to *shape* each other's minds in ways that make them easier to interpret" (2013, xi) [own emphasis]. According to Zawidzki, this "mindshaping" - in the form of capacities and practices such as sophisticated imitation, pedagogy, conformity to norms, narrative self-constitution - is the most important component of human social cognition (ibid). And again, mindshaping is not merely an other-related capacity, but can be self-directed too. Zawidzki calls this sort of mindshaping process *epistemic self-regulation*. And interestingly, such epistemic self-regulation can, according to Zawidzki, occur on the basis of fictional characters and *idealized agents.* He explains the process the following way:

---

[155] Similarly, Brewer and Gardner define the process of self-stereotyping as a "shift in the perception of oneself as an interchangeable exemplar of some social category; to think of oneself as having characteristics representative of a social category" (Brewer & Gardner 1996, 85ff.). If we adapt and modify Brewer and Gardner's definition, *self-directed* role-stereotyping denotes *a shift in the perception of oneself as an interchangeable exemplar of an institutional role-occupant; to think of oneself as having characteristics representative of that institutional role*.

[156] See the special Issue *Folk Psychology: Pluralistic Approaches* (2021) by Andrews, Spaulding & Westra (Eds.) for an overview. In this special issue, see especially Lavelle (2021) for another description of the regulative dimension of social cognition including cultural stereotypes.

> "The idea here is that ritualistically encoded commitments constitute publicly expressed ideals that regulate the behavioral dispositions, i.e., shape the minds, of those who express them, in virtue of the group-level normative attitudes that institute such ideals. Once such communicative and coordinative practices are on the scene, they can trigger a proliferation of virtual social models for mindshaping. These social models are ‚virtual' because they are not necessarily embodied in the behavior of any actual individual; rather, they consist in *publicly and symbolically encoded patterns of behavior, like mythical narratives, that specify ways of playing social roles, e.g., being a parent, that are tacitly sanctioned by a community* […] *Thus, when individuals conceptualize themselves in terms of these roles, these self-conceptualizations play an important mindshaping role: their point is not to describe but to regulate mental states and behavioral dispositions* […] *By conceiving ourselves in terms of virtual social models, and publicly expressing such self-conceptualizations, we set up incentives to shape ourselves to approximate these social models*. These incentives derive in part from the normative attitudes of our group-mates, which support various forms of punishment that enforce adherence to courses of behavior deemed compatible with such public expressions. […] So, for example, *public declarations of self-conceptualizations and commitments to play social roles such as ‚parent', ‚mate', ‚doctor', ‚police officer', ‚president', ‚teacher', etc., alter our incentives in ways which encourage shaping ourselves to fit the expectations members of groups to which we belong hold regarding such categories*" (Zawidzki 2016, 483f.) [own emphasis].

Let me summarize this section. I claimed that R_{ID} is best captured as a social heuristic, where an individual determines an idealized role-occupant and, in a next step, tries to imitate his or her behavior. On this basis, she may *regulate* her own, non-ideal behavior in accordance with the so derived epistemic standard (or narrative) of the prototype. I hold role-*stereotypes* to offer one of the resources by which individuals can achieve this. For one, because role-stereotypes, understood as *collections of beliefs about the traits characterizing typical occupants of institutional roles,* can (co-)influence how idealized role-occupants are determined by the individual. Second, role-stereotypes can encompass a *regulative* dimension in the form of self-stereotyping. They are therefore able to *shape* the minds of those who engage with them. And the ways in which the minds of those individuals are shaped can help us to understand how they come to have a standard against which they can measure their own, non-ideal performance.

Group-level R_{ID}

Let me turn to the third level on which idealized conceptions of role-performances may be determined: on the level of *processes that institutional groups themselves deploy* and which I will broadly summarize under the term of *processes of institutional ideal formation*. The idea here is that institutional groups *themselves* exhibit internal mechanisms that aim to provide a conception of an idealized role performance. Especially large and complexly structured institutional groups are prone to implement such processes of ideal formation. These processes include, or are included in, amongst other things: *onboarding-programs*, *instructional courses* or *manuals*, *mission statements* and *vision statements*, *codes of conduct*, *compliance*

policies, *strategy maps*, *scorecards*, *corporate social responsibility* (CSR) *initiatives*, or internal communication tools like *in-house newspapers*, *bulletins*, *corporate podcasts*, *blogs*, etc. I will give one, real-life example of such a processes by looking at the so called "*Disney look*", i.e., regulatory guidelines for employees at the World Disney World Resort.

But before I continue with a depiction of these processes of institutional ideal formation, let me give you an argument as to why such collective management of R_{ID} is important in the first place, especially for large and complexly structured institutional groups. It can be summarized this way:

(i) The potential heterogeneity of individual role-occupants increases with group-size.

(ii) An increase in the heterogeneity of individual role-occupants correlates with an increase of role-ambiguity.

(iii) from (i)-(ii): Especially large, and complexly structured institutional groups are prone to exhibit role-ambiguity amongst their members.

(iv) Role-ambiguity is a potential threat to Institutional group agency (see Ch. 4).

(v) R_{ID} can moderate the influence of role-ambiguity.

(c) From (iii)-(v): Especially large, and complexly structured institutional groups, e.g., multi-national corporations, exhibit pressures to deploy *processes of institutional ideal formation.* These processes facilitate R_{ID} and secure the agency of institutional groups against the threat of Role-Ambiguity.

Let me briefly elaborate on this: What I mean with the *potential heterogeneity* of individual role-occupants concerns especially the use of discretionary powers in interpreting and applying the tasks and functions of the roles that multiple individuals occupy. Above, I sloganized this use of discretionary powers by stating that individuals "have to make up their own mind" about how to perform the tasks and functions of their roles, most importantly because the design-specifications of institutional roles do not exhaustively prescribe the conduct of role-occupants. Now with an increase of group-size, more and more individuals will be confronted with situations in which they "have to make up their own mind" about how to perform the tasks and functions of their roles. Now my conjecture is simply that the potential heterogeneity of role-occupants correlates with group size *insofar* as the results of discretionary judgements will yield a greater variety, the more individuals are in a position to issue such judgments. More people "making up their own minds" leads to a more diverse, or heterogenous outcome.

I also argued (in Ch. 4.2.) that this variation of output poses a potential threat to institutional agency. Individuals can fail to come up with satisfying ways to contribute to the institutional group's action because of their discretionary judgments. Discretionary powers, especially regarding the interpretation of action-guiding standards, then can turn *toxic.* They can become detrimental to the performance of tasks and functions of institutional roles; as well as to realizing overall institutional group actions. In such situations, the toxic use of discretionary powers can thwart, circumvent, or impede the realization of the institutional group's goal, to which the role-occupant is supposed to contribute to.

Now the way I portrayed this problem in Ch. 4.2. focussed on examples where just *one* individual makes toxic use of her discretionary powers. But note that the threat that such toxic use of discretionary judgment poses to an institutional group's agency *increases* with the number of individuals involved. One "bad apple" so to speak, might not lead to an outright failure of an institutional group to perform a certain action. But the problem seems to reside with the whole basket. So if especially large, and complexly structured institutional groups are prone to exhibit role-ambiguity amongst their members, they are also especially prone to be faced with agency-threatening, i.e. toxic role-performances. And as I've argued in this chapter, Role-Idealizations are able to counter such threats of toxic role-performances by means of *disambiguation*, i.e., by providing the individual role-occupant with heuristic resources and interpretative guidelines on how to perform their roles, especially in cases of uncertainty.

So, I take it, especially large, and complexly structured institutional groups, e.g., multi-national corporations, exhibit pressures to deploy mechanisms aiming to mitigate the threats of toxic role-performances of their members. And if Role-Idealizations provide a standard against which individuals can measure their own, idiosyncratic and non-ideal role-performances, then *actively providing an ideal standard can mitigate the extent to which individuals will deviate from the thus derived idealized role-performance*. Actively prescribed idealized standards then have as their function the direct reduction of role-ambiguity. Rather than each individual orienting herself according to her own idealized understanding, groups are prone to deploy mechanisms by which their members *converge* on an ideal or prototypical role-performance and *gain a shared understanding of what an ideal role-performance amounts to*. This, in turn, mitigates deviation, aligns interpretations and on this basis facilitates cooperation.

I will now argue that *processes of institutional ideal formation* provide such ideals. Such mechanisms, which facilitate R_{ID}, can sometimes be subtle, other times they provide idealized role-performances in rather obvious ways.

Example case: The Disney Look

Instead of going through the myriad of ways in which institutional groups can actively provide ideal standards of role-performances for their members, I will now discuss one, yet ostentatious real-life example of such a process of *institutional ideal formation:* The so called *Disney Look* of the Walt Disney Company. Describing a real-life case inevitably involves some degree of speculation. By doing so, however, I hope that the reader can grasp the idea why other group-level phenomena (e.g., *onboarding-programs*, *instructional courses*, *mission statements*, *code of conducts*, *internal communication* tools like *in-house newspapers* or *corporate podcasts*, etc.), too, function - amongst other things - as tools through which institutional groups can disseminate actively constructed and officially authorized ideal conceptions of role-performances, which members pick up and "plug into" their heuristic strategies.[157]

My example focusses on the Walt Disney Company, a multinational corporation employing 220,000 people as of October 1, 2022 (cf. Walt Disney Company 2022, 2). Amongst other sub-branches, the Walt Disney

[157] Communication- and Management-studies, which are usually interdisciplinary efforts of sociological, psychological and business scientific approaches, provide for a longstanding and theoretical discussion of this phenomenon. For further reading on this subject I especially recommend Hochschild 1983; Kühl 2013, 2018.

Company operates the *Walt Disney World Resort* in Florida. With over 77.000 individuals employed in more than 3000 different job classifications (cf. Walt Disney Company 2020, 1), the Disney World Resort is - as of this date - *the biggest single-site employer in the world*. Operating the Disney World Resort is a real-life case of massive-scale cooperation occurring on a day-to-day basis, including the functional integration of various branches, like e.g., entertainment-operations, administration, hospitality, gastronomy, facility management, including sub-components like security, health services, infrastructure management, etc.[158] The overall goal of the *Walt Disney World Resort* is to generate and maximize profits, primarily by offering its services to visitors in exchange for money. And in order to ensure satisfaction of customers (and thereby to maximize profits), the company has developed ways to coordinate the interaction of the employees with these visitors, i.e., extensive guidelines and training processes that group-members have to go through.[159]

One particular interesting guideline is the so called "Disney Look". This guideline explicitly specifies how individual role-occupants ought to conduct themselves when working in the resort, and how they should interact with visitors. The guideline, e.g., regulates things like the *Clothing Lengths*, *Fabrics and Patterns, Footwear*, *Hair Accessories*, *Headwear*, and *Jewelry* for both "Costumed" and "Non-Costumed" role-occupants (cf. Walt Disney Company 2014). It even specifies what counts as "appropriate undergarments“, which role-occupants ought to wear "at all times" (Walt Disney Company 2014, 7). The Disney Look is also backed up by sanctions for failure to meet the standards of the guideline. It explicitly states that "failure of any Cast Member to adhere to these or any subsequently established or modified standards will result in appropriate disciplinary action, not excluding separation from the company" (Walt Disney Company 2014, 24).

In accordance with what I’ve established so far, running the Disney World Resort is a case of cooperation occurring on a massive scale. It can be described primarily as a matter of each of the 77.000 individuals fulfilling their role-specific tasks and duties, where the structural integration of these thousands of tasks and functions is achieved via the features of institutional agency described in the third chapter (i.e., hierarchical organization of roles to achieve the groups overall goal, the division of tasks into sub-tasks, proxy-actions, etc.). One feature of (some of) these role-specific tasks and duties is that they are to be performed in presence of, or in interaction with visitors of the resort. And so one way to understand the *Disney Look* is to see it as a part of the institutional roles’ *design-specification,* that specifies *how* the tasks and functions of the institutional role-occupants are to be performed because of this feature.

But what I also want to argue for is that the *Disney Look* has - besides its primary function of spelling out design-specifications - another, and perhaps more subtle function. I hold it to actively provide the role-occupants with a shared understanding of what an *ideal role-occupant* at Disney World Resort amounts to. This, in turn, aims to *mitigate the extent to which individuals will deviate from the thus derived idealized role-performance*. Thus, the *Disney Look* is one of the means by which *institutional ideal formation* is

[158] To emphasize the scale of cooperation occurring at Disney World Resort, notice that approximately *58.000.000* customers per year visit the Disney World Resort. The resort includes 30 resort hotels with more than 30,000 rooms, four theme parks, two water parks, a sports complex, four golf courses, 400 dining options, a monorail system, a gondola lift system, a private security service including a fire brigade, etc. (Walt Disney Company 2020).

