Technische Universität Dortmund

Fakultät Erziehungswissenschaft, Psychologie und Bildungsforschung

Understanding Infants' Social-Cognitive Development in the Second Year of Life: The Relation between Imitation, Temperament and Visual Attention

Kumulative Dissertation

zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil)

vorgelegt von

Franziska Sieber

geboren am 03.10.1994 in Dortmund

Erstgutachter: Prof. Dr. Norbert Zmyj

Zweitgutachterin: Prof. Dr. Sabine Seehagen

Februar 2025

Table of Contents

Abstract 1
1 General Introduction 4
1.1 Imitation 4
1.1.1 Definitions of Imitations 4
1.1.1 Theories on the Development of Imitation 7
1.1.2 Functions of Imitation 10
1.2 Temperament 14
1.2.1 Theories on Infant Temperament 14
1.2.2 Measuring Temperament 19
1.3 Imitation and Temperament 22
1.4 Imitation and Attention 26
2 Outline of the Thesis 30
3 Article 1 – Sieber, F., & Zmyj, N. (2022). Stability and structure of infant and toddler temperament in two longitudinal studies in Germany. *Infant Behavior and Development, 67*, 101714 32
4 Article 2 – Sieber, F., Czarnomski, J., Schölmerich, A., Daum, M. M., & Zmyj, N. (2024). The two functions of imitation in the second year of life: A longitudinal study. *Developmental Psychology.* Advance Online Publication 79
5 Article 3 – Sieber, F., Czarnomski, J., Daum, M.M., & Zmyj, N. (2024). The relation between goal-predictive gaze behavior and imitation – A live eye-tracking study in 12-months-olds. *PsyArXiv Preprints* 142
6 General Discussion 166
6.1 Relevance 176
6.2 Limitations and Future Research 179
7 Conclusion 182
References 183
List of Tables 205
List of Figures 208
List of Articles and Own Contribution to the Publications 209
Eidesstattliche Versicherung 210

Abstract

Imitation plays a central role in early human development, emerging within the first year of life and developing rapidly. In order to understand the development of imitative behavior and its variability, it is important not only to consider imitation as an intellectual performance, but also to consider infants' motivation to engage in imitative behavior. Imitation serves both a cognitive function, driven by infants’ instrumental motivation to acquire new skills, and a social function, driven by infants’ social motivation to interact with the model. Individual differences in temperament need to be taken into account to determine infants' different motivations for imitation. It is further suggested that within the second year of life there is a shift between the two functions, from a dominance of the cognitive function at the beginning of the second year of life to a dominance of the social function at the end of the second year of life (Užgiris, 1981). The current thesis aims to systematically investigate the relation between imitation and temperament longitudinally within the second year of life in order to determine whether this shift in the motivation to imitate occurs. Given the important role of attention in the imitation process (Bandura, 1977), a further aim is to investigate infants' gaze direction during the action demonstration. A systematic longitudinal study was therefore designed, using standardized tests of imitation, temperament, and cognitive development in 12-, 18- and 24-month-olds, resulting in three publications.

The first publication of this thesis examines the factor structure and stability of temperament in two longitudinal samples, based on the data from the longitudinal study and on a further dataset of 9- and 12-month-olds. Given the inconclusive evidence for the three-factor structure of temperament as measured by the Infant Behavior Questionnaire-Revised (IBQ-R), particularly in Germany, the factor structure needed to be validated in order to use it as scale and factor level measure. The results revealed a two-factor solution, replicating previous findings with German infants. It is therefore recommended that IBQ-R data are only analyzed at scale level until further validation of the factor structure has been achieved.

The second publication of this thesis examines whether there is a change in the two functions of imitation during the second year of life. Therefore, infants' imitative behavior and social and attentional aspects of their temperament were related at 12, 18 and 24 months of age, while controlling for infants' cognitive developmental status. The longitudinal data revealed no consistent relation between imitation and attentional aspects of temperament at all time points, irrespective of their cognitive developmental status, whereas imitation and social aspects of temperament were related from 12 months onwards. These findings suggest that the shift between the two functions of imitation does not occur in the second year of life, but that imitation serves a social function from early on.

The third publication of this thesis investigates the role of attention in the imitation process. Using a novel methodological approach, 12-months-olds' gaze direction during action demonstration was investigated in a live eye-tracking setting. Infants who imitated the action showed faster goal-predictive gaze shifts than those who did not. The relation between imitation and goal prediction disappeared, when infants' cognitive developmental status was controlled for. This may be due to methodological issues in the first application of this novel approach and does not necessarily indicate that there is no relation between imitation and goal prediction. However, this publication provides a novel approach to investigating infants' goal prediction in a more naturalistic context that resembles real-life scenarios.

Taken together, the findings of the current thesis show that the underlying motivations for imitation do not change in the second year of life, as the social function is present from an early age. The thesis provides a novel approach for investigating infants' gaze behavior during imitation tasks, and thus the role of attention in imitation, a relation that needs to be further explored. The implications of the findings for understanding variability in imitation, the strengths and limitations of the study, and future research directions are discussed in detail.

1 General Introduction

1.1 Imitation

Human's ability to imitate the actions of others plays a crucial role in human cognitive and social development. Emerging during infancy, it enables the transmission of cultural knowledge and practices (Dean et al., 2012; Tomasello, 1999; Whiten, 2005). Imitation is suggested as a precursor to essential social-cognitive skills, including the acquisition of language (Charman et al., 2000; Papoušek & Papoušek, 1989) or understanding others' intentions (Meltzoff, 1995). It also facilitates social interactions and relations (Užgiris, 1981). Studies of atypical populations further emphasize the importance of imitation for social-cognitive development. If imitation is impaired, for example in children with autism, development is seriously affected (Rogers et al., 2003; Ingersoll, 2008).

Over the past few decades, imitation has become a focal point of interest for researchers across various fields. This gave rise to a wealth of theoretical accounts, empirical research and commentaries on imitation from various disciplines within psychology (e.g., developmental, cognitive, comparative) as well as fields beyond it (e.g., philosophy, animal behavior, robotics, neuroscience). The concept and terminology of imitation vary widely both across disciplines and within psychology. Consequently, numerous different labels and definitions of imitation coexist, presenting any study of imitation with the initial challenge of defining the phenomenon.

1.1.1 Definitions of Imitations

Defining imitation proves to be more complex than it firstly seems. Initially, everyone can quickly relate to the term (Carpenter & Call, 2009) associating it, at the very least, with the reproduction of actions following an observation. At the same time, different people associate different behaviors with imitation. Some define it as the copying of actions (e.g., Nielsen, 2009), others emphasize its role in learning new behaviors (e.g., Elsner, 2014) and some understand it as a process of reading intentions through the replication of means and goals (e.g., Tomasello et al., 2005). As a result, definitions of imitation abound in the literature, and researchers have yet to agree on a universally accepted definition.

One of the earliest broad definitions of imitation was proposed by Thorndike (1898) as learning to do an act by witnessing it. This basic assumption - that imitation involves reproducing behaviors previously demonstrated by others - formed the basis for a majority of subsequent research (e.g., Barr et al., 1996; Nielsen, 2009; Over, 2020).

Over time, this definition has been expanded. One expansion was provided by Byrne and Russon (1998). They included the *novelty* of the copied behavior and defined imitation as the acquisition of a novel behavior after observing a demonstrator performing it. According to Byrne and Russon (1998), novelty is essential to the term imitation. If the behavior is already present in the repertoire, other accounts than imitation might also be plausible to explain the reproduction of the modeled behavior (e.g., stimulus enhancement, i.e., social learning about an object through drawing attention to the object of an action without paying attention to what the demonstrator does with it, Spence, 1937). Meltzoff (1988) proposed six criteria to assess whether an action qualifies as novel: 1) it has not been perceived before, 2) it has never been performed before, 3) although it may have been performed, it has not been part of infants' daily behavioral repertoire, 4) it has never been imitated before, 5) it has not been related to a specific object and 6) it is not spontaneously exhibited during free play. Thus, imitation was not only seen as copying a behavior shown by others, but the copied behavior had to be novel in terms of the outlined criteria (Byrne & Russon, 1998). Since criteria 1-5 are mostly unknown to the experimenter, Meltzoff (1998) emphasized the importance of criterion 6 regarding the study of imitation. Emphasizing the relevance of novelty provides an important tool for designing research on imitation: establishing a baseline phase to ensure that infants do not spontaneously show the target behavior.

Another significant expansion to the definition of imitation was introduced by Tomasello (1999) by including the model's *intention*. Following his considerations, imitative behaviors involve infants adopting the specific means employed by the model to achieve a goal. Thus, imitation displays infants' understanding of the model's intention (Tomasello, 1999). Extending both definitions presented above, imitation entails copying a model's novel behavior including the same means, intentions and goals, highlighting the role of means and goals in imitation (e.g., Gergely et al., 2002; Tomasello et al., 2005; Want & Harris, 2002). Based on Tomasello's (1999) definition, related constructs can be derived in an attempt to provide a broader framework for studies of social learning and to better distinguish between the different learning mechanisms. *Emulation* includes reproducing the model's goal using different means than the model (Whiten & Ham, 1992). *Mimicry* contains copying the model's means without the goal, actually without any insights into what goal they serve (Tomasello et al., 1993). These constructs differ from imitation as they do not incorporate the model's intention.

Considering the definitions discussed above, most can be categorized into two main approaches: Broader, behavior-based definitions conceive imitative behavior as copying any observed behavior (e.g., Barr et al., 1996; Paulus, 2011), whereas narrow, intention-based definitions specify imitative behavior as copying the means of an action in order to achieve a specific goal (e.g., Carpenter & Call, 2009; Tomasello, 1999). According to Paulus (2011), a problem of intention-based definitions is that these approaches presuppose a mechanism (i.e., understanding the model's intention) which they actually want to investigate empirically. Therefore, he suggests to restrict the term imitation to the behavioral phenomenon of the similarity of two behaviors without specifying mechanisms that subserve imitation. Following Paulus' (2011) suggestions and in line with current research (e.g., Elsner, 2014; Nielsen, 2009; Over, 2020), a broader definition of imitation as reproduction of the demonstrated behavior, more precisely, any novel action presented by a model, is used for the present dissertation. To ensure the novelty of the actions, a baseline was implemented into the imitation tasks.

1.1.1 Theories on the Development of Imitation

Imitative abilities are present within the first year of life and develop rapidly during infancy (Barr et al., 1996). Over recent decades, numerous theories have been proposed to explain how these abilities develop in infants. To date, it is still discussed whether imitative capacities are present at birth and therefore inborn, or if they emerge later in infancy (Heimann, 2022a). Meltzoff and Moore (1977), picking up early considerations (McDougall, 1908), questioned the theories of imitation as learned behavior, which were widespread at that time (especially Piaget, 1962). In their landmark study on neonatal imitation, they reported that newborns as young as 12-21 days reproduced oral gestures of an adult model. Despite over 40 years of research, the only consistent results on neonatal imitation are those based on tongue protrusion (Anisfeld, 2005; Oostenbroek et al., 2016). As tongue protruding is a common response of newborns to interesting or arousing stimuli, their matching of tongue protruding is more likely an expression of their interest than imitative behavior (Jones, 2009). According to Jones (2009), it is possible that new findings will support the claim of neonatal imitation in the future. However, given the current inconclusive evidence, it is more prudent to assume that imitative abilities are not present at birth but instead develop later in infancy.

Piaget's (1962) cognitive-developmental theory remains one of the most influential theories for understanding the development of imitation. He suggested that infants acquire their imitative abilities in six consecutive stages in the first two years of life. At Stage 1, imitative skills are not present. Newborn's abilities are limited as they only show inborn reflexes. In Stage 2, infants start to reproduce their own acts and therefore engage sporadically in imitation of other's actions. Imitative abilities are still limited to simple gestures already present in their motor repertoire. Reaching Stage 3, infants' start to engage in systematic imitation as they are able to imitate observable actions and effects from two different domains (e.g., clapping: touch and listening), if they are in their own motor repertoire. In Stage 4, these abilities are extended to invisible behaviors, i.e., infants copy movements in body parts that they only observe in others but not in themselves (e.g., the face). Reaching Stage 5 enables infants to imitate unfamiliar actions and behaviors. Finally, in Stage 6, infants are able to imitate based on mental representations. Building a representation from their actual perception allows infants to store the action in memory and reproduce it with a delay. According to Piaget (1962), this ability to defer imitation of observed behavior emerges at the end of the second year of life, not before 18 months.

Tomasello and colleagues (2005) proposed a different perspective on the development of imitation. According to their view, understanding the intentional structure of actions enables imitative behavior. This understanding develops gradually from birth until 14 months. In the first nine months of life, infants do not display any understanding of goals or intention. From nine months onward, infants develop an understanding of goal-directed actions. They begin to understand that another person acts according to their internal goal. Around 14 months, understanding of rational intentions emerges. Infants understand that another person considers different action plans to achieve their goal, chooses one of them and implements it in order to achieve the goal. From 14 months on, the understanding of intentions develops further. Thus, imitation depends crucially on the infants' social-cognitive ability to infer the intention of others (Tomasello & Carpenter, 2005; Carpenter, 2006).

Based on these two theoretical accounts, imitation tasks are typically used in two different contexts (see also Damm et al., 2011; Fagard & Lockman, 2010). Following Piaget (1962), one line of research uses imitation as a method to test the capacity of infants' declarative memory (e.g., Barr et al., 1996; Goertz et al., 2008; Hayne & Herbert, 2004; Seehagen et al., 2015). In these deferred imitation paradigms, a model demonstrates novel actions to the infant. The time interval between the demonstration of the action and its reproduction is systematically varied. It is assumed that infants need to encode and store the demonstrated action in their long-term memory in order to be able to reproduce it after a delay. If infants do not imitate the action, it is usually assumed that they do not have the cognitive abilities to perform the task. Infants are able to solve these tasks within the first year of life. Contrary to Piaget's (1962) assumption that deferred imitation develops between 18 and 24 months, Barr et al. (1996) showed that even six-month-olds imitated after a 24-hour delay when given enough time to interact with the object. Collie and Hayne (1999) replicated this finding with six- and nine-month-olds using different objects. In the course of the second year of life, infants remember longer sequences of actions over extended periods of time. Fourteen-month-olds reproduced two-step actions after a delay of one week (Meltzoff, 1988). Hayne and Herbert (2004) showed that 18-month-olds imitated three-step actions after a delay of four weeks.

Following Tomasello et al.'s (2005) theoretical approach, another line of research utilizes immediate imitation tasks to assess infants' social-cognitive understanding, i.e., how they interpret the goals and intentions of others and reproduce actions accordingly. The ability to imitate is taken as given and infants are assumed to be able to extract information from an observed action. This information is manipulated in these tasks, for example by situational constraints (Gergely et al., 2002) or contextual cues such as the presence or absence of goals (Carpenter et al., 2005). If the infants' behavior does not differ between these conditions, it is assumed that they have not yet developed the respective social-cognitive understanding. Gergely et al. (2002) found that 14-month-olds consider the situational constraints of the model (a blanket that occupies the hands) to choose the most rational alterative for themselves when imitating an unusual action (illuminating a lamp with the head). This evidence of understanding others' attention was even found in 12-month-olds (Schwier et al., 2006) using a slightly different task, but not in 9-month-olds (Zmyj et al., 2009).

To sum up, the debate surrounding the development of imitation remains ongoing, with particular uncertainty about whether imitative abilities are present at birth (for a meta-analysis, see Davis et al., 2021). The diverse concepts of the term imitation explained from the different theoretical perspectives result in inconsistent evidence on the development of imitative skills in the first two years of life. Basically, depending on the definition, the onset of imitation occurs within the first year of life, between zero and 12 months and the ability expands within the second year of life. To understand the development of imitation, however, it is not enough to look only at the cognitive abilities of young children. Imitation not only allows children to learn new skills, but is also a deeply social process enabling children to initiate and maintain social interactions long before they develop verbal communication (Užgiris, 1981). In line with this, McCall and colleagues (1977) argued that cognitive factors determine what infants imitate, while motivational factors will determine whether the infants imitate these behaviors or actions. This raises the question of whether variations in infants' imitative behavior reflect purely intellectual capacities (e.g., capacity of declarative memory or social-cognitive understanding) or whether these behaviors are also shaped by the infants' motivation to engage in imitation.