[159] Again, such mandatory training processes that novice members have to attend can be understood as a way to ensure that individuals undergo the process of Role-Internalization**.**

brought about on a group-level, and it functions as a mechanisms by which the role-occupants ought to *converge* on an ideal or prototypical role-performance. So what brings me to believe that it actually can fulfill this function?

First, besides regulating appearances, the guideline tries to provide the reader with an ideal way in which the employees are ought to *conduct themselves* in a quite general manner. For example, the guideline explicitly states that role-occupants ought to view themselves as *representatives* of the company they work for when fulfilling their tasks and functions:

> "Regardless of your role, when you take pride in your appearance, you become a role model for those around you, and you convey the attitude of excellence that has become synonymous with the Disney name" (Walt Disney Company 2014, 1).

Note that this open-ended appeal does not pertain directly to the institutional role's *tasks and functions*, and it does not directly prescribe how *particular* tasks and functions are to be performed.[160] Rather, it prescribes the conduct of individuals on a *general* level that is supposed to underwrite the performance of the institutional role's *specific* tasks and functions. The guideline prescribes a *general level of conduct* for the employees, but not *specific* ways to perform *particular* tasks.

The management and monitoring of the role-occupant's *general conduct* does not exhaust itself in such encouragements to engage in processes of self-directed role-stereotyping. The guidelines, by asking role-occupants to "convey attitudes of excellence" (ibid) or to be "courteous, conscientious and [to] exhibit good judgment" (Walt Disney Company 2014, 3) also attempt to foster a set of *virtues* that role-occupants are supposed to realize. These virtues aim to prescribe the conduct of individuals on a *general* level, i.e., they aim to *underwrite* the performance of their institutional roles' tasks and functions, rather than specifying them. Additionally, the guideline asks the role-occupants to *counterfactually reason*, whether *any unspecified* action will "be in the best interest of our Disney show" (ibid).

> "The Disney Look appearance guidelines present a standard to be upheld by all Cast Members to ensure the best show to our Guests. The expected behavior of each Cast Member is to be courteous, conscientious and exhibit good judgment at all times to benefit the organization, fellow Cast Members, and Guests. For this reason, *if you are ever in doubt about the appropriateness of your appearance, please keep in mind that anything that could be considered distracting or not in the best interest of our Disney show will not be permitted"* (Walt Disney Company 2014, 3) [own emphasis].

Again, I hold this to aim at influencing the *general conduct* of the employees, from which more specific ways to perform particular tasks may be deduced. By encouraging employees to ask themselves: "is my appearance or action in the best interest to ‚the Disney Show'?", the guideline incentives the individuals to engage in the heuristic strategy of determining what *ideally* would be in the best interest to the "Disney

[160] At least not in the way in which the design-specifications of one's role as a worker at Caterpillar prescribe *how* to deal with broken-down machines (See Ch. 4.2).

Show“, and then either trying to act towards the so deduced envisioned outcome; or to refrain from actions that do not fall into the idealized course of action.

So the upshot is this: operating Disney World Resort is a case of massive-scale cooperation, which requires the coordinated efforts of thousands of individuals. A role-based explanation can lift a lot of the explanatory burden of how such massive-scale institutional action (or as it is an *ongoing* action: institutional *activity*) can come about. Primarily, the Disney World Resort provides its services through individuals fulfilling the tasks and functions of their institutional roles.

But besides this, processes of *institutional ideal formation* also seem to play a function. I identified the *Disney Look* to be one of (several of) these processes by which institutional ideal formation comes about. This guideline is not *exclusively* a process of institutional ideal formation as it can also be understood as part of the design-specifications of the institutional roles. Further, it is not the *only* process of institutional ideal formation, and individual role-occupants at Disney might be subject to more than one of these processes.

But what I wanted to exemplify with this guideline is that institutional groups are prone to deploy processes that manage and govern the actions of role-occupants in a way that goes *beyond* the mere fulfillment of their tasks and functions. By specifying a (role-independent) level of *general, non-situational conduct* of the employees, the guideline can be plausibly said to foster a sense of what *being an employee in general* amounts to, regardless of one’s role, the actual practical implementation of it in the resort, or the circumstances under which its tasks and functions must be performed. And by actively disseminating such an idealized *standard* against which individuals can measure their own non-ideal performances, the guideline enables the individual employee to engage in (counterfactual) reasoning in situations of uncertainty. By being provided with such an idealized standard, individuals may be able to give answers to questions like "*Would an ideal employee at Disney World Resort behave in such and such a way?*", or "*Could this course of action be said to be in the best interest of Disney?",* or *"Is this behavior generally appropriate for an employee at Disney World?“*. Because this idealized conception is created, implemented and maintained by a designated and authorized sub-group, which specializes in issuing such policies (usually the *HR-Department*), it is brought about by an *group-level* process. Being collectively created, maintained by a group-level process, the provided ideal is thus fixed to an *external* standard. It is not up to the individual employee to determine it, e.g. by introspection. Importantly, by establishing an ideal conception of an employee which is shared (in a weak, distributive sense) by the group members, multiple individuals will tend to *converge* on the same answers when they each ask themselves these above stated questions, thereby *disambiguating* role-performances and countering threats of toxic role-performances.

5.4. Summary

The fifth chapter developed my account of Role Agency (RA). In the beginning of this chapter, I argued that RA should be best understood as an exercise of *Role Perspective Taking* (R_{PT}), which is based on the interrelated, dual-mechanism of *Role-Internalization (*R_{IN}) and *Role-Idealization* (R_{ID}). Pace Schmitz’s *role-mode*, I argued that such R_{PT} ultimately is to be analyzed as a matter of the *content* of the individuals’ intentional states. As a consequence, I argued that R_{PT} is ultimately based on an individual’s capacity to adapt the *role-specific reasons for action* that are constitutive of acting within one’s institutional role. To

adapt the perspective of an institutional role is to adapt a set (or framework) of role-specific beliefs, desires, goals. On this basis, I then identified *Role Agency* to constitute *a subtype of individual agency*, which is both "bigger" and "smaller" than individual agency.

The next step of my investigation saw answering the question of how such role R_{PT} could be said to come about in the first place. And in order to explain how role-occupants may come to understand and have control over their institutional role perspective, I argued that the two processes of Role-*Internalization (*R_{IN}) and Role-*Idealization* (R_{ID}) carry this explanatory burden. R_{IN} turned out to be dynamic and multi-dimensional process. I argued for the separation of two dimensions, and on this basis derived four different categories of Role-Internalization: On the one hand, Role-Internalization encompasses a *theoretical* and *practical dimension*. On the other hand, Role-Internalization encompasses both *formal* and *informal* aspects of one's role which need to be internalized. I then argued for the usefulness of keeping these dimensions (formal-theoretical, informal-theoretical; formal-practical, informal-practical) apart from one another. However, it turned out that Role-Internalization is not the whole story of how individuals relate to their Role-Perspectives.

To capture the ways in which individuals come to critically *distance themselves from*, and *reflect on* their roles, I turned to the process of R_{ID}. Here, I argued that R_{ID} should be best thought of as an *Imitate-the-Ideal* Heuristic, where first individuals have to determine an idealized (or prototypical) role-occupant, and in a next step try to imitate this idealized behavior (e.g., action, judgment, choice, decision, preference, or opinion etc.). I argued, *pace Ludwig,* that idealized role performances have to be fixed to an *external* standard in order to be *regulative*, i.e., action-guiding. Finally, I tried to give some insights into the creation, management and monitoring of Role-Idealizations. Here, I argued that this should be thought of as a social and collective process. I identified three levels of the social and collective creation and management of such idealizations: The interpersonal-level of R_{ID} by which individuals make use of (multiple) role models; the environmental-level of R_{ID} in which individuals base such idealizations on stereotypes; and the group-level R_{ID} which is constituted by processes of institutional ideal formation.

Let me briefly meditate on the limitations and scope of my account of RA, and whether I have achieved the goal of providing for a deepened understanding of institutional roles and the way in which individuals relate to them. As to the scope and limitations of my claims, recall that the relation between my account of RA and that of role-based explanations of institutional agency is strictly meant to be *supplementary*. By this, I mean that I do *not* wish to argue that my account is able to *replace* any theory of institutional action. My account of RA is not a theory of institutional agency *per se*, and it certainly does not attempt to explain what institutional group actions are; or how they should be conceptualized. Thus, RA does rely on, and presupposes theories, which provide a role-based explanation of institutional actions. However, it does not *favor* any particular role-based theory. As we saw, accounts such as those of Ritchie (2020a), Miller (2001, 2010) or Ludwig (2017a; 2017b) invoke the concept of roles in order to explain the actions of institutional groups. We also saw that they converge on both the concept of institutional groups *and* of institutional roles.

One reason for why I hold my account to be compatible with such theories, is that my account of RA does not suggest that the existing characterizations of institutional roles are fundamentally *false* or *incorrect*. Institutional roles can indeed be analyzed in terms of their "design specifications", which include the tasks

and functions that role-occupants need to fulfill, and the relations, including the relations of power, in which roles stand towards on another. Rather than revoking this definition of institutional roles, my argument for Role Agency is more modest in nature. By developing the concept of RA, I merely wanted to make the case that such characterizations do not *sufficiently* illuminate the relation between institutional roles and the individual agents who occupy them. Role Agency therefore aims to provide for a more detailed and fine-grained description of institutional roles, which gives us insights into this mostly unexplored relation. So again: Role Agency is an attempt to fill a lacuna in the existing literature, and I hold it to be compatible, rather than competing with role-based explanations of institutional agency.

It does not directly follow from the fact that my account *aims* to fill this lacuna, that the account can *actually improve* on the shortcomings of the existing theories. So let us look at the question whether my account really *can* improve such role-based explanations of institutional agency. I'm declined to answer this question affirmatively.

But let's step back and remind us of these shortcomings. The initial and basic motivation for my account of RA was developed in light of how individual's come to use the *discretionary powers* that are vested in their institutional roles. Now the very use of discretionary powers, in and of itself, is not a problematic aspect of giving a role-based explanation of institutional agency. However, as I tried to show, it *can* eventually impede, or jeopardize an institutional group's capacity for action, either in situations where individuals willfully suspend their discretionary powers; or by individuals being misguided in their application.

My theory can, of course, not change the complex and messy nature of social reality, where such failure will actually occur. So all of this is, of course, not to say that role-occupants will never fail to live up to the standards that their roles prescribe, or that they will never make misguided use of their discretionary powers. The modest goal of my account of Role Agency is to provide us with the means necessary to make these problems *intelligible* in the first place. So I merely hold that, in virtue of my account of RA, both problems, as well as potential strategies that role-occupants may employ to mitigate them, can be addressed more specifically. So how does the explanatory power of my account fare in light of these problems?

The first *Problem of Discretion* can be addressed more specifically primarily because my account of RA allows us to see how, by which means, and to which extent individuals come to understand and control their institutional roles in the first place. The explanatory burden here is carried by the process of Role-Internalization. This is an improvement on the existing literature insofar as Role-Internalization makes intelligible, how occupying an institutional role has both an *interpretative* and *applicatory* dimension; and how it encompasses the acquisition of both *formal* and *informal* knowledge. Making intelligible how individuals come to understand and control their institutional roles is a pre-requisite to understand what's going on (or rather: going wrong) in cases of working to rule, which I argued to cause institutional stupor. Recall that an institutional role-occupant can be said to be working to rule, if she adheres strictly to the prescribed functions and tasks of her assigned role (i.e., if she adheres strictly to the roles formally established design specifications) and meticulously follows the entailed rules governing the performance of her tasks and functions "to the letter". Initially, working to rule could be understood as a straight-up counterexample to the claim that institutional group agency can be reductively explained by individuals fulfilling the tasks and functions of their roles. But this is only apparently so.

Instead, my account of RA gives us the means to analyze how the institutional role's *formal* and *informal* dimensions are inextricably linked to one another. Recall here that an institutional role's formal dimension can be either *symbiotic* with, or *antagonistic* to acting on its formal dimension, and vice versa. *Working to rule* can be understood an instance where such a symbiotic relation between the formal and informal dimensions of performing one's tasks is willfully suspended by the individual, leading the individual to refrain from applying the informal aspects of her role-performance to accompany its formal ones. So being confronted with cases of institutional stupor, we now have the analytical tools to explain what is happening: Is institutional stupor occurring because the individuals involved changed the way in which they carry out their tasks and functions? If so, has the formal, or informal applicatory dimension of their roles changed? Do they "work to rule" and suspend the application of the informal dimension of their roles? Or do certain informal aspects of their role-performance actually hinder them to perform their formally defined tasks and duties?