1.1.2 Functions of Imitation

Most research on infants' motivation to imitate is based on Užgiris' (1981) foundational paper on the two functions of imitation. She proposed that imitation in infancy serves both a cognitive and a social function. The cognitive function is driven by infants' instrumental motivation to acquire new skills. In this context, the imitator learns actions to solve a given problem without having to learn by trial and error (Barr et al., 1996). Simultaneously, imitation fulfills a social function, driven by infants' social motivation to engage in interaction with the model. The imitator communicates non-verbally and conforms to the model by communicating mutuality. The relationship between the imitator and the model is of particular importance, while the specific actions being imitated are secondary. The two functions are not seen as competing, but rather as both being important for children's social-cognitive development.

Following Užgiris' (1981) considerations, the two functions of imitation have been the subject of theoretical approaches examining variations in infants' imitative behavior (Nielsen, 2009; Over & Carpenter, 2013), and empirical studies (e.g., Nielsen et al., 2008; Óturai et al., 2018). Empirical evidence for both functions is widespread. The preceding review of the cognitive function demonstrated how imitation enables infants to quickly acquire new skills. Barr and Hayne (2003), for instance, showed that 12-month-olds acquire one to two new behaviors per day through imitation. Findings on the social function are likewise well-established in the literature (e.g., Nielsen et al., 2008; Over & Carpenter, 2009; for an overview, see Over, 2020). The most common approach to studying social imitation has been to assess the fidelity with which children copy a series of actions demonstrated by a model (e.g., Brugger et al., 2007; Hilbrink et al., 2013; Nielsen, 2006; Nielsen & Blank, 2011; Yu & Kushnir, 2014). In these series of actions, the model demonstrates some actions that are unnecessary to achieve the goal. If infants reproduce these causally irrelevant actions (exact imitation), this is taken as evidence for their social motivation as they are creating mutuality with the model through imitating both causally relevant and irrelevant actions. In contrast, imitating only the causally relevant actions (selective imitation) is viewed as reflecting instrumental motivation, where the child is primarily focused on acquiring new skills.

Užgiris (1981) further proposed a change in infants' motivation to imitate within the second year of life. As their cognitive abilities develop, infants begin to recognize their fundamental similarity to others and develop an understanding of social interactions. Consequently, the respective importance of the two functions for children shifts over this period. Whereas the cognitive function is more important at the beginning of the second year of life and becomes less important over time, the social function becomes more important during the second year of life. Óturai et al. (2018) found that 24-month-olds were more likely to show exact imitation than 18-month-olds. This decrease in infants' selectivity of imitative behavior was interpreted as a shift in their underlying motivation to imitate: given that 24-month-olds predominantly focus on interacting with the model by creating mutuality and not on acquiring new skills, they imitate both causally relevant and irrelevant actions. Hilbrink and colleagues (2013) found a similar decline in selective imitation between 12 months and 15 months of age. In a series of experiments, Nielsen (2006) aimed at investigating possible age-related changes in imitative behavior within the second year of life. Therefore, he provided children with a box containing a desirable toy. In one condition, a model used an object to open the box, while in another, the model used their hand. He found that 12-month-olds predominantly imitated the outcome of an action (opening the box without using the object), whereas 18- and 24-month-olds imitated both, the action and the outcome. He expanded this experiment to further investigate how social cues (i.e., a model acting communicative and engaging vs. a model action aloof) influences 18- and 24-moths-olds' imitative behavior. Eighteen-month-olds opened the same number of boxes in both conditions, but were more likely to imitate the action of the model (i.e., using the object) in the social condition. Twenty-four-month-olds opened more boxes in the social condition than in the non-social condition. These findings align with Užgiris' hypothesis (1981) indicating that while 12-month-olds prioritize skill acquisition, older infants increasingly use imitation to establish social connections. Whereas 12-month-olds focused on learning a new skill (i.e., opening the box), 18- and 24-month-olds used the same means as the model to communicate mutuality. This was further supported by the finding that a socially acting model led to higher imitation rates in 18- and 24-month-olds. Taken together, these findings highlight that the social function gains more importance after the cognitive function in the second year of life.

However, the interpretation of Nielsen's (2006) results as evidence for the changes in infants' underlying motivation to imitate is limited. The experiments were conducted cross-sectionally. Further, in line with the vast majority of research on the onset of the social function, Nielsen (2006) investigated the social influence on imitation by varying the social context of the situation (e.g., communicative vs. noncommunicative cues, Kupán et al., 2017). Over and Carpenter (2013) highlight the problem that, despite Nielsen's (2006) study, little research examines the developmental path of the underlying motivations and legitimately ask "When do the different motivations outlined above emerge in development?" (p. 9). No study to date has systematically examined the development of these motivations throughout the second year longitudinally without altering the context of the imitative situation. (i.e., adding social or instrumental cues). Thus, the question if the two functions change during the second year of life from a dominance of the cognitive function at the beginning to a dominance of the social function at the end remains open.

In order to determine whether a child's imitation is driven by an instrumental or social motivation, it is important to consider individual differences in children's social, emotional and temperamental processes (Heimann, 2002). In line with this, McCall et al. (1977) proposed that temperamental factors account for a significant proportion of the variability in young children's tendency to imitate. To gain a more comprehensive understanding of how imitation functions develop, future research should incorporate infants' individual characteristics in terms of their temperament into studies of young children's imitative behavior.

1.2 Temperament

From the earliest stages of life, individuals exhibit remarkable psychological and physiological differences. For instance, one infant may approach a new situation with curiosity and openness, while another may react with anxiety or hesitation. Similarly, some infants relish social interactions, whereas others may feel uncomfortable in such settings. In other words, infants may react differently to the same situation, even if they have grown up in the same environment. Observations of such striking differences, often initially noted by scientists observing their own children (Rothbart, 2011), laid the foundation for temperament research. Building on these early individual observations, a large number of theories, concepts and measures have been developed (Zentner & Bates, 2008). Four major approaches (i.e., Buss & Plomin, 1975, 1984; Goldsmith & Campos, 1982; Rothbart & Derryberry, 1981; Thomas & Chess, 1977) have dominated the modern research on infant temperament since it became one of the central themes of developmental psychology in the 1980s (Goldsmith et al., 1987; for an overview, see Fu & Pérez-Edgar, 2015; Zentner & Bates, 2008). While there are substantial differences in the conceptualization and measures of temperament between the four approaches, which were captured by all participants in their seminal roundtable discussion (Goldsmith et al., 1987), there is now general agreement that temperament reflects the child's early emerging, relatively stable basic dispositions of activity, attention, affectivity and self-regulation. These are also subject to change as they are the product of interactions between genetic, biological and environmental factors over time (Shiner et al., 2012).

1.2.1 Theories on Infant Temperament

Thomas and Chess (1977) introduced a first theoretical approach to temperament based on the results of their New York Longitudinal Study (NYLS; Thomas et al., 1963). A landmark in the field and the most widely known study in recent temperament research, the NYLS marked the beginning of modern interest in temperament research. Based on response patterns identified from parental interviews of 20 children at 3, 6, and 9 months, Thomas and Chess (1977) conceptualized temperament as reflecting behavioral styles that can be classified along nine dimensions: approach-withdrawal, activity level, regularity, adaptability, threshold of responsiveness, quality of mood, intensity of reaction, distractibility, and attention span/persistence. Based on their scores on the nine dimensions, children can be categorized as “easy”, “difficult” and “slow to warm up” types. They added the concept of the “goodness of fit” to describe the interaction between temperament and environment and its relation to adaptation: optimal development is more likely to occur under conditions of good fit, i.e., infants’ temperament is closely related to the demands, expectations and opportunities provided by their environment (Thomas & Chess, 1977). The theory has provided major contributions to the exploration of infant temperament as the list of dimensions has been seminal. However, concerns have been raised about the overlap between the conceptual definitions of the nine dimensions, as factor analyses of questionnaire items have revealed some redundancy between these dimensions (see Martin et al., 1994), and about their internal consistency, as some scales measuring the dimensions have been found to lack internal reliability (Rothbart & Mauro, 1990) and discriminant validity (Sanson et al., 1994). In addition, the theory has been criticized for its categorical approach, which oversimplifies the complexity of temperament (De Pauw & Mervielde, 2010), which is also reinforced by the fact only 65% of children could be fitted into the three categories on their NLYS (Thomas & Chess, 1980). The approach has further been criticized for its terminology “difficult” and “easy”. According to Putnam and colleagues (2002), this connotation is valuational and risks becoming a self-fulfilling prophecy as for example the classification of a “difficult child” may strengthen the child’s expression of theses “difficult” characteristics.

Buss and Plomin (1975, 1984) introduced a behavioral genetic model of temperament grounded in several defining criteria: temperament traits must be evolutionarily adaptive and phylogenetic relatives. Furthermore, they must appear early in life and be inherited. The final criterion is continuity: temperament traits must be relatively stable during development and therefore be predictive of later behaviors. They identified three dimensions that meet their criteria: 1) emotionality (E), the tendency to get easily distressed and upset, 2) activity (A), the tendency to be energetic or restless, and 3) sociability (S), the tendency to prefer the presence of others and engage in social interactions. This model was groundbreaking in its emphasis on genetic influences on temperament. However, it has been criticized for painting too narrow a picture, leaving out some traits that could reasonably be considered temperamental (Shiner & DeYoung, 2013). Consequently, it has had relatively less influence in contemporary temperament research (Fu & Pérez-Edgar, 2015).

The third major approach was provided by Goldsmith and Campos (1982). They defined temperament as individual differences in the tendency to experience and express all primary emotions (i.e., joy, interest, sadness, anger, fear) and their regulation. In contrast to Buss and Plomin's (1984) concept of emotionality, their definition included positive emotions. These differences manifest through intensity and temporal parameters in behaviors such as facial, vocal, and motor expressions. Building on this initial approach to temperament, Goldsmith and colleagues focused on examining genetic and environmental influences on observed variation in temperament (i.e., emotion reactivity and regulation), particularly the heritable component of temperament dimensions, using twin studies (Goldsmith et al., 1997; 1999). This behavioral genetic approach (as well as that of Buss & Plomin, 1975, 1984) has made important contributions to the understanding of the constitutional basis of temperament. More specifically, these longitudinal behavioral genetic designs allow researchers to determine the relative influence of genetic and environmental components on the development of temperament across age (Saudino, 2009). However, Goldsmith and Campos' (1982) initial concept of temperament is given less attention in current research as focusing on emotions and their regulation does not provide a comprehensive framework for understanding the broader aspects of temperament.

Rothbart and Derryberry (1981) proposed a psychobiological approach to temperament which defines it as constitutionally based individual differences in reactivity and self-regulation, primarily in the domains of affect, activity, and attention (Rothbart & Bates, 2006). There are three important points to this definition. "Constitutionally" refers to the biological basis of infants' temperament that is influenced by heredity, maturation, and experience over time. "Reactivity" refers to an individuals' response in terms of biological arousal to internal and external stimuli in the three domains. This includes a broad range of reactions on physiological and behavioral level, for example the amount of time it takes to return to a normal, calm state after being distressed or the persistence in a task when obstacles occur. "Self-regulation", as the regulatory component of temperament, refers to the neural and behavioral processes that enable individuals to modulate their reactivity. This reflects for example the ability to inhibit a dominant response in order to activate a subdominant response or to shift the attention elsewhere. These individual differences are commonly observed on three broad dimensions of temperament: 1) *Negative affectivity* and 2) *surgency/ extraversion*, reflecting the reactive aspects of temperament, and 3) *effortful control*, reflecting the regulative component of temperament (Rothbart & Bates, 2006; Rothbart & Derryberry, 1981). Using a dimensional approach, each of the three higher-order dimensions consists of multiple lower-order temperament traits (Rothbart & Bates, 2006). Negative affectivity includes scales such as discomfort, anger/frustration, sadness, and fear. Surgency/ extraversion includes scales such as expressing joy during daily activities, activity level, and expressions of pleasure in anticipation of rewards or during high intensity activities. Effortful control encompasses attention orienting and focusing, inhibitory control, and enjoying low intensity stimuli. Even though stability of temperament is one of the basic assumptions of this approach, the manifestation of temperament may change over time, as already noted. In addition to the maturational and experiential changes in the biological processes, the observed changes are also directly related to infants' growing self-regulatory capacities and therefore the growing influence of effortful control on behavior (Rothbart & Bates, 2006). A particularly significant factor in this evolution is the maturation of neural networks involved in executive attention. This maturation enhances the ability to regulate emotional and behavioral responses (Rothbart & Ahadi, 1994). Thus, changes in how temperament is expressed over time are less about shifts in underlying reactivity and more about the development and refinement of regulatory capacities.

Taken together, different approaches to temperament have been proposed, all of which have influenced the understanding of temperament and the direction of research (Fu & Pérez-Edgar, 2015; Gartstein et al., 2024; Shiner et al., 2012). Evolving from a categorical approach of observable behavioral styles (Thomas & Chess, 1977) to dimensional approaches focusing on genetic influences (Buss & Plomin, 1984) or the presence of emotions (Goldsmith & Campos, 1982), Rothbart and Derryberry (1981) proposed a broader dimensional approach focusing on reactivity and self-regulation at both physiological and behavioral levels. They provided a more extensive framework of temperament that allows for a better understanding of the interaction between temperament and environmental factors (Fu & Pérez-Edgar, 2015). Therefore, like much of the current research on temperament (e.g., Bornstein et al., 2015; Gartstein et al., 2017; Nakagawa et al., 2024; see Henderson & Wachs, 2007; see Gartstein et al., 2024), the present dissertation is based on Rothbart and Derryberry's (1981) psychobiological approach to temperament and their definition of temperament as constitutionally based individual differences in reactivity and self-regulation.

1.2.2 Measuring Temperament

In recent decades, various different methods have been developed to assess temperament in infants, primarily using caregiver reports and observational methods on behavioral level. For the latter, a distinction can be made between naturalistic home observations and laboratory observations (for an overview, see Rothbart & Bates, 2006). Each of these methods offer advantages for studying temperament. However, at the same time, they are also subject to error. Caregiver reports provide extensive insights from parents, who observe their infants in diverse situations over extended periods, thus reflecting typical behaviors over weeks or months. Additionally, caregiver reports are practical: they are highly efficient as they are time-saving and relatively inexpensive to develop, administer and analyze. At the same time, caregiver reports are susceptible to parental perceptual or responding biases (Gagne et al., 2011). Social desirability may play a role, as well as parental characteristics that may influence their interpretation of observed behavior and the problem that observed inconsistencies may be ignored (Kagan & Fox, 2006; Rothbart & Goldsmith, 1985). Observational methods, on the other hand, enable a more objective assessment of temperament and facilitate the examination of fine-grained temperament traits (Fu & Pérez-Edgar, 2015; Vroman et al., 2014). Naturalistic observations are characterized by high ecological validity, whereas laboratory observations allow researchers to control the context and therefore identify specific elicitors of infants' behavior. The time course and intensity of the infant's response are revealed (Rothbart & Bates, 2006). On the other hand, observational methods are situation and time specific. Naturalistic observations are relatively expensive and have low day-to-day reliability (Rothbart & Bates, 2006). Laboratory observations are limited as observations may lack ecological validity and the episodes may elicit emotional states rather than assess temperament traits (Goldsmith & Gagne, 2012). To overcome the limitations of the different methods, recent approaches have emphasized the relevance of a multi-method approach to temperament (Fu & Pérez-Edgar, 2015; Gartstein et al., 2024; Majdandžić & van den Boom, 2007; Planalp et al., 2017).

To date, most of the research on infant temperament has relied on caregiver reports. Various parent-report questionnaires have been developed based on the theoretical approaches outlined in section 1.2.1 (Theories on Infant Temperament). These measures are limited at the outset by the definition of their underlying approach of which behaviors constitute temperament (Shiner, 2015). Based on Rothbart and Derryberry's (1981) psychobiological approach, a series of age-related questionnaires have been developed that are among the most influential in current research (Shiner, 2015). One of the most widely used measures in recent research on infant temperament, developed within the psychobiological approach and its understanding of temperament, is the Infant Behavior Questionnaire Revised (IBQ-R; Gartstein & Rothbart, 2003). Within the IBQ-R, the three broad dimensions are assessed among different scales in 3- to 12-month-olds. *Surgency/Extraversion* assesses infants' enjoyment of social situations, their positive emotions in social situations, and their desire to interact with others. It includes scales such as *Activity Level*, which describes infants' gross motor activity; *Approach*, which describes infants' excitement and anticipation of pleasurable activities; and *High-intensity Pleasure*, which refers to the enjoyment of activities characterized by high intensity and novelty. *Negative Affectivity* describes several types of negative emotionality on scales such as *Fear* when facing surprising or changing stimuli, and *Distress to Limitations* after being restricted or unable to finish or perform an action. The factor assessing self-regulation is labeled *Orienting/Regulation* within the IBQ-R. The scales are meant to represent precursors of the later developing capacities of Effortful Control such as *Duration of Orienting,* the ability to direct attention toward a stimulus or situation, or *Falling Reactivity*, the rate of recovery from distress or arousal. The Early Childhood Behavior Questionnaire (ECBQ; Putnam et al.,

2006) is an extension of the IBQ-R for toddlers between 18 and 36 months. The basic assumption of three underlying factors remains, except that the regulation factor is called *Effort Control* within the ECBQ-R. Most of the scales remain identical, although some age-appropriate changes have been made.

The IBQ-R (Gartstein & Rothbart, 2003) and the ECBQ (Putnam et al., 2006) are based on two core assumptions derived from Rothbart and Derryberry's (1981) theoretical framework. First, as noted earlier, the questionnaires are based on a three-factor structure of temperament. In terms of the EBCQ, there is evidence for the validity of the three-factor structure. Various studies suggest the three-factor solution with Surgency/ Extraversion, Negative Affectivity, and Effortful Control as underlying dimensions, for example in American samples (Putnam et al., 2006; Putnam et al., 2008), a Belgian sample (Casalin et al., 2012), a Japanese sample (Sukigara et al., 2015), and a Polish sample (Stępień-Nycz et al., 2018). However, findings regarding the factor structure of the IBQ-R are less consistent. While the original factor structure of the IBQ-R (Gartstein & Rothbart, 2003) was found in an Italian and American sample (Montirosso et al., 2011), other studies revealed three factors as well, but the loadings of the scales on the factors differed from the original version. These divergent findings emerged, for example, in a Polish sample (Dragan et al., 2011), a Portuguese sample (Costa & Figueiredo, 2018), and a more diverse American sample (Bosquet Enlow et al., 2016). Studies with German infants even suggest a two-factor solution as they consistently failed to replicate Orienting/ Regulation factor resembling infants' regulatory capacities (Mink et al., 2013; Vonderlin et al., 2012). Thus, the factor structure underlying infants' temperament remains unclear, and, in line with Kiel et al. (2018), the factor structure of the German IBQ-R (Vonderlin et al., 2012) needs to be validated.

Second, the stability of temperament over time and different contexts is a fundamental aspect of its conceptualization. Within infancy, stability has been found for all scales of the

IBQ-R, for example between 6 and 12 months in Dutch and US infants (Sung et al., 2015) and in German infants (Mink et al., 2013), and between 6, 12, and 18 months in Korean infants (Gartstein et al., 2015). Within toddlerhood, the ECBQ was found to be equally reliable. Stability between 18 and 36 months was found for all scales in an American sample (Putnam et al., 2006) and a Japanese sample (Sukigara et al., 2015), and between 26 and 30 months in a Polish sample (Stępień-Nycz et al., 2017). Evidence for stability between the questionnaires is scarce. Casalin and colleagues (2012) found stability for all three factors in a Belgian sample from infancy (8-13 months) to toddlerhood (20-25 months). Given the difficulties in replicating the factor structure discussed above, a fine-grained approach at the scale level may be more appropriate for determining the stability of temperament between the two questionnaires. Putnam and colleagues (2008) used this approach and found stability at factor level and across the eleven corresponding scales of the IBQ-R and ECBQ between 3-12 month old infants and 18-32 month old toddlers. Aggregating data of children of a wide age range, especially in periods of rapid development as infancy and toddlerhood (Bornstein et al., 2015), leads to a potential loss of information as possible changes or stabilities are overlooked. Thus, in order to draw a more coherent picture of temperament stability in early development from infancy (IBQ-R) to toddlerhood (ECBQ), it is necessary to assess stability at scale level between questionnaires, without data from aggregated age groups.

1.3 Imitation and Temperament

As outlined in section 1.1.3 (Functions of Imitation), individual differences need to be taken into account to determine whether children's imitative behavior is driven by an instrumental motivation to learn new skills or by their social motivation to engage with the model. In light of Rothbart and Derryberry’s (1981) psychobiological approach, infants’ temperament might be related to their motivation to imitate. As outlined, Surgency/Extraversion describes infants’ desire to interact with others and their enjoyment of social situations. Accordingly, this factor might promote the social function of imitation. The self-regulation factor, termed as Orienting/Regulation up to 12 months (Gartstein & Rothbart, 2003) or Effortful Control for toddlers between 18 to 24 months (Putnam et al, 2006), describes children's regulatory capacities related to their attention. These attention capacities include their abilities to direct and sustain attention, and to resist distraction. Accordingly, infants' attention capacities as a prerequisite of learning (Bandura, 1977), might promote the cognitive function. A relation between imitation and temperament has been found on both, biological and behavioral level (for an overview, see Heimann, 2022b).

First, findings from the behavioral genetic perspective can be integrated into the relation between imitation and temperament in terms of the psychobiological approach. In a review on imitation and individual differences, Fenstermacher and Saudino (2006) proposed that variability in imitative behavior reflects a direct genetic influence on children's ability to imitate. As temperament is biological based on its definition, a link between imitation and temperament may reflect overlapping genetic factors underlying infants' imitative behavior. They provided evidence of a genetic influence on imitation in a twin study on deferred imitation in 24-month-olds (Fenstermacher & Saudino, 2007) with correlations of monozygotic twins (MZ) significantly exceeding correlation of dizygotic twins (DZ). Fenstermacher and Saudino (2016) extended their initial findings by investigating the relation between imitation and temperament on a genetic basis in MZ and DZ twins. They found an association between imitation and extraversion, and imitation and task orientation (assessing children's attention span and persistence), which was partly explained by genetic influences. In addition, findings on the serotonin transporter gene 5-HTTLPR suggest a genetic basis for individual differences associated with imitation. The short allele of the 5-HTTLPR has been found to increase the susceptibility to environmental stimuli, while individuals with the long allele are more resilient (Belsky et al., 2009). Schroeder and colleagues (2016) found a link between the short allele of 5-HTTLPR, that is associated with increased social sensitivity, and increased imitative behavior in 2-year-olds. However, this result must be interpreted with caution: the effect of the short allele may depend on another serotonin transporter gene length polymorphism whose occurrence varies worldwide and therefore may not be generalizable to other populations (Schroeder et al., 2016). This is a common problem in many behavioral genetic studies using the one-gene-one-outcome approach, as behavior is often determined by a large number of genes and their interactions (Plomin et al., 1994). This approach has therefore been discouraged in favor of genome-wide association studies, which include a wider range of genes associated with phenotypes (Götz et al., 2021). Interpretation of these studies, which attempt to predict behavior on the basis of individual genes, therefore remains speculative. Taken together, these findings suggest, that imitative performance shares some underlying genetic influences with attentional and social aspects of temperament (see Fenstermacher & Saudino, 2016) in infancy.

Additional evidence linking imitative behavior to attentional and social aspects of temperament comes from studies examining temperament at the behavioral level. Attention, which is a key component of the regulation factor and a precursor to Effortful Control, has been related to infants' imitative behavior from an early age. Zmyj and colleagues (2017) found that attention, as measured by temperament questionnaire and laboratory observation, was associated with imitative behavior in 12-month-olds. The importance of attention for imitation is also evident over the course of the second year of life. Twenty-month-olds with a high attentional focus were less influenced by environmental distractions during an imitation task compared to those with lower attentional focus (Dixon et al., 2006). Effortful Control was related to long-term imitative recall in 21-month-olds (Dixon et al., 2012). However, the relation between social aspects of temperament and imitation did not become evident until later in the second year. While imitation and Extraversion were not correlated at 12 months

(Zmyj et al., 2017), 15-month-olds with higher levels of Extraversion were more likely to imitate faithfully than those with low levels of Extraversion (Hilbrink et al., 2013). Moreover, Extraversion at 15 months was found to predict imitative behavior at 21 months (Dixon et al., 2012). These findings indicate that attention appears to be linked to imitative behavior from 12 months (Zmyj et al., 2017) to the end of the second year of life (Dixon et al., 2006, 2012). In contrast, the relation between Extraversion and imitation emerges from the middle of the second year of life (Hilbrink et al., 2013), but is not present at the beginning (Zmyj et al., 2017).

However, several unresolved issues remain regarding the relation between imitation and temperament during the second year of life. First, no longitudinal study has covered the entire age range of the second year. While some of the studies employed cross-sectional designs (Dixon et al., 2006; Zmyj et al., 2017), Hilbrink and colleagues (2013) focused exclusively on the initial six months of the second year of life, while Dixon and colleagues (2012) concentrated on the subsequent six months. Second, there are limitations in the comparability of the imitation tasks used. While Zmyj and colleagues (2017) used a standardized instrumental task, Hilbrink and colleagues (2013) used a task in which the instrumental cues of the context were varied. Regarding the other studies, information on psychometric properties (i.e., reliability and validity) of the imitation tasks are mostly lacking (Dixon et al., 2006; 2012). In general, comparing imitation studies proves challenging due to methodological inconsistencies, as the tasks differ widely in terms of content, complexity, number of tasks and action steps, and the required retention intervals (within deferred imitation tasks and between deferred and immediate performance; Fagard & Lockman, 2010). Third, temperament was primarily assessed via questionnaire (Dixon et al., 2006; 2012; Hilbrink et al., 2013). Against the background of the discussed multi-method approach to temperament, additional observational measures are necessary.

1.4 Imitation and Attention

According to Bandura’s social learning theory (1977), imitation represents a form of observational learning driven by vicarious reinforcement. In this process, learning occurs through observing other’s behavior and their consequences. The child does not experience the consequences of the behavior itself, but only observes them in the model. He proposed four crucial processes that influence an individual’s response to a modeled behavior and therefore determine whether or not imitation occurs. The first central process to enable the acquisition of a novel behavior is attention to the behavior. The second is retention, by storing the observed behavior in symbolic form in the infant's memory. The third is the motor reproduction process that takes place on a cognitive level. Internal symbolic images of already known behaviors are used as a reference to replicate the relevant behavior internally. In the last phase, motivational processes are considered. Affirmation type, e.g., reinforcement directly experienced, inferred from observations or through verbal explanation, and individual factors, e.g., preferences, determine which behaviors are finally imitated.

According to Bandura's (1977) considerations, attentional processes are a central feature of imitation, not only as part of temperament, but above all in terms of perception. A mere exposure to the model does not ensure that the imitator will perceive the action or behavior in question. In line with the behavior-based definition of imitation (Paulus, 2011), attention to the action is necessary to reproduce it. However, evidence from studies on infants’ looking time at the model, which is often used as a measure of overt attention (Óturai et al., 2013), is lacking. Looking time was not predictive of their imitative behavior at 6 and 9 months (Taylor & Herbert, 2013), at 12 months (Kolling et al., 2014) and at 18 months (Óturai et al., 2013). Sonne and colleagues (2016) even found a negative correlation between looking time and memory performance in 20-month-olds. As simple looking time is not conclusive about the relation between attention and imitation, studies that examine gaze direction using eye-tracking may be more informative in exploring the role of attention in imitating actions.

Interestingly, in another study of 6-, 9- and 12-month-old infants, Taylor and Herbert (2014) found that imitators showed increased attention to the model demonstrating the action than to the puppet where the relevant action was shown. According to Taylor and Herbert (2014), increased attention to the model may have facilitated infants' ability to infer the goal of the action, and thus the imitation of the action. Consequently, infants' ability to understand the model's intention may be related to their imitative behavior (see also Tomasello et al., 2005). Following Carpenter and Call (2009), in order to imitate, infants' need to identify and understand the action goal of the model. Infants' action understanding is typically assessed using their goal-predictive gaze behavior during an ongoing task using eye-tracking (for an overview, see Hunnius & Bekkering, 2014). Goal prediction refers to the ability to anticipate the goal of an observed action before the action is completed, and goal-predictive gaze shifts have been used to measure action understanding in infants (e.g., Falck-Ytter et al., 2006; Melzer et al., 2012). To date, only one previous study has investigated the link between imitation and goal-predictive gaze behavior. Gampe and colleagues (2016) presented 12–30-month-olds with two multi-step actions on a computer screen, one familiar action (i.e., hammering) and one unfamiliar action (i.e., pulling) for this explicit context (i.e., inserting a building block in a box). The faster infants anticipated the action goal in the hammering action, the more likely they were to imitate this action. The results suggest that goal prediction is related to imitation of familiar actions but not unfamiliar actions.

However, three critics can be raised regarding the methodology used by Gampe et al. (2016). First, the actions were presented on a computer screen. Most of the knowledge about infants' anticipatory looking consists of findings from eye-tracking studies of screen-based paradigms (e.g., Cannon et al., 2012; Falck-Ytter et al., 2006; Hunnius & Bekkering, 2010;

Melzer et al., 2012). These controlled paradigms include several advantages, e.g., using "occluders" in the movement path to prevent infants from simply tracing the object (e.g., Paulus et al., 2017) or excluding distractors by only presenting the hand that is moving the object instead of the actor's face (e.g., Cannon & Woodward, 2012). As a result, they maximize the opportunities to anticipate. However, they do not reflect the actual learning environment of infants. In the real-life interactions that constitute infants' learning environment, they are not just confronted with a small snippet that focuses only on the action. Rather, they have to divide their attention between their partner, their partner's movements and the surrounding context. Furthermore, presenting actions on a video screen may have underestimated young children's performance due to the video deficit effect, according to which young children show a reduced ability to transfer learning from videos compared to learning from equivalent real-life experiences (Anderson & Pempek, 2005). The second issue is related to the problem that screen-based paradigms do not represent everyday life well. Another advantage of controlled paradigms is that the relevant actions can be demonstrated more slowly than movements in real life. Consistent with other studies on action prediction, the actions demonstrated by Gampe and colleagues (2016) were relatively slow (i.e., 1000-1960 ms per action step). The relevant action steps in other studies investigating goal prediction are even longer (e.g., approximately 2000ms, Ambrosini et al., 2013). In a recent study by Ganglmayer and colleagues (2019), infants even had 5000ms to predict an action goal. For everyday actions, children have less time to anticipate an action. This raises the question of whether infants are able to predict action goals in everyday life. Monroy and colleagues (2021) were the first to investigate action prediction in real life during free-flowing infant-parent interactions. 9-month-olds anticipated their parents' goal-directed movements. However, these predictions were infrequent, averaging fewer than one anticipation per minute of interaction. Third, Gampe and colleagues (2016) did not control for children's general cognitive development. Accordingly, imitation and anticipatory looking may result from changes in cognitive developmental status rather than action prediction.

Taken together, attentional processes are crucial in imitative behavior (Bandura, 1977; Paulus, 2011). Since simple looking time does not conclusively reveal the relation between attention and imitation, exploring infants' gaze direction could offer more valuable insights (Taylor & Herbert, 2014). Previous research suggests that infants' goal prediction may be related to their imitative behavior (Gampe et al., 2016). However, concerns have been raised regarding the methodological limitations of screen-based paradigms used to investigate goal-predictive gaze behavior, as these approaches differ significantly from real-life presentations in several key aspects.

2 Outline of the Thesis

The preceding review of the literature highlights that understanding infants' imitative behavior necessitates considering not only their cognitive abilities but also their motivation to engage in imitation. While researchers agree upon the fact that imitation servers both a cognitive function, that is instrumentally motivated, and a social function, that is socially motivated, less is known about the development of these two functions in infancy. Specifically, the question whether the two functions change within the second year of life from a dominance of the cognitive function at the beginning to a dominance of the social function at the end, remains open. One of the primary objectives of this thesis was to explore this possible change in the two functions of imitation within the second year of life. In order to understand why infants engage in imitation, in this collaborative project I aimed to explore the extent to which imitation serves as a tool for learning or interacting with others. Irrespective of cognitive abilities, a more extroverted child, who attaches importance to engage with the model, might be more socially motivated to imitate the model, whereas the opposite may be true for a less extroverted child. Therefore, to assess infants' individual motivation independent of their cognitive abilities, the relation between imitation and the social and cognitive aspects of temperament was examined. To the best of my knowledge, to date no other measure of infants' instrumental motivation exists. Furthermore, given the important role of attention in the imitation process (e.g., Bandura, 1977), I aimed to investigate infants' gaze direction during the action demonstration.

Building on these considerations, a systematic longitudinal study in 12-, 18-, and 24-month-olds was designed, using standardized, age-appropriate instrumental imitation tasks, and parent reports and observational tests of temperament. Additionally, the study incorporated cognitive development tasks and implemented eye-tracking during the action demonstration. As discussed in section 1.2.2 (Measuring Temperament), the use of the parent report questionnaire needed to be validated first in order to use it on scale and factor level. Thus, the study design allowed three questions to be investigated.