These questions reveal how occupying an institutional role can be a complex, multi-facetted achievement, and that it can require much time and effort to do so. But these questions also show the necessity of going beyond the established accounts of institutional roles, in order to answer them. My account thus improves our understanding of how individuals adapt their institutional roles against the backdrop of existing explanatory projects. Whereas, e.g., the role-mode of Michael Schmitz rightfully acknowledges that occupying an institutional role can encompass the adaptation of role-specific *vantage points,* my account of Role Agency can explain such a form of perspective taking without invoking the notion of some unspecified "we", or a form of "group mindedness". This form of perspective taking, which I argued to consist of individuals adopting an externally defined framework of beliefs, desires and goals, does not presuppose that individuals represent each other as co-subjects, or that such a role-perspective has to make reference to some other form of collectivity. This, too, is an improvement on the existing literature, as I argued that Schmitz's account was inapt to be applied to especially large and complexly structured institutional groups.

The second *Problem of Discretion*, the problem of *role-ambivalence,* can be addressed more specifically too. To this end, I argued that an individual's functioning within a role can be partially impaired by the discretionary powers vested in the institutional role. I tried to show how discretionary powers may ultimately lead to the *breakdown* of institutional agency, because they open up the possibility for a gap between actual and expected performance of role-occupancy. Recall my exemplificatory case of *racial profiling* by the police. Here, a policewoman's use of her discretionary powers to determine what "suspiciousness" means led to her failure to contribute to her department's goal of lowering crime rates, and it undermined the institution's overall goal of maintaining social calm and upholding social peace in the community. However, *technically speaking*, the policewoman could be said to merely fulfill the tasks of her role. After all, she stops and frisks individuals who *she* interprets to be "suspicious". She does so, however, in a way which impedes or jeopardizes the overall goal she is supposed to contribute to. I think that my concept of Role-Idealization can help us to see what is going on (or rather: going wrong) in her interpretation of how she should ideally perform her tasks.

I think it is fair to stipulate, that our imaginary policewoman, *if she was to gain a critical distance towards her institutional role*, could come to ask herself: "What aspects of my performance *as a policewoman* are based on *my interpretation* of this role? And is my interpretation really congruent with how to *ideally* act as

a police-woman?" Regarding her motives for determining the standard of "suspiciousness", the police woman then might come to ask herself "is stopping and frisking only individuals of one particular ethnic group *really* congruent with how to *ideally* perform this task as a police-woman?" I hold that *this* interpretative level of critical reflection and evaluation of her role-performance could enable her to become aware of her biases, and subtract or suspend the idiosyncratic, and contingent aspects of her role-performance. So in this scenario, we could explain the individual's impairment to function within her role by pointing to a *lack of critical and reflective distance* towards her role-performance. And as I've argued, R_{ID} is the process through which an individual comes to have such a *critical and reflective distance* towards her role-performance. The process of Role-Idealization crucially implies that a role-occupant can come to identify the *discrepancies* between the idealized standard and her own, non-ideal performance. It is on this basis, that she may *regulate* her own, non-ideal behavior in accordance with the so ideal standard.[161] Again, I hold this to improve our understanding of such idealizations, partly because it goes beyond existing approaches that try to emulate such idealizations, e.g., in terms of role-*identification*. With Schmid, e.g., we saw how individuals may keep a *distance* to their institutional roles, and that this role-*distance* constitutes a functional aspect of *acting qua role-occupancy.* But this sort of distance, is only *one* way in which individuals may come to have a reflective, and evaluative attitude towards their roles. With the process of Role-Idealization, we can also examine an individual's capacity to critically reflect and evaluate their own role-performances in light of a *prototypical,* or *ideal* role-performance.

[161] This scenario assumes that the *ideal* that the policewoman is enabled to strive towards is in itself not racist. Rather, it is her own, racially biased interpretation that leads her to determine that certain ethnic groups are more "suspicious looking" than others. But imagine a contrasting case, where the racial profiling of our imaginary policewoman is the result of her trying to imitate an ideal, which is in itself based on a racist ideology. For example, imagine a police officer during South African Apartheid. Now if such a policewoman was to engage in the process of Role-idealization, she might hold her actual, racist police work to be very well justified by such an imagined ideal. To this end, my analysis of the *creation*, *management* and *monitoring* of Role-Idealizations could explain how an individual might end up with a racist ideal of a policewoman's conduct in the first place. After all, the racist ideology in South Africa was ubiquitous and fixed on, as well as promulgated by the interpersonal, cultural, as well as institutional levels of that society. We could, e.g., stipulate that her role-idealizations came about on an interpersonal-level, e.g., by having racist role models within her police-department. Also, such racist ideals of a policewoman's conduct surely were part of her social and cultural surroundings. And it's plausible to assume that her institutional group itself exhibited internal mechanisms that aimed to provide a racist ideal regarding her role performance. So within her police department, there probably were processes of *institutional ideal formation* that provided the policewoman with the ideal of enacting racist policies. The conclusion here is that during Apartheid, trying to be an ideal police officer - sadly - implied that one was to strive towards the (paradoxically sounding) *ideal of being a racist.*

6. Conclusion

The goal of this thesis was to examine and assess theories of institutional group agency, i.e., theories which target the capacity of institutional groups for action. The main effort of this endeavor was to answer questions as to whether, and to which extent, groups could be said to be agents; and how to conceptualize such a form of agency. The result of my investigation was that institutional group agency can be best explained by so called *role-based* explanations. But I also argued that the standard way to describe institutional roles, which is used by such theories, is wanting. This motivated the development of my own account of Role Agency.

Arriving at this conclusion first saw a clarification of some basic concepts at hand, i.e., action, collective action, institutional groups and group agency. The second chapter then explored two broad avenues for explaining institutional group agency. The first path of r*ealist, non-reductive (or inflationary) theories of group agents* saw that the agency of institutional groups is to be explained by arguing for them to be genuine, non-reductive agents in their own right. I here examined four different, yet related theories, which either argued for a *functionalist* or *interpretivist* view on group agents. I tried to show that the examined theories are all vulnerable to reductionistic charges, and that we should therefore refrain from positing the existence of such non-reductive agents.

The second part of the second chapter examined theories arguing for the agency of institutional groups to be *reducible to* the capacity of the institutional groups' members for *collective action*. I claimed that *theories of collective action,* which explain the agency of *institutional* groups, but start out with small-scale cases of collective action, come to face the so called *Upscaling Problem*. As the direct attempts to apply small-scale analyses of collective action to institutional groups are thereby frustrated, I examined the theories of Christopher Kutz and Raimo Tuomela. Both authors explicitly target the agency of large, and complexly structured institutional groups. However, I tried to argue that their accounts have serious shortcomings when it comes to explain compartmentalized and individualistic nature of institutional group action.

In lack of a satisfying explanation of institutional group agency, I then examined so called *role-based* explanations of institutional agency in the third chapter. By *role-based*, I refer to theories that, at their core, argue for the claim that an institutional group action consists of (and consequently can be reduced to) the individual contributory actions of its members, who *act in their assigned* roles, or *qua role-occupancy*. According to such theories, to say that members of an institutional group act in their assigned roles, is to say that they perform the functions and tasks definitive of the roles they occupy. Chapter 3.1. started with the concept of institutional groups and institutional roles employed by such accounts. The upshot of this section was that institutional groups are best viewed as interrelated structures of institutional roles, which in turn are defined through tasks and functions that an individual role-occupant must perform. In Chapter 3.2., I investigated how the concept of institutional roles figures in explaining institutional agency. This required us to overcome two problems that reductive explanations of institutional group agency face: The problem of *Action Integration* and that of *Diachronic Group Constitution*. I argued that role-based explanations can solve these problems by highlighting critical features of institutional roles in relation to institutional agency: First, institutional roles provide the individual with so called *role-based reasons* for

action. Second, institutional roles *specialize* the actions of group members, leading to a differentiation of tasks and a division of labour. Third, the actions of individual role-occupants are functionally integrated into a *layered structure*. Finally, institutional roles allow for *representative*, or *proxy action*. During the course of these two sections, I also tried to convince the reader, that role-based explanations of institutional agency avoid the shortcomings of the theories of collective action examined in Ch. 2.2., all the while refraining from positing the existence of a non-reductive, genuine group-agent endorsed by the theories in Ch. 2.1. With such a role-based explanation of institutional agency at hand, I turned to the question whether it can capture the *anonymity* and *compartmentalization* of institutional group action. In Chapter 3.3. I drew on Seumas Miller's "Collective End Theory" (CET) of joint action and argued that role-performances of individuals can be *functionally integrated* to achieve an end, without the necessity of common knowledge being involved in this process. I argued that such functional integration is possible, because institutional roles are *agent-ambiguous* but *action-specific*. It is on this basis, that cooperation within institutional roles can occur anonymously.

In Chapter 4., I argued that this agent-ambiguity of institutional roles implies that they cannot exhaustively prescribe action. Thus, the performance of one's institutional role necessarily encompasses so called *discretionary powers*. I then claimed that these discretionary powers, which allow for flexibility and adaptability of a role's tasks and functions to specific circumstances, changing environments and unprecedented situations, can also lead to the breakdown of institutional agency. This was demonstrated by the *two Problems of Discretion*, which the existing role-based explanations of institutional agency fail to make sense of. And in order to better understand how individuals come to deal with situations where their discretionary powers must be used to uphold their roles' tasks and functions, closer attention must be paid to the relation between individuals and the institutional roles they occupy.

Hence, the fifth chapter of my thesis developed my account of RA, which aimed to provide such a closer look on this relation. I argued that RA can shed light on the under-theorized and neglected relation that holds between individuals and the institutional roles they come to occupy. So here's a brief summary of my account of RA: RA consists of a modification of individual agency that individuals engage in when acting *qua role-occupancy*. As such, it is fundamentally based on the individual agents' capacity for *Role Perspective Taking* (R_{PT}) which, in turn, can be analyzed in terms of the individuals adopting a framework of role-specific reasons for action. To explain how such forms of R_{PT} can come about, my account relied on the two interrelated processes of Role-Internalization (R_{INT}) and Role-Idealization (R_{ID}). R_{INT} describes the processes by which individuals come to understand and apply the functions and tasks of their institutional roles. R_{INT} can be further analyzed by distinguishing formal and informal aspects of internalizing one's role; and by distinguishing its practical and theoretical aspects from one another. R_{ID}, in turn, does not aim to explain the ways in which individuals come to understand and apply the functions of their roles. Instead, R_{ID} should be understood as a *social heuristic*, that aims to describe what happens when individuals ask themselves how to *ideally* act out their institutional roles. These role-*idealizations* can provide a regulative, action-guiding standard against which individuals can measure their own, non-ideal performances, especially in situations of uncertainty and ambivalence. The content of such ideals can be determined in different ways, including an interpersonal level of role-modeling, a form of self-stereotyping informed by cues in the individual's environment, and by what I called group-level processes of institutional ideal formation.

A brief outlook on future paths of inquiry

While I hold my concept of Role Agency to provide new and useful insights, there is, of course, further work to be done. So let me briefly sketch some further paths of inquiry, which might be taken from here on. One such possible inquiry concerns the question whether my concept of Role Agency could be extended to other social roles. Could Role Agency be said to apply to social roles like, e.g., being a parent, a woman, a spouse, etc.? If so, how, and in which sense might this be possible? One way in which my account focussed on *institutional roles* was that they, in contrast to social roles in general, were conceived of as *agent status roles*, i.e., roles that involve role-occupants performing tasks and functions in order to contribute to an overall group's goal.

However, scholars have argued that some *social* roles, e.g. gender roles such as *woman*, or racial roles such as *being white* (see: Warren 2001) are inherently *performative* too. If e.g., gender roles, like institutional roles, actually require some form of *role performance*, could it be that they also involve some form of Role Agency? Could we, e.g., then compare the processes by which individual *internalize* such social roles with the process of Role-Internalization aimed at capturing *institutional* roles? And could this line of thought be fruitfully applied to the process of Role-*Idealization* too?