The first article examined the factor structure of the German versions of the temperament questionnaires IBQ-R and ECBQ in two samples. In light of the inconclusive evidence regarding the three-factor structure of the IBQ-R, we were able to draw on another longitudinal data set of 9- and 12-month-olds in addition to the underlying longitudinal study.

The second article focused on the developmental changes in the two functions of imitation during infancy. We related infants' imitative behavior to social and cognitive aspects of temperament, both cross-sectional and longitudinally, while controlling for their cognitive development.

The third article offers a methodological contribution. To study infants' attention during action demonstration, we introduce a novel methodological approach for analyzing their gaze direction using a live eye-tracking setting where actions are presented at naturalistic speeds. We related 12-month-olds' imitative behavior to their goal-predictive gaze shifts, controlling for their cognitive development.

3 Article 1 – Sieber, F., & Zmyj, N. (2022). Stability and structure of infant and toddler temperament in two longitudinal studies in Germany. *Infant Behavior and Development, 67*, 101714.

Stability and structure of infant and toddler temperament in two longitudinal studies in Germany

Franziska Sieber & Norbert Zmyj

Department of Psychology, TU Dortmund University

Published in *Infant Behavior and Development, 67,* 101714.

https://doi.org/10.1016/j.infbeh.2022.101714

* **Corresponding author:** Franziska Sieber, TU Dortmund University, 44221 Dortmund, Germany. E-mail address: franziska.sieber@tu-dortmund.de

Abstract

The present study investigated the factor structure and longitudinal stability of infant and toddler temperament measured with the Infant Behavior Questionnaire-Revised (IBQ-R) and the Early Childhood Behavior Questionnaire (ECBQ) in two German samples. Since the few studies using the German IBQ-R failed to replicate the commonly assumed three factors of infant temperament, another exploration of its factor structure was necessary. Whereas previous stability measurements are usually based on groups with relatively large age ranges, we investigated stability between concrete ages. In Study 1, the IBQ-R was applied in 9- and 12-month-old infants. In Study 2, both questionnaires were applied in 12-, 18-, and 24-month-olds. Factor analyses of the IBQ-R in both studies revealed a two-factor solution comprising Surgency/ Extraversion and Negative Affectivity, replicating earlier findings with German infants. The Orienting/ Regulation factor was not replicated for the IBQ-R, whereas analysis of the ECBQ revealed a three-factor solution for toddlers. The results showed stability of temperament ratings within infancy (Study 1) and toddlerhood (Study 2) as well as between the developmental periods within the second year of life (Study 2). Taken together, the present findings indicate that temperament stability emerges early in life and can be reliably measured using the IBQ-R and ECBQ, whereas the factor structure of temperament in infancy requires further validation.

Keywords: temperament, stability, structure, infancy, toddlerhood

Stability and structure of infant and toddler temperament in two longitudinal studies in Germany

Infancy and toddlerhood is characterized by rapid physical and psychological development (Vroman et al., 2014). Infant temperament is among the earliest psychological differences in childhood and thus often used to predict later development, such as personality in school years (Slobodskaya & Kozlova, 2016) and adulthood (Caspi & Shiner, 2006), behavioral problems (Abulizi et al., 2017), and mental health (Kozlova et al., 2020). These predictions are commonly based on the Infant Behavior Questionnaire Revised (IBQ-R, Gartstein & Rothbart, 2003) and rest on two premises. The first assumption is that infant temperament consists of three factors (Surgency/ Extraversion, Negative Affectivity, and Orienting/ Regulation). Although a large body of research found the three-dimensional structure to be adequate (see Bosquet Enlow et al., 2016), the few studies using the German version of the IBQ-R indicated a two-factor structure (Mink et al., 2013; Vonderlin et al., 2012). The second assumption is that infant temperament is stable across time. Previous research investigated stability by aggregating age groups over several months in infancy and toddlerhood (Casalin et al., 2012; Putnam et al., 2008). Since a couple of months can be key for development in early years, measuring temperament at concrete age points could improve the measure of stability.

In the present study, we further investigated the German version of the IBQ-R by testing its factor structure and stability in two samples of longitudinal studies with narrow age ranges at each time point.

1 Temperament

Temperament is composed of different character traits concerning emotions, arousability, and regulation (Gartstein & Rothbart, 2003). It has its roots in infancy (Rothbart & Derryberry,

1981) and forms the basis for later personality (for an overview, see Zentner & Bates, 2008). According to the psychobiological approach by Rothbart and Bates (2006), a child's temperament is defined as constitutionally based individual differences in reactivity and self-regulation. These individual differences can be observed in the domains of affect, activity, and attention (Goldsmith et al., 1987). Constitutional means that the biological basis of temperament is influenced over time by heredity, maturation, and experience. Reactivity refers to an individual's response to changes in the internal and external environment, and includes a broad range of reactions on behavioral and physiological levels (Rothbart & Derryberry, 1981). Self-regulation, as the regulatory component of temperament, refers to neural and behavioral processes functioning to modulate reactivity (Rothbart et al., 1994). As such, self-regulation reflects the ability to inhibit a dominant response in order to activate a subdominant response (Rothbart & Bates, 2006).

1.1 Stability of temperament

The psychobiological approach defines the stability of temperament characteristics over time and different contexts as a central feature. This does not imply that temperament characteristics are immutable, but rather that they develop as a function of maturation and experience (Rothbart & Derryberry, 1981). Whereas reactivity is thought to be present at birth and relatively stable, self-regulation develops during the first year of life. The maturation of neural networks of executive attention leads to an increasing capacity of the child to regulate emotional and behavioral responses (Rothbart et al., 1994). Accordingly, changes in the expression of temperament are attributed to the development of executive functions that allow for the regulation of emotional and behavioral responses, and not to changes in the underlying reactivity (Riese, 1987). Since stability is dynamic (Bornstein et al., 2015), and development is rapid and enduring in childhood (Bornstein et al., 2013), manifestations of temperament change over time (Gartstein & Rothbart, 2003). For example, levels of activity, approach, distress to limitations, and fear increase in the first year of life (Carranza et al., 2000). Thus, developmental changes in the expression of temperament are expected during infancy and toddlerhood.

A series of age-specific questionnaires to assess temperament across childhood have been developed based on the psychobiological approach, for example the Infant Behavior Questionnaire Revised (IBQ-R; Gartstein & Rothbart, 2003), which consists of 14 scales (see Table 1). Factor analyses of the underlying structure of infant temperament suggested three broad dimensions: Surgency/ Extraversion, Negative Affectivity, and Orienting/ Regulation, with Surgency/ Extraversion and Negative Affectivity referring to the reactivity component and Orienting/ Regulation referring to the self-regulatory component. The Early Childhood Behavior Questionnaire (ECBQ; Putnam et al., 2006) is an extension of the IBQ-R for toddlers. Eleven scales are similar between the IBQ-R and the ECBQ, with the IBQ-R scale *Duration of Orienting* labeled *Attention Focusing* in the ECBQ, and the IBQ-R scale *Approach* measured as *Positive Anticipation* in the ECBQ. The IBQ-R scales *Soothability*, *Vocal Reactivity*, and *Smiling and Laughter* are not part of the ECBQ, while seven scales are added, resulting in a total of 18 scales in the ECBQ (see Table 2).

Table 1

Scale definitions of the IBQ-R (Gartstein & Rothbart, 2003)

Activity Level [SE]	Gross motor activity, including movement of arms and legs, squirming and locomotor activity.
Distress to Limitations [NA]	Fussing, crying or showing distress while (a) in a confining place or position; (b) in caretaking activities; (c) unable to perform a desired action.
Fear [NA]	Startle or distress to sudden changes in stimulation, novel physical objects or social stimuli; inhibited approach to novelty.
Duration of Orienting [OR]	Attention to and/or interaction with a single object for extended periods of time
Smiling and Laughter [SE]	Smiling and Laughter during general caretaking and play.
High Intensity Pleasure [SE]	Pleasure or enjoyment related to high stimulus intensity, rate, complexity, novelty, and incongruity.

Low Intensity Pleasure ^{OR}	Amount of pleasure or enjoyment related to low stimulus intensity, rate, complexity, novelty, and incongruity.
Soothability ^{OR}	Reduction of fussing, crying or distress when soothing techniques are used by the caregiver.
Falling Reactivity ^{NA}	Rate of recovery from peak distress, excitement, or general arousal; ease of falling asleep.
Cuddliness ^{OR}	Expression of enjoyment and molding of the body to being held by a caregiver.
Perceptual Sensitivity ^{SE}	Detection of slight, low intensity stimuli from the external environment.
Sadness ^{NA}	Lowered mood and activity related to personal suffering, physical state, object loss, or inability to perform a desired action; general low mood.
Approach ^{SE}	Rapid approach, excitement, and positive anticipation of pleasurable activities.
Vocal Reactivity ^{SE}	Amount of vocalization exhibited by the baby in daily activities.

Note. ^{SE} = Surgency/ Extraversion; ^{NA} = Negative Affectivity; ^{OR} = Orienting/ Regulation.

Table 2

Scale definitions of the ECBQ (Putnam et al., 2006)

Activity Level	Level (rate and intensity) of gross motor activity, including rate and extent of locomotion.
Attentional Focusing	Sustained duration of orienting on an object of attention; resisting distraction.
Attentional Shifting	The ability to transfer attentional focus from one activity/ task to another.
Cuddliness	Child's expression of enjoyment in and molding of the body to being held by a caregiver.
Discomfort	Amount of negative affect related to sensory qualities of stimulation.
Fear	Negative affect related to anticipated pain, distress, sudden events and/or potentially threatening situations.
Frustration	Negative affect related to interruption of ongoing tasks or goal blocking.
High-Intensity Pleasure	Pleasure or enjoyment related to situations involving high intensity, rate, complexity, novelty and incongruity.
Impulsivity	Speed of response initiation.
Inhibitory Control	The capacity to stop, moderate, or refrain from a behavior under instruction.
Low-Intensity Pleasure	Pleasure or enjoyment related to situations involving low intensity, rate, complexity, novelty and incongruity.
Motor Activation	Repetitive small-motor movements; fidgeting.
Perceptual Sensitivity	Detection of slight, low intensity stimuli from the external environment.
Positive Anticipation	Excitement about expected pleasurable activities.
Sadness	Tearfulness or lowered mood related to suffering, disappointment, or loss.
Shyness	Slow or inhibited approach and/or discomfort in social situations involving novelty or uncertainty.
Sociability	Seeking and taking pleasure in interactions with others.
Soothability	Rate of recovery from peak distress, excitement, or general arousal.

Several longitudinal studies have found temperament to be a moderately stable trait over time in the first years of life (Roberts & DelVecchio, 2000). Stability was observed in laboratory behavioral measures in infancy (e.g., Planalp & Goldsmith, 2020), between behavioral measures in infancy and parent report (Rothbart et al., 2000), as well as in many parent report measurements over different periods of time (e.g., Carranza Carnicero et al., 2013; Olafsen et al., 2018), separately for mothers (Komsi et al., 2006) and fathers (Komsi et al, 2008), from infancy to childhood. Thus, stability occurs both within and between developmental periods and across different methods of measurement.

As the IBQ-R (infancy) and the ECBQ (toddlerhood) were developed within the psychobiological approach, stability of temperament is expected using this method of measuring temperament. Indeed, several longitudinal studies have demonstrated stability within the IBQ-R (e.g., Gartstein et al., 2015; Kusanagi et al., 2014) and within the ECBQ (e.g. Putnam et al., 2006). For example, Gartstein et al. (2015) found temperament stability in South Korean infants from 6 to 12 to 18 months using the IBQ-R (see also Desmarais et al., 2019, for US and Dutch infants; Mink et al., 2013, for German infants). To the best of our knowledge, only a few longitudinal studies have reported temperament stability between the IBQ-R and ECBQ and focused on the age range of the first two years of life. Casalin et al. (2012) found stability for all three factors in a Belgian sample (see also Nakagawa & Sukigara, 2014, for a Japanese sample). Nevertheless, the key question of temperament stability could be answered more precisely by analyzing temperament on scale level. This fine-grained approach has several advantages. First, the factor structure differs between different countries and some scales load on different factors in different countries (for an overview, see Bosque Enlow et al., 2016). Especially in Germany, the Orienting/ Regulation factor was not replicated (Mink et al., 2013; Vonderlin et al., 2012). Furthermore, Casalin et al. (2012) left some scales out due to practical reasons. Second, some developmental processes underlying stability may be evident beyond factor-level as some common mechanisms may contribute to diverse behavioral traits (Putnam et al., 2008). Thus, a fine-grained approach is necessary to evaluate stability. The study focusing on longitudinal stability between IBQ-R and ECBQ by Putnam et al. (2008) used a fine-grained approach and found stability on factor level as well as between the eleven corresponding scales of the IBQ-R and ECBQ in an American sample. The sample recruited by Putnam et al. (2008) consisted of infants and toddlers from a large age range (3-12-month-old infants, 18-32-month-old toddlers). Since stability increases after 2 years (Lemery et al., 1999), analyzing children below and above the second birthday as one group may affect the stability coefficients. Casalin et al. (2012) as well recruited infants and toddlers from a larger age range (8-13-month-old infants, 20-25-month-old toddlers). To assess if differences in stability coefficients do occur within these comprised age groups that may reflect developmental changes in specific ages, it is necessary to investigate stability not only over the whole age range of toddlerhood but at different concrete ages using a fine-grained approach.

1.2 Structure of Temperament

In addition to assessing individual differences in temperament characteristics, the measures have been used to identify the structure of infant temperament via factor analysis (e.g., Bosquet Enlow et al., 2016; Casalin et al., 2012; Putnam et al., 2008, see Table 1). The three-factor structure has been replicated in an American and an Italian sample (Montirosso et al., 2011). While other studies also revealed a three-factor solution, there were differences in scale loadings on factors within the US (Bosquet Enlow et al., 2016) as well as across different countries, e.g. Ethiopia (Gartstein et al., 2016), Portugal (Costa & Figueiredo, 2018), and Poland (Dragan et al., 2011). The German translation of the questionnaire (Vonderlin et a., 2012) and the validation of the IBQ-R in Germany revealed only a two-factor solution, with the Orienting/ Regulation factor not being replicated (Mink et al., 2013; Vonderlin et al., 2012). Since the three-factor structure failed to be replicated in Germany, other studies using the German IBQ-R (Zmyj et al., 2017) analyzed infant temperament only on scale level. A current study on the factor structure of the IBQ-R found the three-factor structure to be adequate for Portuguese infants even as young as 2 weeks over the whole first year of life to 12 months (Dias et al., 2021). Thus, as Kiel et al. (2018) stated, there is a need to validate the factor structure of the German IBQ-R.

Findings regarding the factor structure of the ECBQ (Putnam et al., 2006) are more consistent. Different studies suggest the three-factor solution, with Surgency/ Extraversion, Negative Affectivity, and Effortful Control as underlying dimensions (Casalin et al., 2012; Putnam et al., 2006; Putnam et al., 2008).

In the framework of two different longitudinal projects, we investigated infants' temperament at the end of the age group of the IBQ-R and at the beginning of the age group of the ECBQ. The age points were selected within the overarching topics of the two projects. Nevertheless, due to the rapid development in the months around the first year of life, they are of particular interest regarding the question of stability between these different developmental phases. Past research mostly studied stability by using age groups with relatively large age ranges (Casalin et al., 2012; Putnam et al., 2008) or evaluated stability only on factor level focusing on the latent superconstructs (Casalin et al., 2012). The structure of infant and toddler temperament was also validated in age groups with relatively large age ranges in the original IBQ-R (Gartstein & Rothbart, 2003) and ECBQ (Putnam et al., 2006). We investigated the stability and structure of infant and toddler temperament without aggregating age groups over the different age ranges of the two questionnaires focusing on the fine-grained approach as the factor structure of infant temperament, especially in Germany, is not consistent. Thus, we aimed to answer the question whether possible differences within the age groups, that are often comprised, do occur, that may indicate developmental changes.

1.3 The present studies

The first aim of this study was to test the reliability of the German versions of the IBQ-R (Vonderlin et al., 2012) and the ECBQ (Kirchhoff et al., 2013) on the scale and factor level. We were especially interested in the factor structure of the IBQ-R in view of the mixed findings in the past. The second aim was to test the stability of infants' and toddlers' temperament. Therefore, we conducted two longitudinal studies, with Study 1 testing 9- and 12-month-olds and Study 2 testing 12-, 18-, and 24-month olds. In so doing, we were able to test stability both within one questionnaire (Study 1) and across different measures and thus different developmental periods (Study 2). We expected stability to occur within the IBQ-R and within the ECBQ as well between the corresponding scales of the two measures. Due to the exploratory nature of the study in terms of the factor structure of the IBQ-R and ECBQ, no specific predictions were made.

2 Study 1

2.1 Methods

2.1.1 Participants

Data used for this study were taken from a longitudinal study on self-awareness (Klein-Radukic & Zmyj, 2015, 2020). Participants in this study were recruited from a database of families who had previously registered to participate in child development studies. All families lived in one of three medium-sized cities in the Ruhr area of Germany. Parents were given the questionnaires one week prior to the first test phase in the laboratory at 9 months and asked to bring the completed questionnaires along to the laboratory session. For the second test phase at 12 months, the questionnaires were sent to the parents by mail and they were asked to return the completed questionnaires in provided prepaid envelopes.

The original sample at Time 1 (T1) consisted of 109 parents with infants around the age of 9 months. Of these 109 families, 13 were excluded due to more than 20% missing data. The mean age of the remaining 96 children (48 girls) at T1 was 9 months 10 days (*SD* = 10 days; range 262 - 321 days). All children were Caucasian. Most of the parents had either a university degree (mothers 43.8%, fathers 46.9%) or a university entrance-level diploma (mothers 16.7%, fathers 9.4%) as their highest educational attainment. The main native language of mothers and fathers was German (77.1% for mothers, 72.9% for fathers).

Of the 109 initially recruited families, 60 infants (28 girls) around the age of 12 months participated at Time 2. The mean age was 12 months 6 days (*SD* = 8 days; range 352– 391 days). The analyses of internal consistencies and factor structures were conducted with the different sample sizes for each time point.

For longitudinal stability and comparisons of age, we only included children if temperament data were available at both T1 and T2. Therefore, the final sample for longitudinal analyses and developmental changes consisted of 53 children (27 girls) with a mean age of 9 months 9 days (*SD* = 10; range 264 - 321 days) at T1 and 12 months 6 days (*SD* = 8; range 352 - 387 days) at T2.

2.1.2 Material

At both age points, the German version (Vonderlin et al., 2012) of the IBQ-R (Gartstein & Rothbart, 2003) was used to assess infant temperament. The IBQ-R was designed to measure temperament in infants aged between 3 and 12 months. The questionnaire assesses early infant temperament with 191 items that are grouped into 14 scales (see Table 1). Parents are asked to indicate on a 7-point Likert scale (1 = never, 7 = always) how often a specific infant behavior occurred in the past week, or in the past two weeks for some items. Additionally, parents have the possibility to mark an item with "does not apply" if they were not able to observe their child in this specific situation. Internal consistency of the 14 scales of the German version ranged between .74 and .94 (Cronbach's α, Vonderlin et al., 2012). In the current study, Cronbach's αs for the subscales ranged from .64 to .87 at 9 months and from .69 to .91 at 12 months.

2.1.3 Coding and analyses

Items that were marked with "does not apply" as well as omitted items were counted as missing values. Questionnaires with more than 20% missing values were excluded from the analysis. Missing values for the remaining questionnaires were replaced with the mean of the corresponding subscale (see Pauli-Pott et al., 2003; Vonderlin et al., 2012). Stability of IBQ-R scales was analyzed using Pearson product-moment correlations. Paired *t*-tests or Wilcoxon tests were computed to assess developmental changes depending on whether the data were normally distributed. Confirmatory factor analyses (CFA) were conducted to investigate the underlying structure of the respective temperament questionnaire. The three-factor model described by Gartstein and Rothbart (2003) and the two-factor model obtained on the German version of the IBQ-R by Vonderlin et al. (2012) were tested, both without sporadic occurred double loadings. All analyses were made using the structural equation modeling software (SEM) IBM SPSS AMOS 28.0.0.

2.2 Results

2.2.1 Factor structure

Before examining the stability and possible developmental changes, the factor structure of infant temperament was analyzed for each age. Both models did not match the data at 9 months. The match statistics for the three-factor model were $\chi^2(74, N = 96) = 168.70$, $p < .001$; comparative fit index (CFI) = .68; Tucker-Lewis coefficient (TLI) = .61 ; root-mean-square error approximation (RMSEA) = .12 ; Akaike information criterion (AIC) = 258.70. For the two-factor model, the statistics were the following: $\chi^2(76, N = 96) = 177.21$, $p < .001$; CFI = .66; TLI = .59; RMSEA = .12; AIC = 263.21. Due to the poor fit of both models, we additionally reconstructed the factor structure using exploratory factor analyses (for a similar procedure, see Dragan et al., 2011). We first used Velicer's minimum average partial (MAP) test and parallel analysis (O'Conner, 2000) to determine the number of factors (for details, see Supplementary methods). Subsequently, a principal axis factoring with an oblimin rotation was carried out. Whereas the MAP test revealed a two-factor solution, a three-factor solution emerged from parallel analysis of the IBQ-R scales at 9 months (see Supplementary results Study 1). Therefore, we conducted the principal axis extraction analysis with an oblimin rotation for both possible numbers of factors. The suggested two-factor solution was nearly identical to that of Vonderlin et al. (2012) (see Table 3), whereas the number of iterations in the three-factor solution was too large for the principal axis extraction analysis. The analysis with three factors could not be conducted for this sample. Parallel analysis tends to overextract the number of factors (O'Connor, 2000). Thus, data of this sample indicate a two-factor rather than a three-factor solution.

At 12 months, the models tested with confirmatory factor analyses again did not match the data. The match statistics for the three-factor model (Gartstein & Rothbart, 2003) were $\chi^2(74, N = 60) = 141.35$, $p < .001$; CFI = .56; TLI = .46; RMSEA = .12; AIC = 231.35. Even though the two-factor model had a poor fit as well ($\chi^2(76, N = 60) = 136.35$, $p < .001$; CFI = .61; TLI = .53; RMSEA = .12; AIC = 222.35), comparing both models AIC coefficients with a smaller AIC of the German model rather suggests a two than a three-factor solution. Nevertheless, because both models showed a poor fit, we again conducted principal axis factoring with an oblimin rotation.

A two-factor solution was suggested by the MAP test and parallel analysis (see Supplementary results Study 1). The loadings on the first factor were inversely related to the Negative Affectivity factor at 9 months and to the Negative Affectivity factor found by Vonderlin et al. (2012) (see Table 3), whereas the Surgency/ Extraversion factor was nearly identical to that at 9 months. Although the CFA model fit was poor, the two-factor model outperformed the three-factor model. Thus, the current analysis did not support the third factor, Self-regulation, found by Gartstein and Rothbart (2003).

Table 3

IBQ-R exploratory factor loadings at 9 and 12 months.

	9 months		12 months	
IBQ-R scale	Surgency/ Extraversion	Negative Affectivity	Negative Affectivity reversed	Surgency/ Extraversion
Activity Level	.22	.42	-.62	.28
Distress to Limitations		.60	-.77	
Fear		.24	-.22	
Sadness		.70	-.62	
Falling Reactivity		-.60	.48	
Cuddliness	.24	-.30	.46	
Duration of Orienting	.58			.38
Smiling and Laughter	.54		.28	.68
High Intensity Pleasure	.56			.53
Low Intensity Pleasure	.68	-.22	.45	.43
Soothability	.38	-.32	.33	
Perceptual Sensitivity	.53	.30	-.26	.52
Approach	.54			.46
Vocal Reactivity	.65			.52

Note: Loadings <.20 are not included.

2.2.2 Stability and developmental change

Due to the inconsistent factor structure, we only analyzed temperament stability on the scale level. Table 4 shows the Pearson correlations between the IBQ-R subscales measured at 9 and 12 months. All scales showed a moderate (r = .34 for Falling Reactivity) to high (r = .71 for Fear and Duration of Orienting) stability during this period of three months.

Table 4

Descriptive Statistics and Correlations of the IBQ-R scales at 9 and 12 months.

IBQ-R scale	9 months	12 months	
	M (*SD*)	*M* (*SD*)	*r*
Activity Level	3.86 (0.86)	3.98 (0.88)	.64**
Distress to Limitations	3.86 (0.87)	4.30 (0.78)	.57**
Fear	2.70 (1.09)	2.91 (1.07)	.71**
Duration of Orienting	2.92 (0.95)	3.15 (0.73)	.71**
Smiling and Laughter	4.55 (0.88)	4.68 (0.92)	.70**
High Pleasure	5.96 (0.55)	6.05 (0.47)	.58**
Low Pleasure	5.02 (0.76)	4.96 (0.74)	.44**
Soothability	5.16 (0.70)	5.27 (0.84)	.42**
Falling Reactivity	5.18 (0.93)	5.26 (0.87)	.34*
Cuddliness	5.49 (0.81)	5.34 (0.89)	.64**
Perceptual Sensitivity	4.23 (1.15)	4.43 (1.04)	.55**
Sadness	3.27 (0.97)	3.44 (0.96)	.69**
Approach	5.65 (0.69)	5.73 (0.65)	.49**
Vocal Reactivity	4.54 (0.82)	5.07 (0.76)	.44**

Note. N = 53. ** $p < .01$. * $p < .05$, all p-values are two-tailed.

To assess developmental changes, paired t-tests and Wilcoxon tests were computed. For descriptive purposes, means and standard deviations are shown in Table 4. Scores increased significantly for Distress to Limitations (9 months: M = 3.86, SD = 0.87; 12 months: M = 4.30, *SD* = 0.78; $p < .001$), Duration of Orienting (9 months: *M* = 2.92, *SD* = 0.95; 12 months: *M* = 3.15, *SD* = 0.73; $p < .05$), and Vocal Reactivity [t (52) = -4.60, $p < .001$].

2.3 Discussion

The results of Study 1 provide support for the stability of parental ratings in the IBQ-R over a period of three months. Since this period is relatively short, and other studies (e.g. Gartstein & Rothbart, 2003) investigated 9-12-month-olds as one group, the question arises whether comparable stability coefficients occur over a longer period of time. Increasing the age range would also allow us to investigate stability not only within one questionnaire but also across two different questionnaires. Descriptive statistics of the IBQ-R scales are similar to those reported for the German sample (Vonderlin et al., 2012) and the American sample (Gartstein & Rothbart, 2003).

The findings regarding the factor structure of infant temperament are only partially consistent with previous work. Whereas the three-factor solution was found for US-American (Gartstein & Rothbart, 2003), Italian (Montirosso et al., 2011), Portuguese (Costa & Figueiredo, 2018) and Russian (Gartstein et al., 2005) infants, the results of the factor analysis at 9 months confirmed the two-factor solution found for German infants (Mink et al., 2013; Vonderlin et al., 2012). At 12 months, the findings also suggested a two-factor solution, but the content-based comparisons of the second factor revealed some differences.

Study 2 was designed not only to investigate temperament stability over a longer period of time and across two different measurement tools, but also to examine the underlying factor structure of infant temperament in a different sample.

3 Study 2

3.1 Methods

3.1.1 Participants

Data were collected as part of an ongoing longitudinal study. The participating families were recruited from a database in which they had previously registered to participate in child development studies. The families came from one of two medium-sized cities in the Ruhr area of Germany. The parents received the questionnaires one week prior to the laboratory session at the university, and were asked to bring the completed questionnaires to the session.

The initial sample consisted of 147 twelve-month-olds (77 girls) with a mean age of 11 months 24 days (*SD* = 9 days; range 344-387). Ten additional children had to be excluded due to more than 20% missing values. All children were Caucasian. Most of the parents had either a university degree (mothers 60.5%, fathers 56.5%) or a university entrance-level diploma (mothers 17.7%, fathers 13.6%) as their highest educational attainment.

Questionnaires for the second laboratory appointment at 18 months were obtained from the families of 126 infants (66 girls) with a mean age of 17 months 27 days (*SD* = 10 days; range 530 – 572 days). Questionnaires for the third laboratory appointment at 24 months were obtained from the families of 126 infants (65 girls) with a mean age of 23 months 27 days (*SD* = 9 days; range 708 – 751). We analyzed internal consistencies and underlying factor structures with each of the previous respective samples at each time point. In order to analyze longitudinal stability and developmental changes, children were only included if temperament data were available at all three time points. Therefore, the final sample for longitudinal analyses consisted of 110 children (58 girls) who participated at 12, 18, and 24 months. The mean age was 11 months 24 days (*SD* = 8; range 347 - 381) at T1, 17 months 27 days (*SD* = 11; range 531 - 572) at T2, and 23 months 27 days (*SD* = 9; range 708 - 751) at T3.

3.1.2 Material

As in Study 1, at T1, the German version (Vonderlin et al., 2012) of the IBQ-R (Gartstein & Rothbart, 2003) was employed. For the sample of Study 2, Cronbach's αs for the subscales of the IBQ-R ranged between .76 and .88. At T2 and T3, we used the German translation (Kirchhoff et al., 2013) of the ECBQ (Putnam et al., 2006) to assess the toddlers' temperament. The ECBQ was designed to measure temperament in toddlers aged between 18 and 36 months. Similar to the IBQ-R, parents are asked to indicate on a 7-point Likert scale how often a specific behavior occurred in the past two weeks (1 = never, 7 = always). The ECBQ contains 201 items that are grouped into 18 scales. With the exception of the scale Activity Level, which scored below .60, the internal consistencies of the 18 scales of the German translation ranged from .61 to .87 (Cronbach's α, Kirchhoff et al., 2013). At 18 months, Cronbach's αs for the 18 scales in the current study ranged from .69 to .88. At 24 months, Cronbach's αs ranged from .68 to .88.

3.1.3 Coding and analysis

Coding and statistical analyses were similar to Study 1. Stability of infant temperament over the period of six and 12 months was assessed via the eleven similar scales of the two questionnaires. Stability of the additional seven ECBQ scales was also computed using Pearson product-moment correlations.

3.2 Results

3.2.1 Factor Structure

We first analyzed the underlying structure of the infants' and toddlers' temperament. Similar to Study 1, the results of the confirmatory factor analyses were not convincing, since the three-factor model (Gartstein & Rothbart, 2003) and the two-factor model (Vonderlin et al., 2012) did not match the data for the IBQ-R scales at 12 months (for detailed test statistics, see Supplementary results Study 2). Comparing the AIC = 309.43 of the three-factor model with the AIC = 294.10 of the two-factor model rather indicates the two over the three-factor solution. Due to the poor fit of both models, we again decided to additionally conduct an exploratory factor analyses. Prior to principal axis factoring, we determined the number of factors using parallel analysis and the MAP test (O'Conner, 2000). Whereas the MAP test revealed a two-factor solution, a three-factor solution emerged from parallel analysis of the IBQ-R scales at 12 months (see Supplementary results Study 2). Therefore, we conducted principal axis extraction analysis with an oblimin rotation with both possible numbers of factors and compared the derived factors to the US-American three-factor solution (Gartstein & Rothbart, 2003) and the German two-factor solution (Vonderlin et al., 2012). Loadings of the scales in the three-factor solution suggested by our sample did not correspond to the loadings of the scales of the American version. Loadings of the two-factor solution of the present sample were identical to the factor structure suggested by Vonderlin and colleagues (2012). Although the CFA model fit was poor, the two-factor model outperformed the three-factor model. Thus, the data from the present sample suggest a two-factor rather than a three-factor solution (see Table 5).

Table 5

IBQ-R exploratory factor loadings at 12 months.

IBQ-R scale	Surgency/ Extraversion	Negative Affectivity
Activity Level		.51
Distress to Limitations		.82
Fear		.31
Sadness		.70
Falling Reactivity		-.50

Cuddliness	.22	-.42
Duration of Orienting	.46	
Smiling and Laughter	.58	
High Intensity Pleasure	.60	
Low Intensity Pleasure	.54	-.30
Soothability	.38	
Perceptual Sensitivity	.55	.22
Approach	.60	
Vocal Reactivity	.70	

Note: Loadings <.20 are not included.

Since prior findings regarding the factor structure of the ECBQ were consistent, we only tested the suggested three-factor model by Putnam et al. (2006) in CFA. At both age point, the model fit was poor (18 months: $\chi^2(132, N = 126) = 336.03$, $p < .001$; CFI = .61; TLI = .55; RMSEA = .11; 24 months: $\chi^2(132, N = 126) = 301.92$, $p < .001$; CFI = .66; TLI = .60; RMSEA = .10). Since there were no other theoretical derivations, no other model was tested. To verify the factor structure, we again conducted principal axis factoring. At 18 and 24 months, a three-factor solution was derived from our sample for the ECBQ (see Table 6) using the MAP test and parallel analysis (see Supplementary results Study 2). The loadings of the scales on the factors is very similar to in the ECBQ (Putnam et al., 2006). The only difference found at 24 months was the primary loading of Perceptual Sensitivity. This scale primarily loaded positively on the factor Effortful Control, but still had secondary positive loadings on Negative Affectivity.