Another open question is how multiple, different roles relate to one another in terms of the Role Agency they provide for their occupants. Sociologists have long pointed to the fact, that individuals can experience intra- and inter-role conflicts, i.e., situations where the demands and obligations of different institutional roles clash with one another. However, I think it would also be of interest to investigate the relations of social and institutional roles not only in terms of such conflicts, but also in terms of mutual, *symbiotic benefits*. In what way could the Role Agency of an individual be said to "*travel*" outside of an institutional group context? Is there, e.g., a mutual influence of *being a mother* and *occupying the role of a judge*? If so, is this influence *symbiotic,* or *antagonistic*?

Lastly, my account of Role Agency did not answer questions regarding the *ethical* dimension of occupying institutional roles. Throughout this thesis I argued that the agency of institutional groups is reducible to the agency of individuals who *act qua role-occupancy*. Given that this is a reasonable claim to accept, we would expect institutional roles to figure prominently in the debates about the *ethical* dimensions of institutional agency, e.g., concerning questions regarding the *moral responsibility* of institutional groups for their actions. Surprisingly, Barber & Cordell (2023) recently asserted that this does not seem to be the case, and that there exists, to this date, no "ethical theory of social roles as such" (Barber & Cordell 2023, 1). One way to advance the debate about the ethics of institutional roles would see an investigation of my account of Role Agency in terms of the moral quality of the rights, duties and responsibilities that individuals possess in virtue of their institutional roles, or the moral character of the rules that role-occupants must follow. For example, Collins (2023) recently argued that, in situations where the moral obligations and the obligations we have *qua role-occupants* clash with one another, we have *moral obligations to alter our role-obligations*.

Another way to assess the moral nature of institutional roles would be to look at them from a consequentialist point of view in terms of the moral quality of the subsequent actions they give rise to. These questions are often debated in light of the (forward- or backward-looking) moral responsibility that we attribute to institutional groups for their actions. Are, for example, members of institutional groups

morally responsible for their immediate individual actions only? Or should we attribute moral responsibility for the large-scale actions which result from the coordinated contributory actions of the group members too (cf. Skerker 2020, 275)? Generally, higher levels of power and influence on a group's action seem to suggest a higher level of moral responsibility that we attribute to an individual for her role-related actions. One of the merits of giving a role-based explanation of institutional agency is that it allows for a deepened analysis of the ways in which relations of power are distributed among institutional groups. But my discussion of role-based cooperation may also raise questions how to attribute moral responsibility in cases where individuals do not explicitly know the group's overall goal they are contributing to; or cases in which they have been deceived to believe that they are contributing to something else. Should we, e.g., hold the Calutron Girls accountable for contributing to the goal of building the atomic bomb, although they were unaware of what their actions contributed to?

A third, and promising option to examine the moral status of institutional roles would be to ask whether my account of Role Agency could be connected to, and thus be analyzed in terms of *virtues*, or *practical wisdom*. One of the merits of my account of Role Agency is that it lets us consider how individuals can spend a great amount of time and energy on "mastering" their roles. Role Agency allows us to see how individuals come to gradually internalize and consequently balance their roles' tasks and functions against each other, and how they come to "grow into" the roles they occupy. My account of Role Agency also allows us to acknowledge that some individuals will be "better" (or "worse") than others at occupying a given institutional role. Saying that acting qua role-occupancy is simply performing the tasks definitive of the role does not, e.g., allow us to see what exactly the difference between an *experienced* role-occupant and a *novice* amounts to. Now given that we can observe such differences, and given that some institutional roles, like *judges*, or *police officers* clearly have great moral relevance, one way to think about Role Agency could be in terms of virtuous (or vicious) character traits, or dispositions that individuals develop and possess in virtue of their institutional roles. A plausible conjecture here would be that an *expert* role-occupant acquired a form of *phronesis*, i.e., practical wisdom which allows her, but not the well-intending novice, to actually make good decisions. If this turned out to be a fruitful approach to the ethics of institutional roles, then subsequent questions could be investigated: Could such role-related virtues be taught? And if so, what would be effective methods for teaching them? And are current methods by which institutional groups enable role-occupants to acquire the capacity for Role Agency actually effective at fostering such practical wisdom? Does, e.g., modern police-training allow individuals to take up the Role Agency of *being a police officer* in a virtuous manner? Could certain *vices* also be taught along the way?

All of this seems to suggest that my concept of Role Agency could yield for fruitful adaptations and extensions. It could provide further clarity for not only understanding the way in which we relate to our institutional roles, but also the ethical ways in which we might *morally ought* to relate to them.

Final remarks

Often times, the social reality, including the reality of institutional groups will turn out to be messy, complex and confusing. Introducing basic concepts, such as institutional roles, is a necessary and welcomed step to make progress in our understanding of it. But more often than not, reality does not oblige to the concepts through which we try to capture it. I hope to have shown that this is the case with the ways in which institutional roles have traditionally been tried to conceptualize. Such attempts to "carve social reality by its joints" will eventually fail to do justice to their multi-facetted and complex nature. I think that our philosophical theorizing should pay respect to the complexity and intricacy of the institutional roles we all occupy, but also to the far-reaching impacts they have on our personal lives. We can get lost in their perspectives, their forms of reasoning, and in the microcosms of norms, rules and ideals that they provide for us. So rather than thinking about them merely in terms of tasks and functions that we ought to perform, we should think of institutional roles in terms of our abilities to act under pretense, to reason counterfactually, to adapt external viewpoints, and to creatively engage with our surroundings. Institutional roles can become deeply immersive and transformative. Any analysis of institutional roles should try to pay respect to the fact that they have the power to deeply change not only the paths our lives take, but also the power to transform ourselves along the way. With my account of Role Agency, I hope to have provided us with a starting point for such an analysis.

This page was intentionally left blank

Bibliography

Abbasianchavari, Arezou & Moritz, Alexandra (2021): The impact of role models on entrepreneurial intentions and behavior: a review of the literature. *Manag Rev Q, 71*, 1–40. https://doi.org/10.1007/s11301-019-00179-0

Adibifar, Karam & Monson, Melissa (2020): Workplace Subjective Alienation and Individuals' well-being, *Journal of Economic Development Environment and People*, *9(3)*, 22-37. DOI:10.26458/jedep.v9i3.669

Aguilar, Jesús H. & Buckareff, Andrei A. (2010): *The Causal Theory of Action: Origins and Issues*, in Aguilar, J. & Buckareff, A (Eds.): Causing human actions: New perspectives on the causal theory of action. Cambridge/ London: MIT Press, 1-26.

Akgün, Ali E., Lynn, Gary S., & Byrne, John C. (2003): Organizational Learning: A Socio-Cognitive Framework. *Human Relations*, *56(7)*, 839-868. https://doi.org/10.1177/00187267030567004

Almquist, Elisabeth M., & Angrist, Shirley S. (1971): Role model influences on college women's career aspirations. *Merrill-Palmer Quarterly of Behavior and Development, 17,* 263–279.

Alonso, Facundo M. (2017): *Reductive Views of Shared Intention*, in: Jankovic, Marija & Ludwig, Kirk (Eds.): The Routledge Handbook of Collective Intentionality, New York: Routledge, 34-44.

Alvarez, Maria & Way, Jonathan (2024): Reasons for Action: Justification, Motivation, Explanation, in Zalta, Edward N. & Nodelman, Uri (Eds.): *The Stanford Encyclopedia of Philosophy* (Fall 2024 Edition). Access via: URL = <https://plato.stanford.edu/archives/fall2024/entries/reasons-just-vs-expl/>. Accessed: 26.02.2025.

Alvesson, Mats (2015): Organizational Culture. In Edgell, Stephen, Gottfried, Heidi & Granter, Edward (Eds.): *The SAGE Handbook of the Sociology of Work and Employment*, London: Sage, 262–82.

Ames, Daniel R. (2004): Inside the mind reader's tool kit: projection and stereotyping in mental state inference. *Journal of personality and social psychology*, *87(3),* 340-353.

Andersson, Åsa (2007): *Power and Social Ontology*. Malmö: Bokbox

Andrews, Kristin, Spaulding, Shannon & Westra, Evan (Eds.) (2021): Special Issue: Folk Psychology: Pluralistic Approaches. *Synthese, 199*, 1685–1700. https://doi.org/10.1007/s11229-020-02837-3

Anscombe, G. E. M. (1958): On Brute Facts. *Analysis*, *18(3)*, 69–72. https://doi.org/10.2307/3326788

- (1963): *Intention*, Cornell University Press, New York

Aristotle: Metaphysics, Book I [Met.], in: Barnes, Jonathan (Ed.): *The Complete Works of Aristotle: The Revised Oxford Translation,* Princeton, N.J.: Princeton University Press, 1984.

Backes, Marvin (2021): Can groups be genuine believers? The argument from interpretationism. *Synthese 199*, 10311–10329. https://doi.org/10.1007/s11229-021-03246-w

Baker, Lynne Rudder (2012): From Consciousness to Self-Consciousness, *Grazer Philosophische Studien, 84*, 19–38.

Balkaran, Stephen (1999): Mass Media and Racism, *The Yale Political Quarterly, 21(1),* 10-13.

Barber, Alex, and Sean Cordell (Eds.) (2023): *The Ethics of Social Roles*, Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780192843562.001.0001.

Beeghly, Erin (2015): What is a Stereotype? What is Stereotyping? *Hypatia*, *30(4)*, 675–691. http://www.jstor.org/stable/24541975.

Bennion, Francis A. R. (2009): *Understanding Common Law Legislation: Drafting and Interpretation*, Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199564101.003.0015

Berkshire Hathaway Inc. (2024): *Official Home Page*. Access via: https://www.berkshirehathaway.com/. Accessed: 07.03.2024

Bermúdez, Jose Luis (2005): *Philosophy of Psychology. A Contemporary Introduction*. London: Routledge.

Berrios, German, E. (1981): Stupor: A conceptual history. *Psychological Medicine, 11(4),* 677-688. doi:10.1017/S0033291700041179

Bix, Brian (1991): H. L. A. Hart and the "Open Texture" of Language. *Law and Philosophy*, *10(1)*, 51–72. http://www.jstor.org/stable/3504835

Blackman, Reid (2023): Explaining Role-Based Reasons, in Barber, Alex & Cordell, Sean (Eds.): *The Ethics of Social Roles*, Oxford: Oxford University Press, 156-174. https://doi.org/10.1093/oso/9780192843562.003.0008.

Bloch, Marc J., & Moorman, Scott A. (1993): Working to Rule and Other Alternate Job Actions. *The Labor Lawyer*, *9(2),* 169–188. http://www.jstor.org/stable/40862200

Blomberg, O. (2018): Review of Kirk Ludwig, From Individual to Plural Agency, Collective Action: Volume 1. *Philosophical Quarterly, 68(272)*, 626-628. https://doi.org/10.1093/pq/pqx048

Blum, Lawrence (2004): Stereotypes And Stereotyping: A Moral Analysis, *Philosophical Papers, 33(3),* 251-289. DOI: 10.1080/05568640409485143.

Bosetzky, Horst (2019): *Mikropolitik: Netzwerke und Karrieren.* Wiesbaden: Springer VS.

Bratman, Michael E. (1987): *Intention, Plans, and Practical Reason,* Cambridge, MA.: Harvard University Press.

-(1992): Shared Cooperative Activity. *The Philosophical Review*, *101(2)*, 327–341. https://doi.org/10.2307/2185537

-(1993): Shared Intention. *Ethics*, *104(1),* 97–113. http://www.jstor.org/stable/2381695.

-(1999): *Faces of Intention. Selected Essays on Intention and Agency,* Cambridge/New York: Cambridge University Press.

-(2006): What Is the Accordion Effect? *The Journal of Ethics*, *10(1/2*), 5–19. http://www.jstor.org/stable/25115848.

-(2009): Shared Agency, in Mantzavinos, C (Ed.): *Philosophy of the social sciences : philosophical theory and scientific practice*, Cambridge, New York : Cambridge University Press, 41–59.

-(2014): *Shared Agency. A Planning Theory of Acting Together*, Oxford/New York: Oxford University Press.

-(2018): Review: From Plural to Institutional Agency: Collective Action II, *Notre Dame Philosophical Reviews, Access via :*https://ndpr.nd.edu/reviews/from-plural-to-institutional-agency-collective-*action-ii/. Accessed: 03.01.2023.*

-(2021): Shared Intention, Organized Institutions, in: Shoemaker, David (Ed.): *Oxford Studies in Agency and Responsibility Volume 7*, 54-80. https://doi.org/10.1093/oso/9780192844644.003.0004.