Table 6

ECBQ exploratory factor loadings at 18 and 24 months.

ECBQ scale	18 months			24 months		
	Negative Affectivity	Surgency/ Extraversion	Effortful Control	Negative Affectivity	Surgency/ Extraversion	Effortful Control
Activity Level		.66	-.27		.72	
High Pleasure		.64			.64	
Impulsivity		.48			.64	
Positive Anticipation	.23	.40		.21	.50	
Sociability		.49	.21		.40	.34
Attentional Focusing			.56			.41
Attentional Shifting		.31	.69			.65
Cuddliness			.37		-.25	.32
Inhibitory Control		-.29	.53		-.30	.58
Low Pleasure			.53			.72
Perceptual Sensitivity	.44		.38	.36	.22	.44
Discomfort	.60			.59		
Fear	.68	-.27		.76	-.27	
Frustration	.59	.25		.60	.36	
Motor Activation	.25			.31	.26	
Shyness	.56	-.42		.50	-.40	
Soothability	-.47		.20	-.35		.26

Note: Loadings <.20 are not included.

3.2.2 Stability and developmental changes

Stability coefficients are presented in Table 7. With the exception of the corresponding scales Approach (IBQ-R) and Positive Anticipation (ECBQ), all correlations for the six-month span from 12 to 18 months were statistically significant. The lowest stability coefficient was found for Sadness ($r = .24$, $p < .05$). Correlations between ECBQ scales from 18 to 24 months were all above $r = .50$. These findings indicate a high normative stability of parental temperament ratings in this six-month span. With the exception of Sadness, the corresponding scales of the IBQ-R and ECBQ showed stability over the one-year period from 12 to 24 months.

Since factor analyses of this sample revealed a more consistent factor structure of infants' and toddlers' temperament, we additionally explored stability on the factor level. The factors Surgency/Extraversion and Negative Affectivity were found at all three age time points. With one exception, all cross-age correlations for the aforementioned factors were significant and moderate to large in magnitude. We did not find stability of Surgency/ Extraversion over the period of one year, but the six-month spans showed significant correlations ($r = .33$, $p < .01$ for 12 to 18 months; $r = .72$, $p < .01$ for 18 to 24 months). Stability coefficients for Negative Affectivity were significant at $p < .01$, with $r = .46$ from 12 to 18 months, $r = .69$ from 18 to 24 months and $r = .41$ from 12 to 24 months. The regulation factor Effortful Control, which emerged at 18 and 24 months, showed significant stability ($r = .64$, $p < .01$). No comparable factor was found at 12 months.

To assess developmental changes, repeated measures ANOVAs were computed. For descriptive purposes, Table 7 shows means and standard deviations for all three age time points. We first compared the eleven corresponding scales at all three ages. Significant age effects were observed for all scales apart from Low Pleasure (see Table 7). We subsequently analyzed significant age effects using post-hoc tests. All post-hoc tests were Bonferroni-adjusted and significant at $p < .05$.

Paired *t*-tests or Wilcoxon tests were computed for the remaining seven ECBQ scales, depending on whether the data were normally distributed. Significant age effects are shown in Table 7 (for details, see Supplementary results Study 2).

Table 7

Descriptive Statistics, Correlations and significant age effects of the IBQ-R and ECBQ scales at 12, 18, and 24 months.

IBQ-R/ECBQ scale	12 months	18 months	24 months	12-18	18-24	12-24	ANOVA	
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*r*	*r*	*r*	*F*	*df*
Activity Level/ Activity Level	3.84[c, d] (0.79)	4.71[c] (0.79)	4.72[d] (0.76)	.38**	.69**	.27**	82.19***	1.662, 181.182
Distress to Limitations/ Frustration	4.23[c, d] (0.84)	3.64[c] (0.94)	3.71[d] (0.79)	.32**	.61**	.27**	25.6***	1.803, 196.579
Fear/ Fear	2.79[c, d] (0.91)	2.05[c] (0.64)	2.30[d] (0.70)	.37**	.58**	.38**	44.6***	1.689, 184.078
Duration of Orienting/ Attentional Focusing	3.06[c, d] (0.93)	4.20[c, e] (0.91)	4.46[d, e] (0.82)	.36**	.66**	.38**	142.93***	1.721, 187.573
High Pleasure/ High Pleasure	6.06[c, d] (0.62)	5.12[c] (0.78)	5.11[d] (0.81)	.31**	.56**	.34**	99.69***	2, 218
Low Pleasure/ Low Pleasure	4.87 (1.00)	4.99 (0.81)	5.03 (0.70)	.46**	.57**	.45**	1.95	1.780, 194.001
Soothability/ Soothabilty	5.28[c, d] (0.85)	5.91[c] (0.74)	5.81[d] (0.69)	.33**	.61**	.36**	37.3***	1.712, 186.717
Cuddliness/ Cuddliness	5.31[c] (0.78)	5.04[c] (0.85)	5.13 (0.75)	.45**	.50**	.45**	6.12**	2, 218
Perceptual Sensitivity/ Perceptual Sensitivity	4.46 (1.12)	4.42[c] (1.05)	4.67[c] (0.92)	.51**	.67**	.42**	3.95*	1.773, 193.224
Sadness/ Sadness	3.25[c, d] (0.86)	2.76[c] (0.83)	2.83[d] (0.80)	.24*	.68**	.17	16.94***	1.608, 175.234
Approach/ Positive Anticipation	5.85[c, d] (0.67)	4.82[c, e] (0.99)	5.30[d, e] (0.81)	.17	.55**	.39**	67.21***	1.764, 192.254
Falling Reactivity [a]	5.10 (0.99)							
Vocal Reactivity [a]	4.92 (1.00)							
Smiling and Laughter [a]	4.67 (0.86)							
Attentional Shifting [b]		4.47 (0.63)	4.54 (0.66)		.55**			
Discomfort [b]		2.31[c] (0.88)	2.63[c] (0.80)		.50**			
Impulsivity [b]		5.08 (0.75)	5.04 (0.70)		.55**			
Inhibitory Control [b]		3.49[c] (0.95)	3.84[c] (0.95)		.64**			
Motor Activation [b]		2.35 (0.83)	2.27 (0.75)		.69**			
Shyness [b]		3.13[c] (0.85)	3.36[c] (0.93)		.52**			
Sociability [b]		5.31 (0.94)	5.45 (1.00)		.54**			

Note. $N = 110$. *** $p < .001$; ** $p < .01$; * $p < .05$, all p-values are two-tailed. [a] Scales are only included in IBQ-R; there are no equivalents in ECBQ. [b] Scales are only included in ECBQ; there are no equivalents in IBQ-R. With the exception of High Pleasure and Cuddliness, all significant repeated measures ANOVAs are Greenhouse-Geisser corrected. Means sharing common superscripts ([c, d, e]) are significantly different.

3.3 Discussion

The results of Study 2 demonstrated stability of temperament ratings between infancy and toddlerhood and within toddlerhood. Ten out of eleven correlation coefficients of the corresponding scales between the IBQ-R and ECBQ for the six-month span ranged between .24 and .51, indicating moderate to high stability of temperament ratings between infancy and toddlerhood. Only the scale Approach/ Positive Anticipation failed to reach significance (r = .17). The correlation coefficients between 12 and 24 months likewise suggest temperament stability between the IBQ-R and ECBQ. With the exception of the scale Sadness (r = .17), the correlations for the other ten corresponding scales ranged between .27 and .45, indicating a moderate stability of temperament ratings. Apart from the aforementioned scales, the correlation coefficients for the 12-month span are comparable to the coefficients for the six-month span, indicating that temperament is relatively stable even over the longer period of one year. The correlations over the six-month span from 18 to 24 months were all large in magnitude, indicating a high stability of temperament ratings within the ECBQ.

Stability of the factor Surgency/ Extraversion was not present over the period of one year from 12 to 24 months, but was found for the six-month span. This finding is only partially consistent with pervious work, since the two longitudinal studies focusing on stability between the IBQ-R and ECBQ did find stability of Surgency (Casalin et al., 2012; Putnam et al., 2008). This difference might be explained by methodological reasons. Whereas our sample differentiated between 18- and 24-month-olds, the sample of Putnam et al. (2008) comprised an approximately equal distribution of toddlers aged between 18 and 32 months, and Casalin et al. (2012) analyzed one age group of toddlers across an interval of six months. The correlation coefficient of Surgency within toddlerhood was large in magnitude (r = .72), suggesting a significant correlation and therefore stability between infancy and toddlerhood if 18- and 24-

month-olds had been treated as one group. Given the stability of the factor over six months, the results of our study do suggest stability of Surgency from infancy to toddlerhood. Stability of Negative Affectivity was found from infancy to both ages in toddlerhood and also within toddlerhood. These findings are consistent with the results of the two aforementioned longitudinal studies. While no regulation factor was found at 12 months, Effortful Control did show high stability within toddlerhood. Thus, the results of Study 2 provide support for the stability of temperament on the scale and factor level between different developmental periods and different measures as well as within the ECBQ in toddlerhood.

As the factor structure of the IBQ-R was analogous to that of infant temperament in Study 1, and again confirmed the two-factor solution in German infants (Mink et al., 2013; Vonderlin et al., 2012), below, we discuss possible reasons and explanations for both studies together.

4 General Discussion

The present study evaluated the factor structure and longitudinal stability of temperament from infancy to toddlerhood within the theoretical framework and the conceptual and operational definitions of the psychobiological approach to temperament (Rothbart, 1981).

The first aim of this study was to test the reliability of the German versions of the IBQ-R (Vonderlin et al., 2012) and the ECBQ (Kirchhoff et al., 2013) on the scale and factor level with a special focus on the factor structure of the IBQ-R. The results of both studies provide support for the reliability of both questionnaires on the scale level. The Cronbach's αs for the 14 scales of the IBQ-R in Study 1 and Study 2 were all above .60, indicating adequate internal consistencies. Whereas the German version of the IBQ-R (Vonderlin et al., 2012) was validated in 7-9-month-olds, the present study demonstrates reliability of the questionnaire across 9- and

12-month-olds and in two different samples. Overall, the internal consistencies and the average descriptive scores of the 14 scales in Study 1 and Study 2 are comparable to those reported for the original sample of the German IBQ-R (Vonderlin et al., 2012) as well as for 12-month-old German infants (Zmyj et al., 2017). The Cronbach's αs for the 18 scales of the ECBQ were also all above .60, with just two scales scoring lower than .70, thus indicating high internal consistencies of the translated ECBQ scales. In the German translation by Kirchhoff et al. (2013), the scale Activity Level was removed as its internal consistency was below .60. The present finding of a Cronbach's α of .75 at 18 and 24 months indicates an adequate internal consistency and a homogenous scale. Thus, we suggest retaining this scale in the German ECBQ. The internal consistencies of the other scales were similar to those of the translation (Kirchhoff et al., 2013). Overall, the average descriptive scores of the 18 scales at 18 and 24 months are comparable to those reported by Kirchhoff et al. (2013). As such, the present study provides support for the reliability of the German versions of the IBQ-R and ECBQ on the scale level and therefore for its application to investigate temperament using parental ratings in German infants and toddlers.

Due to discrepancies in the factor structure of the IBQ-R found in previous research (see Bosque Enlow et al., 2016), we were especially interested in its factor structure. The results of this study confirmed the two-factor solution of infant temperament reported in previous studies with German samples (Mink et al., 2013; Vonderlin et al., 2012). All German samples failed to replicate the Orienting/ Regulation factor. In the original factor structure of the IBQ-R (Gartstein & Rothbart, 2003), the Orienting/ Regulation factor contained primary loadings on the scales Low Pleasure, Cuddliness, Duration of Orienting, and Soothability. With the exception of the scale Duration of Orienting, these scales loaded on Surgency/Extraversion as well as on Negative

Affectivity in our study. Vonderlin et al. (2012) found comparable double loadings. Other studies, which confirmed the three-factor solution, found notable secondary loadings of the scales of the regulation factor on the other two factors (Gartstein et al., 2016; Montirosso et al., 2011). In addition, intercorrelations between the Orienting/ Regulation factor and the other two factors emerged in previous research that confirmed the three-factor solution, whereas Surgency and Negative Affectivity were found to be relatively independent factors (Gartstein & Rothbart, 2003; Putnam et al., 2001). Thus, it is questionable whether the regulation factor represents a third autonomous factor within the first year of life that can be detected independently of the affectivity factors, especially given that self-regulatory capacities develop at the end of the first year of life and in the transition to the second (Rothbart, 1981). Evans and Rothbart (2009) even suggested a higher-order two-factor model of temperament against the background of robust convergence between the three-factor model of childhood temperament and the Big Five within adult temperament. In line with the results of our factor analyses, the authors suggested the two factors Extraversion/ Positive Emotionality and Negative Affect, which are found to be consistent over the life span (Evans & Rothbart, 2009).

Another possible explanation for the differences in factor structure lies in cultural differences. While the IBQ-R was developed in the US, the perception and occurrence of everyday situations as detailed in its 191 items might differ between countries (see Mink et al., 2013). Differences in cultural values even between two western oriented cultures (Germany and the U.S.) do occur. Regarding most of the dimensions in Hofstede’s cultural value system, Germany and the U.S. were found to be very similar. However, differences occurred for example on the dimension Individualism/ Collectivism, with Germans being less individualistic (Hofstede, 2001). These differences in cultural values result in differences in temperament description and parenting behaviors. In observations of mother-child-interactions of German and American mother-infant dyads by Kirchhoff (2015), mothers from the U.S. showed higher levels of intensity. They were more stimulating regarding the number of toys used, the noise level and the level of exuberance. Furthermore, they expressed more enthusiasm and positive emotions. Whereas mothers from Germany confirmed to focus on rearing their children as being empathic and helpful to others, U.S. American mothers highlighted positive emotionality with a friendly and easy attitude (Kirchhoff, 2015). Due to these differences in parenting, some of situations described in the items might occur more often and are perceived differently in Germany than in the U.S., for example the items “When tossed around playfully how often did the baby smile?” and “When tossed around playfully how often did the baby laugh?”. Since U.S. American mothers are show more exuberance and highlight positive emotionality, these situations might occur more often in the U.S. than in Germany. Thus, these cultural differences in parenting might also result in different factor structures. Parents engage in caregiving practices that enhance characteristics in line with their respective cultural values and those of their cultural group, leading to cultural differences in the development of temperament (see also Bosque Enlow et al., 2016; Gartstein et al., 2006). Finally, parents might evaluate their infants’ temperament characteristics against the background of their culture-specific standards, including aspects such as social desirability (Mink et al., 2013). Several studies (e.g. Desmarais et al., 2019; Dragan et al., 2011; Gartstein et al., 2016; Sung et al., 2015) reporting mean level differences in temperament ratings between different cultures support these arguments of cultural differences as a possible explanation for differences in temperament ratings, also for German and American infants (Kirchhoff, 2015).

Whereas the reliability of the three-factor structure of infant temperament in the first year of life is questionable, the results regarding the factor structure within toddlerhood are consistent with previous studies. The three-factor solution with Surgency/ Extraversion, Negative Affectivity, and Effortful Control is analogous to that of the original ECBQ (Putnam et al., 2006) and of the German translation (Kirchhoff et al., 2013). In turn, these three factors are analogous to models of adult personality with the constructs Extraversion, Neuroticism, and Constraint/ Conscientiousness (see Putnam et al., 2001). Thus, the results of this study provide support for the reliability and consistency of the three-factor structure of toddlers' temperament.

The second aim of the present study was to test the stability of infants' and toddlers' temperament. As expected, stability occurred within the IBQ-R and within the ECBQ as well as between the corresponding scales of the IBQ-R and ECBQ. These results are in line with previous research on stability of temperament in German infants (Mink et al., 2013). Moreover, these findings of stability can be integrated into the psychobiological approach of temperament (Rothbart, 1981), suggesting that temperament traits, especially the reactivity components, are relatively stable very early in life (Rothbart & Derryberry, 1981). The correlation coefficients for all 18 scales of the ECBQ indicate a high stability of temperament within the first six months of toddlerhood. The values are comparable to those obtained over similar time spans in previous studies on temperament (e.g., Putnam et al., 2006).

A major advantage of the present study is that stability was tested not only within infancy and within toddlerhood, but also across these two developmental periods. The only subscales in which no stability in parental ratings from infancy to toddlerhood was found were Approach/ Positive Anticipation between 12 and 18 months and Sadness between 12 and 24 months. The lack of stability of these two scales may be related to increasing regulatory capacities, including emotion regulation (Rothbart et al., 2011). Since self-regulation emerges around the first year of life and increases in the second year of life, the expression of excitement and positive emotions may be more controlled in toddlers than in infants. This possible explanation is supported by the lack of stability of Surgency over the period of one year in Study 2. Casalin et al. (2012) found a similar lack of absolute stability, with decreasing levels of Surgency/ Extraversion between infancy and toddlerhood. In line with our results, Putnam et al. (2006) found Positive Anticipation to score the lowest stability of all 18 ECBQ scales. Nevertheless, a moderate stability within one year was found for Approach/ Positive Anticipation. Thus, the lack of stability of Approach between 12 and 18 months does not indicate that the underlying temperament structure has changed, but rather that the expression of this temperament trait changes in adjacent developmental periods.

In addition, at around the age of 12 months, infants gain the ability to downregulate emotions, especially negative emotions such as sadness, more effectively (Planalp et al., 2017). Accordingly, toddlers might be better able to control their expression of sadness compared to infants, thus explaining the lower stability of this subscale. Previous research has also found comparable low levels of stability of sadness (Putnam et al., 2008). For all other subscales, stability of temperament ratings between infancy and toddlerhood over six and 12 months was found, which is in line with earlier findings (Putnam et al., 2008).

With the exception of the aforementioned lack of stability of the Surgency factor, all cross-age correlations of the factors suggest temperament stability on the factor level. This finding replicates previous reports of high stability of Negative Affectivity and Effortful Control (Casalin et al., 2012; Putnam et al., 2008). Lower levels of stability of Surgency between the IBQ-R and ECBQ compared to the other two factors were also found in a Belgian (Casalin et al.,

2012) and a Japanese sample (Nakagawa & Sukigara, 2014), indicating that the expression of positive emotions decreases from infancy to toddlerhood. In the second half of the first year of life, infants gain more capacities to explore the world, and typically show their excitement about this. In toddlerhood, this expression of excitement decreases, explaining the lower levels of stability in Surgency/ Extraversion.

As expected within the psychobiological approach of temperament (Rothbart, 1981), there were significant changes in average ratings within infancy and toddlerhood, and especially between infancy and toddlerhood. The developmental changes found in infancy are consistent with previous research. At 12 months, temperament ratings were higher for Distress to Limitations, Duration of Orienting, and Vocal Reactivity. Higher scores for Distress to Limitations have been reported previously (Carranza et al., 2000; Gartstein & Rothbart, 2003; Mink et al., 2013). Increases in Distress to Limitations may be related to increased capacities for locomotion combined with more parental limitations on exploratory behavior (Gartstein & Rothbart, 2003). The small number of age differences within infancy found in the present study compared to some previous research (Gartstein & Rothbart, 2003; Mink et al., 2013) could be related to the short time interval of three months assessed in this study. Mink et al. (2013) measured differences between 6- and 12-month-olds, and Gartstein & Rothbart (2003) even combined 9- and 12-month olds into one group, which they compared with other age points.

Within toddlerhood, significant increases emerged for Attentional Focusing, Perceptual Sensitivity, Positive Anticipation, Shyness, Discomfort, Fear, and Inhibitory Control. Putnam et al. (2006) reported similar increases within the second and third year of life.

Age-related increases between infancy and toddlerhood were found for Activity Level, Duration of Orienting/ Attentional Focusing and Soothability, whereas decreases emerged for

Distress to Limitations, High Pleasure, Sadness, Cuddliness, Approach/ Positive Anticipation, and Fear. No changes were observed for Low Pleasure and Perceptual Sensitivity. To the best of our knowledge, only one previous study has examined longitudinal stability of the IBQ-R and ECBQ on the scale level (Puntam et al., 2008). Since the authors did not analyze mean level differences of average temperament ratings, our study is the first to compare developmental changes in temperament ratings from infancy to toddlerhood using the IBQ-R and ECBQ. The results of our study suggest that considerable developmental changes do occur more between infancy and toddlerhood than within each of the two developmental phases. With the exception of the scale Approach/ Positive Anticipation, all decreasing scales are related to Surgency/ Extraversion and Negative Affectivity. This decrease in positive emotions is in line with previous research (Casalin et al., 2012).

The present findings need to be interpreted in the context of some limitations. First, one aim of this study was to validate the German translation of the ECBQ (Kirchhoff et al., 2013). Although the results suggest reliability of the German version, our sample only included toddlers aged from 18 to 24 months. Since the ECBQ was developed to assess toddlers' temperament between 18 and 36 months, further validation should include the whole age range in order to enable reliable statements for the third year of life. Second, the sample size of Study 1, especially at 12 months, was possibly too small to obtain trustworthy results using exploratory factor analysis (EFA). The sample size in an EFA is essential for the interpretation of the results, and small sample sizes are a common problem (Jung & Lee, 2011). Since McNeish (2017) stated that an EFA with less than 100 observations and missing data does not produce trustworthy results, the sample size of $N = 60$ at T2 is presumably too small for a meaningful interpretation of the findings. Thus, special caution is warranted when interpreting the factor structure at 12 months in

Study 1. Since the number of items (191) is relatively high for this small sample size, the inverse loadings of the scales on Negative Affectivity and the inconsistent factor may be due to a statistical problem rather than representing a different factor. Furthermore, to obtain a clearer understanding of the factor structure of infant temperament, especially in Germany, a gradual analysis in a larger sample is necessary. The IBQ-R should be applied to different age groups across infancy to investigate when the Orienting/ Regulation factor emerges and whether it already exists in infancy. Third, the present study assessed temperament only via parental report. Since parental expectations might influence their ratings, and the overt manifestation of some temperament traits might be more discreet (Putnam et al., 2008), stability of infant and toddler temperament should additionally be assessed using behavioral measures as well. Nevertheless, it is important to note that satisfactory convergence between parental ratings and behavioral measures has been previously reported (e.g. Carranza Carnicero et al., 2000; Vroman et al., 2014).

Finally, a major advantage of this study is that we investigated toddlers at specific ages rather than examining a sample comprising the whole age range of toddlerhood together. This allows us to draw a more precise picture of the developmental changes from infancy to toddlerhood. Since lower levels of stability were observed across longer developmental periods (Putnam et al., 2008), and we assessed temperament only within the first two years of life, future research might include toddlers older than 24 months in a longitudinal investigation in order to ascertain between which precise ages changes occur. A more profound understanding of the developmental changes of temperament between infancy and toddlerhood might in turn allow us to identify possible environmental influences and maturation processes responsible for these changes.

In sum, the results of the present study provide support for the reliability of the German versions of the IBQ-R and the ECBQ on the scale level. Thus, we support the application of the German versions of the IBQ-R and ECBQ in research on temperament in infants and toddlers. As our findings regarding the factor structure of temperament in infancy add to the number of previously reported inconsistent results, suggesting that it might not be possible to reliably determine the structure of infant temperament in the first year of life, we recommend analyzing the results of the IBQ-R on the scale level. It is questionable whether the regulation factor is detectable independently of the affectivity factors within the first 12 months. Thus, it will only be possible to reliably interpret results on the scale level once further validation of the factor structure of the IBQ-R has been achieved. The three-factor structure found within toddlerhood was reliable and comparable to that of adults. Furthermore, the results of this study provide support for the stability of temperament ratings on the scale and factor level within the developmental periods of infancy and toddlerhood, as well as between these developmental periods and the corresponding measures.

5 References

Abulizi, X., Pryor, L., Michel, G., Melchior, M., van der Waerden, J., & EDEN Mother-Child Cohort Study Group. (2017). Temperament in infancy and behavioral and emotional problems at age 5.5: The EDEN mother-child cohort. *PLoS ONE, 12*(2), Article e0171971. https://doi.org/10.1371/journal.pone.0171971

Bornstein, M. H., Arterberry, M. E., & Lamb, M. E. (2013). *Development in infancy: A contemporary introduction* (5th ed.). Psychology Press. https://doi.org/10.4324/9780203758472

Bornstein, M. H., Putnick, D. L., Gartstein, M. A., Hahn, C. S., Auestad, N., & O'Connor, D. L. (2015). Infant temperament: Stability by age, gender, birth order, term status, and socioeconomic status. *Child Development*, *86*(3), 844-863.

Carranza Carnicero, J. A., González-Salinas, C., & Ato, E. (2013). A longitudinal study of temperament continuity through IBQ, TBAQ and CBQ. *Infant Behavior and Development*, *36*(4), 749-761. https://doi.org/10.1016/j.infbeh.2013.08.002

Carranza Carnicero, J. A., Pérez-López, J., Del Carmen González Salinas, M., & Martínez-Fuentes, M. T. (2000). A longitudinal study of temperament in infancy: Stability and convergence of measures. *European Journal of Personality*, *14*(1), 21-37. https://doi.org/10.1002/(SICI)1099-0984(200001/02)14:1<21::AID-PER367>3.0.CO;2-A

Casalin, S., Luyten, P., Vliegen, N., & Meurs, P. (2012). The structure and stability of temperament from infancy to toddlerhood: A one-year prospective study. *Infant Behavior and Development*, *35*(1), 94-108. https://doi.org/10.1016/j.infbeh.2011.08.004

Caspi, A., & Shiner, R. L. (2006). Personality Development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 300–365). John Wiley & Sons, Inc..

Costa, R., & Figueiredo, B. (2018). Infant behaviour questionnaire–revised version: a psychometric study in a Portuguese sample. *Journal of Reproductive and Infant Psychology*, *36*(2), 207-218. https://doi.org/10.1080/02646838.2018.1436752

Desmarais, E., Majdandžić, M., Gartstein, M. A., Bridgett, D. J., & French, B. F. (2019). Cross-cultural differences in temperament: comparing paternal ratings of US and Dutch infants. *European Journal of Developmental Psychology*, *16*(2), 137-151. https://doi.org/10.1080/17405629.2017.1356713

Dias, C. C., Costa, R., Pinto, T. M., & Figueiredo, B. (2021). The Infant Behavior Questionnaire-Revised: psychometric properties at 2 weeks, 3, 6 and 12 months of life. *Early Human Development*, *153*, 105290. https://doi.org/10.1016/j.earlhumdev.2020.105290

Dragan, W. Ł., Kmita, G., & Fronczyk, K. (2011). Psychometric properties of the polish adaptation of the Infant Behavior Questionnaire—revised (IBQ-R). *International Journal of Behavioral Development*, *35*(6), 542-549. https://doi.org/10.1177/0165025411422181

Evans, D. E., & Rothbart, M. K. (2009). A two-factor model of temperament. *Personality and Individual Differences*, *47*(6), 565-570. https://doi.org/10.1016/j.paid.2009.05.010

Gartstein, M. A., Bogale, W., & Meehan, C. L. (2016). Adaptation of the Infant Behavior Questionnaire-Revised for use in Ethiopia: Expanding cross-cultural investigation of temperament development. *Infant Behavior and Development*, *45*(Pt. A), 51-63. https://doi.org/10.1016/j.infbeh.2016.09.004

Gartstein, M. A., Gonzalez, C., Carranza, J. A., Ahadi, S. A., Ye, R., Rothbart, M. K., et al. (2006). Studying cross-cultural differences in the development of infant temperament: People's Republic of China, the United States of America, and Spain. *Child Psychiatry and Human Development, 37,* 145–161. https://doi.org/10.1007/s10578-006-0025-6

Gartstein, M. A., Knyazev, G. G., & Slobodskaya, H. R. (2005). Cross-cultural differences in the structure of infant temperament: United States of America (US) and Russia. *Infant Behavior and Development*, *28*(1), 54-61. https://doi.org/10.1016/j.infbeh.2004.09.003

Gartstein, M. A., Putnick, D. L., Kwak, K., Hahn, C. S., & Bornstein, M. H. (2015). Stability of temperament in South Korean infants from 6 to 12 to 18 months: Moderation by age, sex, and birth order. *Infant Behavior and Development*, *40,* 103-107. https://doi.org/10.1016/j.infbeh.2015.05.002

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behavior and Development*, *26*(1), 64-86. https://doi.org/10.1016/S0163-6383(02)00169-8

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., ... & McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child Development*, *58*(2), 505-529. https://doi.org/10.2307/1130527

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations* (2nd ed.). Sage.

Jung, S., & Lee, S. (2011). Exploratory factor analysis for small samples. *Behavior Research Methods*, *43*(3), 701–709. https://doi.org/10.3758/s13428-011-0077-9

Kiel, N., Bruckdorfer, R., Petermann, F., & Reinelt, T. (2018). Temperament in der frühen Kindheit und die Entwicklung externalisierender Störungen: Implikationen für die klinische Diagnostik. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*. https://doi.org/10.1024/1661-4747/a000354

Kirchhoff, C., Gartstein, M., Iverson, S., Reed-Smith, H., Brekke, A., & Cress, E. (2013). *Temperament and behavioral difficulties across cultures: A comparison between German and US toddlers*. Paper presented at the Occasional Temperament Conference, Salt Lake City, UT, USA.

Kirchhoff, M. C. (2015). *Temperament and behavioral difficulties across cultures: A comparison between German and U.S. toddlers applying the concept of the 'developmental niche'* (Order No. 10043063) [Doctoral dissertation, Washington State Universtiy]. Publicly Available Content Database.

Klein-Radukic, S., & Zmyj, N. (2015). Dropout in looking time studies: The role of infants' temperament and cognitive developmental status. *Infant Behavior and Development*, *41*, 142-153. https://doi.org/10.1016/j.infbeh.2015.10.001

Klein-Radukic, S., & Zmyj, N. (2020). No evidence for infants' contingency detection as a precursor of toddlers' self-recognition. *Infant Behavior and Development*, *60*, 101459. https://doi.org/10.1016/j.infbeh.2020.101459

Komsi, N., Räikkönen, K., Heinonen, K., Pesonen, A. K., Keskivaara, P., Järvenpää, A. L., & Strandberg, T. E. (2008). Continuity of father-rated temperament from infancy to middle childhood. *Infant Behavior and Development*, *31*(2), 239-254. https://doi.org/10.1016/j.infbeh.2007.10.002

Komsi, N., Räikkönen, K., Pesonen, A. K., Heinonen, K., Keskivaara, P., Järvenpää, A. L., & Strandberg, T. E. (2006). Continuity of temperament from infancy to middle childhood. *Infant Behavior and Development*, *29*(4), 494-508. https://doi.org/10.1016/j.infbeh.2006.05.002

Kozlova, E. A., Slobodskaya, H. R., & Gartstein, M. A. (2020). Early temperament as a predictor of child mental health. *International Journal of Mental Health and Addiction*, *18*(6), 1493-1506. https://doi.org/10.1007/s11469-019-00181-3

Kusanagi, E., Nakano, S., & Kondo-Ikemura, K. (2014). The development of infant temperament and its relationship with maternal temperament. *Psychologia*, *57*(1), 31-48. https://doi.org/10.2117/psysoc.2014.31

Lemery, K. S., Goldsmith, H. H., Klinnert, M. D., & Mrazek, D. A. (1999). Developmental models of infant and childhood temperament. *Developmental Psychology*, *35*(1), 189-204. https://doi:10.1037/0012-1649.35.1.189

McNeish, D. (2017). Exploratory factor analysis with small samples and missing data. *Journal of Personality Assessment*, *99*(6), 637–652. https://doi.org/10.1080/00223891.2016.1252382

Mink, D., Henning, A., & Aschersleben, G. (2013). Visual habituation tasks: The role of infant temperament. *Infant Behavior and Development*, *36*(3), 377-390. https://doi.org/10.1016/j.infbeh.2013.03.005

Montirosso, R., Cozzi, P., Putnam, S. P., Gartstein, M. A., & Borgatti, R. (2011). Studying cross-cultural differences in temperament in the first year of life: United States and Italy. *International Journal of Behavioral Development*, *35*(1), 27-37. https://doi.org/10.1177/0165025410368944

Nakagawa, A., & Sukigara, M. (2014). The effects of soothing techniques and rough-and-tumble play on the early development of temperament: A longitudinal study of infants. *Child Development Research*, *Vol. 2014*, Article ID 741373. http://dx.doi.org/10.1155/2014/741373

Olafsen, K. S., Ulvund, S. E., Torgersen, A. M., Wentzel-Larsen, T., Smith, L., & Moe, V. (2018). Temperamental adaptability, persistence, and regularity: parental ratings of

Norwegian infants aged 6 to 12 months, with some implications for preventive practice. *Infant mental health journal*, *39*(2), 183-197. https://doi.org/10.1002/imhj.21697

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396-402. https://doi.org/10.3758/BF03200807

Pauli-Pott, U., Mertesacker, B., & Beckmann, D. (2003). Ein Fragebogen zur Erfassung des «frühkindlichen Temperaments» im Elternurteil. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, *31*(2), 99-110. https://doi.org/10.1024/1422-4917.31.2.99

Planalp, E. M., Van Hulle, C., Gagne, J. R., & Goldsmith, H. H. (2017). The infant version of the Laboratory Temperament Assessment Battery (Lab-TAB): Measurement properties and implications for concepts of temperament. *Frontiers in Psychology*, *8*, Article 846. https://doi.org/10.3389/fpsyg.2017.00846

Planalp, E. M., & Goldsmith, H. H. (2020). Observed profiles of infant temperament: Stability, heritability, and associations with parenting. *Child development*, *91*(3), e563-e580. https://doi.org/10.1111/cdev.13277

Putnam, S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In A. Eliasz & A. Angleitner (Eds.), *Advances in research on temperament* (pp. 165–182). PabstScience Publishers.