-(2022): *Shared and Institutional Agency: Toward a Planning Theory of Human Practical Organization.* Oxford / New York: Oxford University Press. https://doi.org/10.1093/oso/9780197580899.001.0001.

Brentano, Franz (2009): *Psychology from an Empirical Standpoint*, London/New York: Routledge.

Brewer, Marilyn B., & Gardner, Wendy (1996): Who is this "We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71(1),* 83–93. https://doi.org/10.1037/0022-3514.71.1.83

Caputi, Jane (2007): Green Consciousness: Earth-Based Myth and Meaning in Shrek. *Ethics and the Environment*, *12(2)*, 23–44. http://www.jstor.org/stable/40339139

Casati, Roberto & Varzi, Achille (2020): Events, in Zalta, Edward N. (Ed.): *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition). Access via: URL = <https://plato.stanford.edu/archives/sum2020/entries/events/>. Accessed: 12.09.2023.

Chalmers, David (2008): Strong and Weak Emergence, in Philip Clayton & Davies, Paul (Eds.): *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion,* Oxford: Oxford University Press, 244-254. https://doi.org/10.1093/acprof:oso/9780199544318.003.0011.

Chant, Sara Rachel (2007): Unintentional Collective Action. *Philosophical Explorations*, *10,* 245–256.

-(2017): Collective Action and Agency, in: Jankovic, Marija & Ludwig, Kirk (Eds.): *The Routledge Handbook of Collective Intentionality*, New York: Routledge, 13-24.

Chant, Sara R., Hindriks, Frank & Preyer, Gerhard (2014): Introduction: Beyond the Big Four and the Big Five, in Chant, Sara R., Hindriks, Frank & Preyer, Gerhard (Eds.): *From Individual to Collective Intentionality: New Essays*, Oxford / New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199936502.003.0001, accessed 25 Jan. 2024.

Chao, Georgina (2008): Mentoring and organizational socialization: networks for work adjustment, in Ragins, Belle R. & Kram, Kathy E. (Eds.): *The Handbook of Mentoring at Work: Theory, Research, and Practice*, London: SAGE Publications, Inc., 179-196. https://doi.org/10.4135/9781412976619

Collins, Stephanie (2023): Role Obligations to Alter Role Obligations, in Barber, Alex & Cordell, Sean (Eds.): *The Ethics of Social Roles*, Oxford: Oxford University Press, 200-216. https://doi.org/10.1093/oso/9780192843562.003.0010

Copp, David (1979): Collective Actions and Secondary Actions. *American Philosophical Quarterly*, *16(3)*, 177–186. http://www.jstor.org/stable/20009757

Crone, Katja (2021): Foundations of a we-perspective. *Synthese, 198*, 11815–11832. https://doi.org/10.1007/s11229-020-02834-6

-(2025): Collective intentionality: why content matters. *Inquiry*, 1–23. https://doi.org/10.1080/0020174X.2025.2451688

Crone, Katja & Gab, Max (Eds.) (forthcoming): Special Issue: Minimal Forms of Shared Intentionality, *Philosophical Psychology*.

Dahrendorf, Ralf (2010): *Homo Sociologicus. Ein Versuch zur Geschichte, Bedeutung und Kritik der Kategorie der sozialen Rolle*, Wiesbaden: Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92592-9.

Davidson, Donald (2001): *Essays on Actions and Events*, Oxford: Oxford University Press.

-(2001a): *Actions, Reasons, and Causes*. In Essays on Actions and Events. Oxford: Oxford University Press.

-(2001b): *The Logical Form of Action Sentences.* In Essays on Actions and Events, Oxford: Oxford University Press.

-(2001c): *How is Weakness of the Will Possible?* In Essays on Actions and Events, Oxford: Oxford University Press.

-(2001d): *Mental Events*. In Essays on Actions and Events, Oxford: Oxford University Press.

-(2001e): *Agency*. In Essays on Actions and Events, Oxford: Oxford University Press.

-(2001f): *Freedom to Act*. In Essays on Actions and Events, Oxford: Oxford University Press.

-(2001g): *Intending*. In Essays on Actions and Events, Oxford: Oxford University Press.

-(2001h): *The individuation of Events*. In Essays on Actions and Events, Oxford: Oxford University Press.

Davis, Wayne (2010): *The Causal Theory of Action*, in O'Connor, Timothy & Sandis, Constantine (Eds.): *A Companion to the Philosophy of Action*, Malden/Oxford: Wiley-Blackwell, 32-39. DOI:10.1002/9781444323528.ch5.

De Botton, Alain (2008): *The Pleasures and Sorrows of Work*, New York: Pantheon Books.

Dennett, Daniel (1971): Intentional Systems. *The Journal of Philosophy, 68(4)*, 87–106. https://doi.org/10.2307/2025382.

-(1981): *Brainstorms. Philosophical Essays on Mind and Psychology*, Cambridge, MA: Bradford Books.

-(1987): True Believers in Dennett, Daniel (Ed.): *The Intentional Stance*, Cambridge, MA: The MIT Press.

-(2009): Intentional Systems Theory, in Beckermann, Ansgar, McLaughlin, Brian & Walter, Sven (Eds.): *The Oxford Handbook of Philosophy of Mind*, Oxford: Oxford University Press, 339-350. https://doi.org/ 10.1093/oxfordhb/9780199262618.003.0020

Dovidio, John, Hewstone, Miles, Glick, Peter, & Esses, Victoria (2010): Prejudice, stereotyping and discrimination: theoretical and empirical overview, in Dovidio, John, Hewstone, Miles, Glick, Peter & Esses, Veronica (Eds.): *The Sage Handbook of Prejudcie, Stereotyping and Discrimination*, Washington: SAGE Publications Ltd, 3-28. https://doi.org/10.4135/9781446200919,

Dworkin, Ronald (1975): Hard Cases. *Harvard Law Review*, *88(6)*, 1057–1109. https://doi.org/10.2307/1340249

Economic Times (2023): Apple's m-cap bigger than most countries' GDP. Access via: https://economictimes.indiatimes.com/markets/stocks/news/apples-m-cap-bigger-than-most-countries-gdp/articleshow/101557439.cms?from=mdr. Accessed: 07.03.2024

Elder-Vass, Dave (2010): *The Causal Power of Social Structures. Emergence, Structure and Agency*, Cambridge: Cambridge University Press.

Ende, Michael (1960): *Jim Knopf und Lukas, der Lokomotivführer.* Thienemann, Stuttgart.

European Unions Directorate-General for Communication (2024): Institutions and Bodies. Access via: https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies_en. Accessed: 01.02.2024.

Fassio, David & Logins, Artus (2023): Justification and gradability. *Philos Stud, 180*, 2051–2077. https://doi.org/10.1007/s11098-023-01945-3

Fiebich, Anika (2019): Social Cognition, Empathy and Agent-Specificities in Cooperation, *Topoi, 38*, 163–172.

Filstad, Catherine (2004): How newcomers use role models in organizational socialization, *Journal of Workplace Learning*, *16(7),* 396-409. https://doi.org/10.1108/13665620410558297

Fiske, Susan & Taylor, Shelley (2013): *Social Cognition: From Brains to Culture (3rd Edition)*, Thousand Oaks, CA: Sage Publication.

Frankfurt, Harry G. (1971): Freedom of the Will and the Concept of a Person, *The Journal of Philosophy*, *68(1)*, 5–20. https://doi.org/10.2307/2024717

-(1978): The Problem of Action. *American Philosophical Quarterly*, *15(2),* 157–162. http://www.jstor.org/stable/20009708

-(1988): *The Importance of What We Care About,* Cambridge: Cambridge University Press, 159-176.

French, Peter (1979): The Corporation as a Moral Person, *American Philosophical Quarterly*, *16(3)*, 207–215. http://www.jstor.org/stable/20009760

-(1995): *Corporate Ethics*. Orlando: Hartcourt Brace.

-(1996): Integrity, Intentions, and Corporations, *American Business Law Journal, 34*, 141-156. https://doi.org/10.1111/j.1744-1714.1996.tb00693.x

-(2015): Corporate Moral Agency, in Clubb, Colin & Imam, Shahed (Eds.): *Wiley Encyclopedia of Management,* Malden / Oxford: Wiley, 1-3. https://doi.org/10.1002/9781118785317.weom020062

-(2020): Types of Collectives and Responsibility, in: Bazargan-Forward, Saba & Tollefsen, Deborah P. (Eds.): *The Routledge Handbook of Collective Responsibility*, New York: Routledge, 9-23.

Fricker, Miranda (2007): *Epistemic Injustice: Power and the Ethics of Knowing,* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

García-Crespo, Ángel, Colomo-Palacios, Ricardo, Gómez-Berbís, Juan M. & Tovar-Caro, Edmundo (2008): The IT Crowd: Are We Stereotypes?, *IT Professional*, *10(6)*, 24-27. doi: 10.1109/MITP.2008.134.

Garcia-Godinez, Miguel (2020): What Are Institutional Groups?, in M. Garcia-Godinez, Miguel, Mellin, Rachael & Tuomela, Raimo (Eds.): *Social Ontology, Normativity and Law*, Berlin, Boston: De Gruyter, 39-62. https://doi.org/10.1515/9783110663617-004

- (2023): Review: Shared and Institutional Agency: Toward a Planning Theory of Human Practical Organization, *The Philosophical Quarterly, 73(3)*, 837–840. https://doi.org/10.1093/pq/pqad017

Giddens, Anthony (1984): *The Constitution of Society. Outline of the Theory of Structuration,* Cambridge: Polity Press.

Gigerenzer, Gerd, Todd, Peter M., et al. (1999): *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press.

Gilbert, Margaret (1987): Modeling Collective Belief, *Synthese*, *73(1)*, 185–204. http://www.jstor.org/stable/20116447

-(1990): Walking Together: A Paradigmatic Social Phenomenon, *Midwest Studies in Philosophy, 15,* 1–14. https://doi.org/10.1111/j.1475-4975.1990.tb00202.x

-(2002): Belief and Acceptance as Features of Groups, *Protosociology, 16*, 35–69.

-(2006): *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society.* Oxford: Oxford University Press.

-(2014): *Joint Commitment: How We Make the Social World*. Oxford: Oxford University Press.

Gilligan, Denis J. (1990): *Discretionary Powers. A Legal Study of Official Discretion*. Oxford: Clarendon Press

Ginet, Carl (1990): *On Action.* Cambridge: Cambridge University Press.

Goffman, Erving (1959): *The presentation of self in everyday life*. New York: Doubleday

-(1961): *Encounters: Two studies in the sociology of interaction.* Indianapolis: Bobbs-Merrill.

Goldman, Alvin I (1970): *A Theory of Human Action*. Englewood Cliffs: Prentice-Hall.

Goodwin, Charles (1994): Professional vision, *American Anthropologist, 96(3)*, 606–33.

Graeber, David (2018): *Bullshit Jobs: A Theory*. New York: Simon & Schuster.

Greif, Avner & Kingston, Christopher (2011): Institutions: Rules or Equilibria? in Schofield, Norman & Caballero, Gonzalo (Eds.): *Political Economy of Institutions, Democracy and Voting*, Berlin: Springer, 13-43.

Grewendorf, Günther & Meggle, Georg (Eds.) (2002): Speech Acts, Minds, and Social Reality. Discussions with John R. Searle, *Studies in Linguistics and Philosophy, 79*.

Guala, Francesco (2016): *Understanding Institutions, The Science and Philosophy of Living Together.* Princeton N.J.: Princeton University Press. https://doi.org/10.1515/9781400880911

Gunnemyr, Mattias (2017): Review: Kirk Ludwig: From Individual to Plural Agency: Collective Action, Volume I. *Ethic Theory Moral Prac, 20*, 915–918. https://doi.org/10.1007/s10677-017-9811-4

Hale, Bob (2017): Rule-Following, Objectivity, and Meaning, in Hale, Wright, and Miller (Eds.): *A Companion to the Philosophy of Language*, Malden / Oxford: Wiley, 619–648. https://doi.org/10.1002/9781118972090.ch24

Hammond, Paul (2016): Distinguishing joint actions from collective actions, *Synthese*, *193(9)*, 2707–2720. http://www.jstor.org/stable/24897930

Hart, Herbert L.A. (1994): *The Concept of Law, 2nd Ed.* New York/Oxford: Clarendon Press.

Hédoin, Cyril (2021): The Beliefs-Rules-Equilibrium Account of Institutions: A Contribution to a Naturalistic Social Ontology, *Journal of Social Ontology, 7(1)*, 73-96. https://doi.org/10.1515/jso-2020-0001

Hegel, Georg Wilhelm Friedrich (2018): *The Phenomenology of Spirit. Translated and edited by Terry Pinkard*, Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139050494

Hertwig, Ralph & Hoffrage, Ulrich (2012): *Simple Heuristics in a Social World.* Oxford: OUP.

Hess, Kendy M. (2020): Assembling the Elephant. Attending to the Metaphysics of Corporate Agents, in Bazargan-Forward, Saba & Tollefsen, Deborah P. (Eds.): *The Routledge Handbook of Collective Responsibility,* London / New York: Routledge, 113-126.

Higgins, Vaughan & Larner, Wendy (2010): From Standardization to Standardizing Work, in Higgins, Vaughn & Larner, Wendy (Eds.): *Calculating the Social.* London: Palgrave Macmillan. https://doi.org/10.1057/9780230289673_12.

Hindriks, Frank (2009): Constitutive Rules, Language, and Ontology, *Erkenntnis, 71(2)*, 253–75. http://www.jstor.org/stable/40267433.

-(2015): Review: Raimo Tuomela, Social Ontology: Collective Intentionality and Group Agents, *Economics and Philosophy, 31(2)*, 341-348. doi:10.1017/S0266267115000036

-(2017): Institutions and Collective Intentionality, in Jankovic, Marija & Ludwig, Kirk (Eds.): *The Routledge Handbook of Collective Intentionality*, New York: Routledge, 353-262.

-(2023): Ontological Holism Without Mental Holism: Bratman on Institutional Agency. *Journal of Social Ontology, 9(1)*. Access via: https://www.researchgate.net/publication/374584247_Ontological_Holism_Without_Mental_Holism_Bratman_on_Institutional_Agency. Accessed: 20.03.2024.

Hindriks, Frank & Guala, Francesco (2015): Institutions, Rules, and Equilibria: A Unified Theory, *Journal of Institutional Economics, 11(3),* 459–80.

Hobbes, Thomas (2000): *Leviathan*. *APA 7th Edition.* Hoboken, N.J.: Generic NL Freebook Publisher.

Hochschild, Arlie (1983): *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press.

Honneth, Axel (1994): *Kampf um Anerkennung. Zur moralischen Grammatik sozialer Konflikte.* Frankfurt a.M.: Suhrkamp.

-(2007): Recognition as ideology*,* in van den Brink, Bert & Owen, David (Eds.): *Recognition and power: Axel Honneth and the tradition of critical social theory*, Cambridge: Cambridge University Press, 323-347.

Honneth, Axel & Fraser, Nancy (2003): *Umverteilung oder Anerkennung? Eine politisch-philosophische Kontroverse.* Frankfurt a.M.: Suhrkamp.

Huebner, Bryce (2014): *Macrocognition. A Theory of Distributed Minds and Collective Intentionality.* Oxford: Oxford University Press.

Ikäheimo, Heikki & Laitinen, Arto (Eds.) (2011): Recognition and Social Ontology, *Social and Critical Theory, 11*.

Jacob, Pierre (2019): Intentionality, in Zalta, Edward N. (Ed.): *The Stanford Encyclopedia of Philosophy (Winter 2019 Edition)*, Access via: URL = <https://plato.stanford.edu/archives/win2019/entries/intentionality/>. Accessed: 25.03.2024.

Jankovic, Marija & Ludwig, Kirk (Eds.) (2017a): *The Routledge Handbook of Collective Intentionality.* New York / London: Routledge.

- (2017b): Introduction, in: Jankovic, Marija & Ludwig, Kirk (Eds.): *The Routledge Handbook of Collective Intentionality,* New York / London: Routledge, 1-5.

Jansen, Ludger (2016): *Gruppen und Institutionen. Eine Ontologie des Sozialen.* Wiesbaden: Springer.

Khanna, Parag (2016): These 25 Companies are more powerful than many Countries, *foreignpolicy.com*. Access via: https://foreignpolicy.com/2016/03/15/these-25-companies-are-more-powerful-than-many-countries-multinational-corporate-wealth-power/ Accessed: 22.03.2024.

Keeley, Michael (1981): Organizations as non-persons, *JValue Inquiry, 15*, 149–155. https://doi.org/10.1007/BF00147112

Kiernan, Denise (2013): *The Girls of Atomic City. The Untold Story of the Women Who Helped Win World War II.* New York/ London: Touchstone Book.

Koepsell, David & Moss, Laurence S. (Eds.) (2003): *John Searle's Ideas About Social Reality. Extensions, Criticisms, and Reconstructions,* Malden / Oxford: Wiley-Blackwell.

Kogon, Eugen (2006): *The Theory and Practice of Hell. The German Concentration Camps and the System Behind Them. With a New Introduction by Nikolaus Wachsmann*, New York: Macmillan Publishers.

Kubbe, Ina & Engelbert, Annika (Eds.) (2018): *Corruption and Norms. Why Informal Rules Matter.* London: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-66254-1

Kühl, Stefan (2007): Formalität, Informalität und Illegalität in der Organisationsberatung: Systemtheoretische Analyse eines Beratungsprozesses. *Soziale Welt, 58*, 269–91.

-(2013): *Organizations. A Systems Approach*. Farnham: Gower.

-(2014): *Ordinary Organizations. The Sociology of the Holocaust*, Berlin / Frankfurt a.M.: Suhrkamp.

-(2018): Rollen als Grundlagenthema im Coaching, in Greif Siegfried, Möller Heidi & Scholl Wolfgang (Eds.): *Handbuch Schlüsselkonzepte im Coaching*. *Springer Reference Psychologie*. Berlin: Springer, 495-502.

-(2022): *Useful Illegality. The Benefits of Breaking the Rules in Organizations.* Princeton/Hamburg: Organizational Dialogue Press.

Kutz, Christopher (2000): Acting Together, *Philosophy and Phenomenological Research*, *61(1)*, 1–31. https://doi.org/10.2307/2653401

Lackey, Jennifer (2021): *The Epistemology of Groups.* Oxford/New York: Oxford University Press. DOI:10.1093/oso/9780199656608.003.0003

Lampland, Martha & Star, Susan L. (2009): *Standards and their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life.* Ithaca: Cornell University Press.

Lavelle, Jane S. (2021): The impact of culture on mindreading, in Andrews, Kristin, Spaulding, Shannon & Westra, Evan (Eds.): *Folk Psychology: Pluralistic Approaches, Synthese, 199*, 6351–6374.

-(2022): *Mindreading and Social Cognition*, Cambridge, M.A.: Cambridge University Press. DOI:10.1017/9781108946766.

Lavin, Douglas (2013): Must There Be Basic Action?, *Noûs, 47(2)*, 273–301.

Ledeneva, Alena V (2018): *The Global Encyclopedia of Informality: Volume 1*. London: UCL Press.

Levin, Janet (2023): Functionalism, in Zalta, Edward N. & Nodelman, Uri (Eds.): *The Stanford Encyclopedia of Philosophy (Summer 2023 Edition)*. Access via URL = <https://plato.stanford.edu/archives/sum2023/entries/functionalism/>. Accessed: 25.03.2024.

Lewis, David Kellogg (1969): *Convention: A Philosophical Study.* Cambridge, M.A.: Harvard University Press.

Lieberstein, Paul (2007): *Money 1&2*, (S4,E7-8) [TV series episode], in Silverman, Ben & Daniels, Greg (Executive Producers): *The Office*, Universal Television.

Lieberthal, Mill (1976): TV and Movie Images of Workers, Reinforcing the Stereotypes, *Labour Studies Journal, 1, 162-169.*

List, Christian (2018): What is it Like to be a Group Agent?, *Noûs, 52*, 295-319. https://doi.org/10.1111/nous.12162

List, Christian & Pettit, Philip (2011): *Group Agency: The Possibility, Design, and Status of Corporate Agents.* Oxford: Oxford University Press.

Lizardo, Omar (2021): Culture, Cognition, and Internalization, *Sociological Forum, 36,* 1177-1206. https://doi.org/10.1111/socf.12771.

Lombard, Lawrence B. (1986): *Events: a Metaphysical Study.* London: Routledge.

Loon, Siu H. & McShane, Steven (2010): Structural and informal knowledge acquisition and dissemination in organizational learning: An exploratory analysis, *The Learning Organization*, *17(4)*, 364-386. https://doi.org/10.1108/09696471011043117

Ludwig, Kirk (2007): Collective Intentional Behavior from the Standpoint of Semantics, *Noûs, 41,* 355-393. https://doi.org/10.1111/j.1468-0068.2007.00652.x

-(2010): Adverbs of Action and Logical Form, in: O'Connor, Timothy & Sandis, Constantine (Eds.): *A Companion to the Philosophy of Action*, Malden/Oxford: Wiley-Blackwell, 40-49.

-(2014): Proxy Agency in Collective Action, *Noûs, 48(1)*, 75-105.

-(2015a): Is Distributed Cognition Group Level Cognition? *Journal of Social Ontology, 1(2)*, 189-224. https://doi.org/10.1515/jso-2015-0001

-(2015b): What are Conditional Intentions?, *Methode: Analytical Perspectives, 4(6),* 30–60.

-(2017a): *From Individual to Plural Agency: Collective Action I*. Oxford: Oxford University Press.

- -(2017b): *From Plural to Institutional Agency: Collective Action II*. Oxford: Oxford University Press.

-(2018a): Proxy Assertion*,* in Goldberg, Sanford (Ed.): *The Oxford Handbook of Assertion*, Oxford: Oxford University Press, 306-326. https://doi.org/10.1093/oxfordhb/9780190675233.013.13.

-(2018b): Proxy Agency, in: Jankovic, Marija & Ludwig, Kirk (Eds.): *The Routledge Handbook of Collective Intentionality*, New York: Routledge, 58–67.

-(2020a): The Social Construction of Legal Norms, in: Garcia-Godinez, Mellin & Tuomela (Eds.): *Social Ontology, Normativity and Law*, Berlin, Boston: De Gruyter, 179-208.

-(2020b): What Is Minimally Cooperative Behavior? in Fiebich, Anika (Ed.): *Studies in the Philosophy of Sociality 11*, Cham: Springer Nature, 9-39.

-(2020c): From Individual to Collective Responsibility: There and Back Again, in: Bazargan-Forward, Saba & Tollefsen, Deborah P. (Eds.): *The Routledge Handbook of Collective Responsibility*, London / New York: Routledge, 78-93.

Ludwig, Kirk & Jankovic, Marija (2022): Conventions and Status Functions, *The Journal of Philosophy*, *119(2),* 89-111.https://doi.org/10.5840/jphil202211926

Luhmann, Niklas (1964): *Funktionen und Folgen formaler Organisation,* Berlin: Duncker & Humblot.

- (1973): *Zweckbegriff und Systemrationalität*, Frankfurt a. M.: Suhrkamp.

Mäkelä, Pekka (2007): Collective Agents and Moral Responsibility, *Journal of Social Philosophy*, *38(3),* 456-468. https://doi.org/10.1111/j.1467-9833.2007.00391.x

Martens, Judith (2018): Exploring the Relation between the Sense of Other and the Sense of Us: Core Agency Cognition, Emergent Coordination, and the Sense of Agency. *Journal of Social Philosophy, 49(1), 38-60.* https://doi.org/10.1111/josp.12223

Marx, Karl (1867/1968): Das Kapital, Band. I, in: *Werke, Band 23*, Berlin: Dietz Verlag.

Marx, Karl & Engels, Friedrich (1848/1955): *The Communist Manifesto. With selections from The Eighteenth Brumaire of Louis Bonaparte and Capital, Edited by Samuel H. Beer.* New York: Appleton-Century-Crofts.

Mathiesen, Kay (2006): We're All in This Together: Responsibility of Collective Agents and Their Members, *Midwest Studies in Philosophy, 30*, 240-255. https://doi.org/10.1111/j.1475-4975.2006.00137.x

McCann, Hugh, J. (1983): Individuating Actions: The Fine-Grained Approach. *Canadian Journal of Philosophy*, *13(4)*, 493–512. http://www.jstor.org/stable/40231336

McLeod, Amber, Forgasz, Helen & Lang, Catherine (2013): IT stereotypes in television shows, in Deng, Hepu & Standing, Craig (Eds.): *Proceedings of the 24th Australasian Conference on Information Systems (ACIS), Melbourne, Australia, 4-6 December, 2013*, 1-10. Access via: https://www.researchgate.net/publication/267041676_IT_stereotypes_in_television_shows. Accessed: 12.03.2024.

Mead, George Herbert (1934): *Mind, self, and society. From the standpoint of a social behaviorist*. Chicago: University of Chicago Press.

Meijers, Anthonie (2003): Can Collective Intentionality Be Individualized?, *The American Journal of Economics and Sociology*, *62(1)*, 167–183. http://www.jstor.org/stable/3487966

Mele, Alfred R. (2003): *Motivation and Agency*, Oxford: Oxford University Press.

Merton, Robert K. (1957): *Social theory and social structure*. New York: Free Press.

Miller, Alexander & Sultanescu, Olivia (2022): Rule-Following and Intentionality, in: Zalta, Edward N. (Ed.): *The Stanford Encyclopedia of Philosophy (Summer 2022 Edition)*. Access via: URL = <https://plato.stanford.edu/archives/sum2022/entries/rule-following/>. Accessed: 25.03.2024.

Miller, Seumas (1998): *Authority, Discretion and Accountability. The Case of Policing*, in: Public sector ethics: finding and implementing values, Leichhardt: Federation Press, 37-53.

-(2001): *Social Action: A Teleological Account.* New York: Cambridge University Press.

-(2010): *The Moral Foundations of Social Institutions: A Philosophical Study.* Cambridge: Cambridge University Press.

-(2013): Police Ethics, *International Encyclopedia of Ethics,* 1-8 https://doi.org/10.1002/9781444367072.wbiee215

-(2019): *S*ocial Institutions, in: Zalta, Edward (Ed.): *The Stanford Encyclopedia of Philosophy (Summer 2019 Edition)*. Access via: URL = <https://plato.stanford.edu/archives/sum2019/entries/social-institutions/>. Accessed: 25.03.2024.

-(2021): Review: Kirk Ludwig, "From Plural to Institutional Agency: Collective Action II, *Philosophy in Review, 41(1), 37-39, https://doi.org/10.7202/1076215ar*

Miller, Seumas & Blackler, John (2005): *Ethical Issues in Policing.* Farnham/Burlington: Ashgate.

Miller, Seumas & Mäkelä, Pekka (2005): The Collectivist Approach to Collective Moral Responsibility, *Metaphilosophy,* 36(5), 623-651.

Moen, Lars J.K. (2023): Eliminating Group Agency, *Economics and Philosophy, 39(1)*, 43-66. doi:10.1017/S0266267121000341

Morgenroth, Thekla, Ryan, Michelle K. & Peters, Kim (2015): The Motivational Theory of Role Modeling: How Role Models Influence Role Aspirants' Goals*, Review of General Psychology, 19(4), 465-483.*

MSD Manual Consumer Version (2022): *Stupor and Coma*. Access via: URL = https://www.msdmanuals.com/home/quick-facts-brain,-spinal-cord,-and-nerve-disorders/coma-and-impaired-consciousness/stupor-and-coma. Accessed: 20.04.2023.

Nguyen, C. Thi (2019)*:* Games and the Art of Agency, *The Philosophical Review, 128(4),* 1-35. Access via*: https://philpapers.org/archive/NGUGAT-2.pdf Accessed:* 20.03.2024*.*

-(2020): *Games: Agency As Art.* Oxford: Oxford University Press.

Osrecki, Fran (2014): Autonomie von der Abweichung her denken: Zur Wiederentdeckung einer Theoriefigur, in Franzen, Martina et al. (Eds.): *Autonomie revisited: Beiträge zu einem umstrittenen Grundbegriff in Wissenschaft, Kunst und Politik*, Weinheim: Beltz Juventa, 400-423.

Ouyang, Guangwei & Shiner, Roger A. (1995): Organisations and agency, *Legal Theory, 1(3)*, 283–310. doi:10.1017/S1352325200000288

Oxford English Dictionary (2023)*: s.v. "MacGyver (v.),*" Access via: https://doi.org/10.1093/OED/7550886457. Accessed: 21.03.2024.

Pacherie, Elisabeth (2013): Intentional joint agency: shared intention lite, *Synthese*, *190(10)*, 1817–1839. http://www.jstor.org/stable/41931971

Paice, Elisabeth, Heard, Shelley & Moss, Fiona (2002): How important are role models in making good doctors?, *BMJ: British Medical Journal, 325,* 707–710. http://dx.doi.org/10.1136/bmj.325.7366.707

Paternotte, Cederik (2011): Being realistic about common knowledge: a Lewisian approach, *Synthese, 183*, 249–276. https://doi.org/10.1007/s11229-010-9770-y

Paul, Sarah (2020): *Philosophy of Action: A Contemporary Introduction.* London/New York: Routledge.

Perry, Beth R.N. (2009): Role modeling excellence in clinical nursing practice, *Nurse Educ Pract, 9(1),* 36-44. doi: 10.1016/j.nepr.2008.05.001

Perry, John (1994): Intentionality, in Guttenplan, Samuel (Ed.): *A Companion to the Philosophy of Mind. Blackwell Companions to Philosophy*, Malden/Oxford: Blackwell Publishing, 386-395.

Pettit, Philip (1996): *The Common Mind: An Essay on Psychology, Society, and Politics.* Oxford/ New York: Oxford University Press.

-(2003): Groups with Minds of Their Own, in Schmitt, Frederick F. (Ed.): *Socializing Metaphysics. The Nature of Social Reality*. Lanham: Rowman & Littlefield, 167-193.

-(2007): Rationality, Reasoning and Group Agency, *Dialectica*, *61*, 495-519. https://doi.org/10.1111/j.1746-8361.2007.01115.x

-(2018): Corporate Agency. The Lesson of the Discursive Dilemma, in: Jankovic, Marija & Ludwig, Kirk (Eds.): The Routledge Handbook of Collective Intentionality, New York: Routledge, 249-262.

Pettit, Philip & Schweikard, David (2006): Joint Actions and Group Agents, *Philosophy of the Social Sciences, 36(1),* 18–39. https://doi.org/10.1177/0048393105284169

Piñeros Glasscock, Juan S. and Tenenbaum, Sergio (2023): Action, in Zalta, Edward N. & Nodelman, Uri (Eds.): *The Stanford Encyclopedia of Philosophy (Spring 2023 Edition)*. Access via: URL = <https://plato.stanford.edu/archives/spr2023/entries/action/>. Accessed: 25.03.2024.

Poljanšek, Tom (2015): Choosing Appropriate Paradigmatic Examples for Understanding Collective Agency, in Misselhorn, Catherine (Ed.): *Collective Agency and Cooperation in Natural and Artificial Systems,* Philosophical Studies Series 122, 185-204. DOI 10.1007/978-3-319-15515-9_10

Preyer, Gerhard & Peter, Georg (Eds.) (2017): *Social Ontology and Collective Intentionality: Critical Essays on the Philosophy of Raimo Tuomela with his Responses. Studies in the Philosophy of Sociality 8*, Cham: Springer Nature.

Priest, Maura (2014): Review: Raimo Tuomela, Social Ontology: Collective Intentionality and Group Agents. *Ethics, 125(1),* 293-298,.

Quinton, Anthony (1975): Social Objects, *Proceedings of the Aristotelian Society, 76,* 1–27.

Rawls, John (1955): Two Concepts of Rules, *The Philosophical Review, 64(1),* 3–32. https://doi.org/10.2307/2182230

-(1999): *A Theory of Justice, Revised Edition.* Cambridge, M.A.: Harvard University Press.

Read, Leonard E. (1958): *I, Pencil: My Family Tree. As told to Leonard E. Read*, New York: Foundation for Economic Education, Inc. Access via: https://oll-resources.s3.us-east-2.amazonaws.com/oll3/store/titles/112/Read_0202_EBk_v6.0.pdf. Accessed: 17.01.2024.

Rhodes, Richard (1987): *The Making of the Atomic Bomb*. New York: Simon & Schuster.

Ritchie, Katherine (2013): What are Groups?, *Philosophical Studies, 166(2)*, 257–272.

-(2015): The Metaphysics of Social Groups, *Philosophy Compass*, *10*, 310– 321. doi: 10.1111/phc3.12213.

-(2016): Review: Groups as Agents, *Journal of Social Ontology*, *2(1)*, 173-175. https://doi.org/10.1515/jso-2015-0031

-(2020a): Minimal Cooperation and Group Roles, in Fiebich, Anika (Ed.): *Minimal Cooperation and Shared Agency, Studies in the Philosophy of Sociality 11*, Cham: Springer Nature, 93-109.

-(2020b): Social Structures and the Ontology of Social Groups, *Philos Phenomenol Res, 100,* 402-424. https://doi.org/10.1111/phpr.12555

Ritzer, George (1983): The "McDonaldization" of Society, *Journal of American Culture, 6(1),* 100-107. https://doi.org/10.1111/j.1542-734X.1983.0601_100.x

Romdenh-Romluc, Komarine (2017): Hermeneutical Injustice and the Problem of Authority, *Feminist Philosophy Quarterly, 3(3),* 1-22. doi:10.5206/fpq/2017.3.1.

Rönnegard, David (2013): How Autonomy Alone Debunks Corporate Moral Agency, *Business & Professional Ethics Journal*, *32(1),* 77-106.

-(2015): *The Fallacy of Corporate Moral Agency,* Dordrecht: Springer. https://doi.org/10.1007/978-94-017-9756-6

Ross, Tara (2019): Media and Stereotypes*,* in Ratuva, Steven (Ed.): *The Palgrave Handbook of Ethnicity,* London: Palgrave Macmillan. https://doi.org/10.1007/978-981-13-0242-8_26-1

Roth, Abraham S. (2017): Shared Agency, in Zalta, Edward N. (Ed.): *The Stanford Encyclopedia of Philosophy (Summer 2017 Edition).* Access via: URL = <https://plato.stanford.edu/archives/sum2017/entries/shared-agency/>. Accessed: 25.03.2024.

Rovane, Carol (1997): *The Bounds of Agency: An Essay in Revisionary Metaphysics.* Princeton, N.J.: Princeton University Press.

Ryle, Gilbert (1949): *The Concept of Mind.* Harmondsworth: Penguin.

Schmid, Hans Bernhard (2014): Plural self-awareness, *Phenomenology and the Cognitive Sciences, 13(1),* 7-24.

-(2023): *We, Together: The Social Ontology of Us.* Oxford: Oxford University Press.

Schmid, Hans Bernhard & Schweikard, David P. (2009): Einleitung: Kollektive Intentionalität. Begriff, Geschichte, Probleme, in Schmid, Hans B. & Schweikard, David P. (Eds.): *Kollektive Intentionalität. Eine Debatte über die Grundlagen des Sozialen,* Frankfurt a.M.: Suhrkamp, 11-65.

Schmidt am Busch, Hans-Christoph & Zurn, Christopher (Eds.) (2010): *The Philosophy of Recognition: Historical and Contemporary Perspectives*, Lanham: Lexington Books.

Schmitt, Frederick F. (2003): Joint Action: From Individualism to Supraindividualism, in Schmitt, Frederick F. (Ed.): *Socializing Metaphysics. The Nature of Social Reality*, Lanham: Rowman & Littlefield, 129-165.

-(2017): Collective Belief and Acceptance, in: Jankovic, Marija & Ludwig, Kirk (Eds.): *The Routledge Handbook of Collective Intentionality,* New York: Routledge, 90-103.

Schmitz, Michael (2017): What is a Mode Account of Collective Intentionality? In Preyer, Gerhard & Peter, Georg (Eds.): *Social Ontology and Collective Intentionality. Studies in the Philosophy of Sociality*, 8, Cham: Springer, 37-70. https://doi.org/10.1007/978-3-319-33236-9_3

-(2023): From we-mode to role-mode, in Garcia-Godinez, Miguel & Mellin, Rachael (Eds.): *Tuomela on Sociality*, London/ Cham: Palgrave Macmillan, 177-200.

Schweikard, David P. & Hans Bernhard Schmid (2021): Collective Intentionality, in Zalta, Edward N (Ed.): *The Stanford Encyclopedia of Philosophy (Fall 2021 Edition).* Access via: URL = <https://plato.stanford.edu/archives/fall2021/entries/collective-intentionality/>. Accessed: 25.03.2024.

Scott, James C. (1998): *Seeing like a State. How Certain Schemes to Improve the Human Condition Have Failed.* New Haven/London: Yale University Press.

Searle, John R. (1964): How to derive "ought" from "is", *The Philosophical Review, 73,* 43-58.

-(1969): *Speech acts: An essay in the philosophy of language*. Cambridge, M.A.: Cambridge University Press.

-(1978): Literal Meaning, *Erkenntnis, 13(1),* 207–224. http://www.jstor.org/stable/20010627

-(1983): *Intentionality. An Essay in the Philosophy of Mind.* Cambridge, M.A.: Cambridge University Press.

-(1990): Collective Intentions and Actions, in Cohen, Philip R., Morgan, Jerry & Pollack, Martha (Eds.): *Intentions in Communication*, Cambridge, M.A.: MIT Press, 401-415.

-(1992): *The Rediscovery of the Mind*, Cambridge, M.A.: MIT Press/Bradford Books.

-(1994): Intentionality, in Guttenplan, Samuel (Ed.): *A Companion to the Philosophy of Mind. Blackwell Companions to Philosophy*, Malden/Oxford: Blackwell Publishing, 379—386.

-(1995): *The Construction of Social Reality.* New York: The Free Press.

-(1998): *Mind, Language and Society.* New York: Basic Books.

- (2005): What Is an Institution?, *Journal of Institutional Economics, 1(1)*, 1–22. doi:10.1017/S1744137405000020

-(2010): *Making the Social World: The Structure of Human Civilization.* Oxford: Oxford University Press.

Shapiro, Scott J. (2014): Massively Shared Agency, in Vargas, Manuel and Yaffe, Gideon (Eds.): *Rational and Social Agency: Essays on the Philosophy of Michael Bratman*, New York: Oxford University Press, Yale Law School, Public Law Research Paper No. 581. Access via: https://ssrn.com/abstract=2839482. Accessed: 21.03.2024.

Skerker, Michael (2020): Individual Responsibility for Collective Actions, in: Bazargan-Forward, Saba & Tollefsen, Deborah P. (Eds.): *The Routledge Handbook of Collective Responsibility*, New York: Routledge, 274-284.

Smith, Thomas H. (2012): Review: Group Agency: The Possibility, Design, and Status of Corporate Agents, by Christian List and Philip Pettit, *Mind*, *121(482),* 501–507, https://doi.org/10.1093/mind/fzs073

Sofsky, Wolfgang (1996): *The Order of Terror. The Concentration Camp.* Princeton, N.J.: Princeton University Press.

Spaulding, Shannon (2018): *How We Understand Others: Philosophy and Social Cognition.* London/New York: Routledge.

Spears, Russel., Doosje, Bertjan., & Ellemers, Naomi (1997): Self-Stereotyping in the Face of Threats to Group Status and Distinctiveness: The Role of Group Identification, *Personality and Social Psychology Bulletin*, *23(5)*, 538-553. https://doi.org/10.1177/0146167297235009

Sperber, Dan & Wilson, Dierde (1986): *Relevance.* Cambridge: Harvard University Press.

Stohl, Cynthia, & Redding, Charles W. (1987): Messages and message exchange processes*,* in Jablin, Frederic L. et al. (Eds.): *Handbook of organizational communication: An interdisciplinary perspective*, London: Sage Publications, 451–502.

Stoll, Julia (2023): *Average daily time spent watching TV per capita in the United States from 2009 to 2022, by age group,* statista.com*. Access via:* https://www.statista.com/statistics/411775/average-daily-time-watching-tv-us-by-age/#statisticContainer*.* Accessed: 07.12.2023.

Strohmaier, David (2020): Two theories of group agency, *Philos Stud, 177*, 1901–1918. https://doi.org/10.1007/s11098-019-01290-4

Styhre, Alexander (2010): Knowledge work and practices of seeing: Epistemologies of the eye, gaze, and professional vision, *Culture and Organization, 16(4)*, 361-376. DOI: 10.1080/14759551.2010.519931

Sugden, R. (2005): *The Economics of Rights, Cooperation and Welfare*, *2nd ed*. London / New York: Palgrave Macmillan.

Sylvian, Kurt (2012): How to be a redundant realist, *Episteme, 9(3),* 271-282. doi:10.1017/epi.2012.16.

Tanney, Julia (2021): Some Problems in Contemporary Work on Knowing-How and Knowing-That, in Zalta, Edward N. (Ed.): *The Stanford Encyclopedia of Philosophy (Summer 2022 Edition).* Access via: URL = https://plato.stanford.edu/entries/ryle/knowing-how.html. Accessed: 25.03.2024.

Taylor, Frederick W. (1911): *The Principles of Scientific Management.* New York/London: Harper & Brothers. Access via: https://archive.org/details/principlesofscie00taylrich Accessed: 14.11.2023.

The Guardian (2013): *A history of Nokia's from paper mills to Gorbachev.* Access via: https://www.theguardian.com/technology/2013/apr/01/history-nokia Accessed: 07.03.2024.

Theiner, Georg (2018): Groups as Distributed Cognitive Systems, in: Jankovic, Marija & Ludwig, Kirk (Eds.): *The Routledge Handbook of Collective Intentionality*, New York: Routledge, 233-248.

Tollefsen, Deborah P. (2002a): Collective Intentionality and the Social Sciences, *Philosophy of the Social Sciences, 32(1),* 25–50. https://doi.org/10.1177/004839310203200102

-(2002b): Organizations as True Believers, *Journal of Social Philosophy, 33,* 395-410. https://doi.org/10.1111/0047-2786.00149

-(2015): *Groups as Agents.* Cambridge/Malden: Polity Press.

Townsend, Leo (2013): Being and Becoming in the Theory of Group Agency, *Abstracta, 7(1),* 39–53. https://doi.org/10.24338/abs-2013.231

-(2015): Review: Raimo Tuomela, Social Ontology: Collective Intentionality and Group Agents, *Journal of Social Ontology*, *1(1),* 183-187. https://doi.org/10.1515/jso-2014-0040

-(2020): Groups with Minds of Their Own Making, *Journal of Social Philosophy*, *51*, 129-151. https://doi.org/10.1111/josp.12295

Tuomela, Raimo (1992): Group Beliefs, *Synthese*, *91(3)*, 285–318. http://www.jstor.org/stable/20117028

-(2002): *The Philosophy of Social Practices: A Collective Acceptance View.* Cambridge: Cambridge University Press.

-(2003): Collective Acceptance, Social Institutions, and Social Reality, *The American Journal of Economics and Sociology, 62(1)*, 123-165.

-(2007): *The Philosophy of Sociality: The Shared Point of View*. Oxford/ New York: Oxford University Press.

-(2011): An Account of Group Knowledge, in Schmid, Hans B., Sirtes, Daniel & Weber, Marcel (Eds.): *Collective Epistemology*, Berlin/ Boston: De Gruyter, 75-118. https://doi.org/10.1515/9783110322583

-(2013): *Social Ontology. Collective Intentionality and Group Agents.* Oxford/ New York: Oxford University Press.

-(2017): Review: From Individual to Plural Agency: Collective Action: Volume 1, *Notre Dame Philosophical Reviews.* Access via: https://ndpr.nd.edu/reviews/from-individual-to-plural-agency-collective-action-volume-1/. Accessed: 25.03.2024

Tuomela, Raimo & Miller, Kaarlo (1988): We-intentions, *Philosophical Studies, 53*, 367–389. https://doi.org/10.1007/BF00353512

Turner, John C. (1985): Social Categorization and the Self-Concept: A Social Cognitive Theory of Group Behavior, in Lawler, Edward J. (Ed.): *Advances in Group Processes: Theory and Research*. Greenwich, CT: JAI Press.

Turner, Ralph H. (2001): Role theory, in Turner John H. (Ed.): *Handbook of sociological theory*, New York: Springer, 233–254.

Tsohatzidis, Savas L. (Ed.) (2007): *John Searle's Philosophy of Language: Force, Meaning and Mind.* Cambridge: Cambridge University Press.

Ungan, Mustafa C. (2006): Standardization through process documentation, *Business Process Management Journal*, *12(2),* 135-148. https://doi.org/10.1108/14637150610657495

United States Department of Labor (2023): *Bureau of Labor Statistics: Occupational Outlook Handbook, Firefighters*. Access via: https://www.bls.gov/ooh/protective-service/firefighters.htm. Accessed: 05.06.2023.

United States Department of War (1941): *Basic Field Manual. Solider's Handbook*. Washington D.C.: United States Government Printing Office. Access via: https://www.ibiblio.org/hyperwar/USA/ref/FM/PDFs/FM21-100.pdf. Accessed: 21.05.2024.

United States Office of Strategic Services [OSS] (1944): *Simple Sabotage Field Manual*, Washington. Access via: https://www.cia.gov/static/5c875f3ec660e092cf893f60b4a288df/SimpleSabotage.pdf. Accessed: 27.03.2024.

Vanderschraaf, Peter & Sillari, Giacomo (2022): Common Knowledge, in Zalta, Edward N. & Nodelman, Uri (Eds.): *The Stanford Encyclopedia of Philosophy (Fall 2022 Edition)*. Access via: URL = <https://plato.stanford.edu/archives/fall2022/entries/common-knowledge/>. Accessed 25.03.2024.

van Dijk, Rozemarijn E. (2023): Playing by the rules? The formal and informal rules of candidate selection, *Women's Studies International Forum, 96*. https://doi.org/10.1016/j.wsif.2022.102669.

van Riel, Raphael & van Gulick, Robert (2024): Scientific Reduction, in Zalta, Edward N. & Nodelman, Uri (Eds.): *The Stanford Encyclopedia of Philosophy (Spring 2024 Edition).* Access via: URL = <https://plato.stanford.edu/archives/spr2024/entries/scientific-reduction/>. Accessed: 15.05.2024.

Velasquez, Manuel G. (1983): Why Corporations Are Not Morally Responsible for Anything They Do, *Business & Professional Ethics Journal*, *2(3),* 1–18. http://www.jstor.org/stable/27799793

von Hayek, Friedrich A. (1945): The Use of Knowledge in Society, *The American Economic Review*, *35*(4), 519–530. http://www.jstor.org/stable/1809376

von Kleist, Heinrich (1986): *Prinz Friedrich von Homburg.* Stuttgart: Reclam.

Wallace, Kathleen, A. (1999): Anonymity. *Ethics and Information Technology, 1*, 21–31. https://doi.org/10.1023/A:1010066509278

Walt Disney Company (2014): *The Disney Look.* Access via: https://cepfranco.files.wordpress.com/2016/04/disney-look-book.pdf. Accessed: 13.12.2023

-(2020): *Fact Sheet*. Access via: https://web.archive.org/web/20200220211438/https://dpep.disney.com/wp-content/uploads/2020/02/fact_sheet_walt_disney_world_resort_2020_Q1.pdf. Accessed: 13.12.2023.

-(2022): *Fiscal Year 2022 Annual Financial Report*. Access via: https://thewaltdisneycompany.com/app/uploads/2023/02/2022-Annual-Report.pdf. Accessed: 13.12.2023.

Warren, John T. (2001): Doing whiteness: On the performative dimensions of race in the classroom, *Communication Education, 50(2)*, 91-108, DOI: 10.1080/03634520109379237

Westmarland, Louise (2008): Police cultures, in Newburn, Tim (Ed.): *Handbook of Policing (2nd ed.).* Cullompton, Devon: Willan Publishing, 253–280.

Westra, Evan (2019): Stereotypes, theory of mind, and the action-prediction hierarchy, *Synthese, 196(7),* 2821-2846.

Wittgenstein, Ludwig (2001): *Philosophische Untersuchungen*. *Kritisch-genetische Edition. Herausgegeben von Joachim Schulte.* Frankfurt a.M: Wissenschaftliche Buchgesellschaft.

Zahle, Julie (2016): Methodological Holism in the Social Sciences, in Zalta, Edward N. (Ed.): *The Stanford Encyclopedia of Philosophy (Winter 2021 Edition)*. Access via: URL = <https://plato.stanford.edu/archives/win2021/entries/holism-social/>. Accessed: 25.03.2024.

Zawidzki, Tadeusz W. (2013): *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, M.A.: MIT Press.

-(2016): Mindshaping and self-interpretation, in Kiverstein, Julian (Ed.): *The Routledge Handbook of Philosophy of the Social Mind.* New York / London: Routledge, 479-497. https://doi.org/10.4324/9781315530178.

List of Figures, Tables, and Boxes

Index

of Subjects & Names

Technical terms of the author's are in **boldface type**, as are pages that define/introduce the term