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development*, *29*(3), 386-401. https://doi.org/10.1016/j.infbeh.2006.01.004

Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and Child Development: An International Journal of Research and Practice*, *17*(4), 387-405. https://doi.org/10.1002/icd.582

Riese, M. L. (1987). Temperament stability between the neonatal period and 24 months. *Developmental Psychology*, *23*(2), 216-222. https://doi.org/10.1037/0012-1649.23.2.216

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3-25. https://doi.org/10.1037//0033-2909.126.1.3

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52*(2), 569-578. https://doi.org/10.2307/1129176

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (6th ed., Vol.3, pp. 99–166). John Wiley & Sons, Inc..

Rothbart, M. K. & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds.), *Advances in Developmental Psychology* (Vol. 1, pp. 37–86). Erlbaum.

Rothbart, M. K., Derryberry, D., & Hershey, K. (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and Personality Development Across the Life Span* (pp. 85–119). Erlbaum.

Rothbart, M. K., Derryberry, D. & Posner, M. I. (1994). A psychobiological approach to the development of temperament. In J. E. Bates & T. D. Wachs (Eds.), *Temperament:*

Individual differences at the interface of biology and behavior (pp. 83–116). American Psychological Association. https://doi.org/10.1037/10149-003

Rothbart, M. K., Sheese, B. E., Rueda, M. R., & Posner, M. I. (2011). Developing mechanisms of self-regulation in early life. *Emotion Review*, *3*(2), 207-213. https://doi.org/10.1177/1754073910387943

Slobodskaya, H. R., & Kozlova, E. A. (2016). Early temperament as a predictor of later personality. *Personality and individual differences*, *99*, 127-132. https://doi.org/10.1016/j.paid.2016.04.094

Sung, J., Beijers, R., Gartstein, M. A., de Weerth, C., & Putnam, S. P. (2015). Exploring temperamental differences in infants from the USA and the Netherlands. *European Journal of Developmental Psychology*, *12*(1), 15-28. https://doi.org/10.1080/17405629.2014.937700

Vonderlin, E., Ropeter, A., & Pauen, S. (2012). Assessment of temperament with the Infant Behavior Questionnaire Revised (IBQ-R)-the psychometric properties of a German version. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, *40*(5), 307-314. https://doi.org/10.1024/1422-4917/a000187

Vroman, L. N., Lo, S. L., & Durbin, C. E. (2014). Structure and convergent validity of children's temperament traits as assessed by experimenter ratings of child behavior. *Journal of Research in Personality*, *52*, 6-12. https://doi.org/10.1016/j.jrp.2014.06.002

Zentner, M., & Bates, J. E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. *International Journal of Developmental Science*, *2*(1-2), 7-37. https://doi.org/10.3233/DEV-2008-21203

Zmyj, N., Schölmerich, A., & Daum, M. M. (2017). The relationship between attention and deferred imitation in 12-month-old infants. *Infant Behavior and Development*, *48*(Pt. B), 175-183. https://doi.org/10.1016/j.infbeh.2017.02.001

Supplementary Materials

1. Supplementary methods

1.1 Parallel analysis

Parallel analysis provides the eigenvalues from a factor analysis of randomly permuted datasets with the same number of variables and participants. The eigenvalues of the random permutations from the 95th percentile are then compared with the real data. Any factor from the real data set with eigenvalues exceeding the randomly generated values are considered relevant (O'Connor, 2000).

1.2 MAP test

The MAP tests includes a complete principal components analysis followed by the examination of a series of matrices of partial correlations (O'Connnor, 2000). In the first step, the first principal component (factor) is partialled out of the correlations between the variables and the average squared partial correlation is computed. In the next step, the first two variables are partialled out of the original correlation, than the first three, etc. The number of factors is determined by the step number in the analyses that resulted in the minimum average partial correlation. This number of factors is related to systematic, rather than unsystematic variance in the original correlation matrix. Optimal decisions are likely to be made after the results of both procedures have been considered.

2. Supplementary results

2.1 Study 1

We first conducted MAP test and parallel analysis with 9 months in order to investigate the factor structure. The average squared partial correlations associated with the first four components were .057, .040, .035 and .037. For the MAP test, the smallest average partial correlation is associated with the second component (.035), suggesting a two-factor solution.

For the parallel analysis, we used the SPSS syntax by O'Connor (2000); 9000 random data sets with 96 cases and 14 variables were computed, for which the first eigenvalues were 1.71, 1.52, 1.39 and 1.28. At root 4 the actual eigenvalue (1.12) fell below the mean random data eigenvalue (1.28), indicating a three-factor structure.

With 12 months, the average squared partial correlations in the MAP test associated with the first four components were .050, .044, .034 and .042. The smallest average partial correlation is associated with the second component (.034), suggestion a two-factor solution. Using SPSS syntax by O'Connor (2000), 9000 random data sets with 96 cases and 14 variables were computed, for which the first eigenvalues were 1.92, 1.68, 1.50 and 1.35. At root 4 the actual eigenvalue (1.46) fell below the mean random data eigenvalue (1.50), indicating an optimal two-factor structure by parallel analysis.

2.2 Study 2

2.2.1 Confirmatory factor analyses

Both tested models did not match the data at 12 months. The match statistics for the three-factor model were $\chi^2(74, N = 147) = 219.43$, $p < .001$; comparative fit index (CFI) = .72; Tucker-Lewis coefficient (TLI) = .65 ; root-mean-square error approximation (RMSEA) = .12 ; Akaike information criterion (AIC) = 309.43. For the two-factor model, the statistics were the following: $\chi^2(76, N = 147) = 208.10$, $p < .001$; CFI = .74; TLI = .69; RMSEA = .11; AIC = 294.10.

2.2.2 MAP test and parallel analysis

We first conducted MAP test and parallel analysis with 12 months. 9000 random data sets with 147 cases and 14 variables were computed, for which the first eigenvalues were 1.56, 1.42, 1.32 and 1.23. At root 4 the actual eigenvalue (0.95) fell below the mean random data eigenvalue (1.23), indicating a three-factor structure by parallel analysis. The average squared partial correlations in the MAP test associated with the first four components were

.065, .040, .029 and .031. The smallest average partial correlation is associated with the second component (.029), suggesting an optimal two-factor solution.

With 18 months, the smallest average partial correlation in the MAP test is associated with the third component (.026), as well suggesting three factors underlying toddler temperament in the ECBQ. Using SPSS syntax by O'Connor (2000), 9000 random data sets with 126 cases and 18 variables were computed. At root 4, the actual eigenvalue (1.21) fell below the mean random data eigenvalue (1.36), indicating an optimal three-factor structure by parallel analysis.

Analogous to 18 months, 9000 random data sets with 126 cases and 18 variables were computed with 24 months. At root 4, the actual eigenvalue (1.33) fell below the mean random data eigenvalue (1.36), indicating an optimal three-factor structure by parallel analysis. The smallest average partial correlation in the MAP test is again associated with the third component (.023), as well suggesting three factors with 24 months.

2.2.3 Developmental changes

We conducted paired *t*-tests and Wilcoxon tests for the seven ECBQ scales, depending on whether the data were normally distributed. Ratings for Shyness increased from 18 to 24 months [$t(109) = -2.86$, $p < .01$] as well as for Discomfort (18 months: $M = 2.31$, $SD = 0.88$; 24 months: $M = 2.63$, $SD = 0.80$; $p < .001$), and Inhibitory Control (18 months: $M = 3.49$, $SD = 0.95$; 24 months: $M = 3.84$; $SD = 0.95$; $p < .001$). No significant changes were observed for Attentional Shifting, Impulsivity, Motor Activation and Sociability.

References

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396-402.

4 Article 2 – Sieber, F., Czarnomski, J., Schölmerich, A., Daum, M. M., & Zmyj, N. (2024). The two functions of imitation in the second year of life: A longitudinal study. *Developmental Psychology*. Advance Online Publication.

The two functions of imitation in the second year of life: A longitudinal study

Franziska Sieber[1,] Jan Czarnomski[1], Axel Schölmerich[2], Moritz M. Daum[3] & Norbert Zmyj[1]

[1]TU Dortmund University, Dortmund, Germany

[2] Ruhr University Bochum, Bochum, Germany

[3]University of Zurich, Zurich, Switzerland

Published in *Developmental Psychology*. Advance online publication.

https://doi.org/10.1037/dev0001856

Corresponding author at: Institute of Psychology, TU Dortmund University, 44227 Dortmund, Germany. *E-mail address*: franziska.sieber@tu-dortmund.de (F.Sieber).

Declarations of interest: none.

Abstract

Infant imitation serves a cognitive and a social function. As part of their temperament, infants' attention and social orientation mirror these two functions. This longitudinal study investigated the development of the two functions within the second year of life in German infants (N = 136, 74 female), using standardized tests at the ages of 12, 18, and 24 months, conducted in 2018 and 2019. We measured temperament using two established parental questionnaires (IBQ-R, ECBQ) and behavioral observation (Lab-TAB), imitation using the Frankfurt Imitation Test, and cognitive development using the Cognitive Scale of the Bayley Scales. Hierarchical regressions revealed an association between imitation and social orientation from 12 months onwards, whereas no clear relation emerged between imitation and attentional variables independently of infants' cognitive development. The findings suggest that imitation serves a primarily social function early in life.

Keywords: imitation, temperament, social function, cognitive function, infants, attention, social orientation

Public significance statement

This longitudinal study suggests that infants' and toddlers' imitative behavior depends on interindividual differences in their social orientation. This finding supports the notion that infants and toddlers imitate not only to learn new skills but also to engage in social interaction.

The two functions of imitation in the second year of life: A longitudinal study

Imitation is central to early social-cognitive development and knowledge acquisition, and has been widely studied in the past few decades, with a large amount of research focusing on the onset of imitation and on external factors such as the model's age (Seehagen & Herbert, 2011) or language (Buttelmann et al., 2013). Fewer studies focused on infants' characteristics that might influence whether they imitate. These studies are primarily based on the assumption that imitation serves two functions in infancy (Nielsen, 2009; Over & Carpenter, 2013; Užgiris, 1981): a cognitive function to learn new skills and a social function to communicate non-verbally and to share and create experiences with an interaction partner. Užgiris (1981) suggested that during the second year of life, the social function of imitation becomes more important, while the opposite is true for the cognitive function. The present study aimed to investigate this hypothesis.

Imitation

Definitions of imitation vary between studies depending on the context and goals of the studies. Some definitions include understanding the model's intention as a necessary condition (e.g., Carpenter & Call, 2009). Definitions based on intention reading include the possible underlying mechanism that subserves imitation. Because the goal of this study is not to answer the underlying mechanisms of imitation, we adopt a minimal definition of imitation as the reproduction of any observed behavior (Paulus, 2011).

Infants differ in the rigor of their imitation: Some imitate with precision irrespective of whether an action is causally relevant (*exact imitation*), while others imitate based on whether an action is causally relevant (*selective imitation*; e.g., Brugger et al., 2007; Hilbrink et al., 2013). Research has shown that 15-month-olds are more likely to show exact imitation than 12-month-olds (Hilbrink et al., 2013; see also Gellén & Buttelmann, 2019), with this decrease in the selectivity of imitative behavior interpreted as a shift in the motivation underlying imitation.

Two functions of imitation

In a seminal paper, Užgiris (1981) postulated that imitation in infancy serves a cognitive and a social function. The cognitive function is driven by infants' instrumental motivation to learn new skills: The imitator learns actions in order to solve a given problem without trial-and-error learning. The social function is driven by infants' social motivation to interact with the model: The imitator communicates non-verbally and identifies or conforms with the model by communicating mutuality. The two functions are not seen as competing with one another, but are rather intricately related in children's social-cognitive development.

Užgiris (1981) further assumed that through imitation and changes in infants' cognitive abilities within the second year of life, their motivation to imitate also changes, as they begin to understand their fundamental similarity to others and thus start to develop an understanding of social interactions. Accordingly, the respective importance of the two functions changes during the second year of life, with the social function becoming more important and the cognitive function becoming less important (Užgiris, 1981; see also Over & Carpenter, 2013). Other researchers assume, more extensively, that imitation is always embedded in a social context and thus consider the social function as key to understanding imitation (Over & Carpenter, 2013).

The seminal work of Užgiris (1981) formed the basis for empirical studies (e.g., Hilbrink et al., 2013; Óturai et al., 2018) and theoretical approaches addressing differences in infants' imitative behavior (Nielsen, 2009; Over & Carpenter, 2013). For instance, the aforementioned trend from selective towards exact imitation within the second year of life (e.g., Óturai et al., 2018) can be integrated into Užgiris' (1981) theoretical framework: Given that 24-month-olds focus predominantly not on acquiring new skills but on interacting and creating mutuality with the model, they imitate both causally relevant actions and irrelevant actions.

Similar approaches to that of Užgiris (1981) are the dual-engine approach (Legare & Nielsen, 2015) and the bifocal-stance theory (Jagiello et al., 2022). The dual-engine approach provides a framework to explain how children's cultural learning is driven by imitative fidelity and innovation. Specifically, according to this approach, instrumental skills and cultural conventions need to be transmitted with high-fidelity imitation, and at the same time, innovation is necessary to ensure adaptation to new environments (Legare & Nielsen, 2015). The bifocal-stance theory (Jagiello et al., 2022) extends the ideas of the dual-engine approach, suggesting that the ability to adopt different motivational stances results in the co-existence of high-fidelity imitation and innovation. Both of these approaches provide a framework for cultural evolution, while the two processes involved differ in whether or not an individual imitates the model exactly. By contrast, Uzgiris' approach focuses on developmental shifts in the reasons for imitative acts within the second year of life. Irrespective of the differences between these approaches, all share the assumption that imitative behavior is driven by motivational differences (instrumental vs. social) in when and what to learn from others.

Imitation tasks are typically used in two different contexts. One line of research uses such tasks to test the capacity of infants' declarative memory (Barr et al., 1996). The assumption here is that infants need to encode and store the demonstrated action in their long-term memory in order to reproduce it after a delay. If infants do not imitate the demonstrated action, it is typically assumed that they do not have the cognitive abilities required to solve this task. Another line of research uses imitation tasks to test social-cognitive understanding. Here, the ability of imitation is taken as given, and additional information is manipulated, such as the model's intentionality during the demonstration (Carpenter et al., 1998) or the situational constraints (Gergely et al., 2002). If infants fail to show a difference in their imitative behavior between conditions, it is assumed that they have not developed the social-cognitive ability in question.

Against the background of the two functions of imitation, the question arises of whether imitation is simply an intellectual performance (e.g., the capacity of declarative memory or social-cognitive understanding) or whether infants are differently motivated to engage in imitation. To examine this question, we aimed to investigate the extent to which imitation is used to learn or to interact with others. For example, irrespective of cognitive abilities, a less extroverted child, who attaches little importance to engaging in contact with the model, might be less socially motivated to imitate the action demonstrated by the model, whereas the opposite might be true for a more extroverted child. Thus, with the present study, we sought to investigate how infants' cognitive and social skills contribute to their imitative behavior.

Temperament

An infant's behavior in general, and imitation in particular, can be described in terms of function and context. However, behavior is further influenced by the infant's individual characteristics, such as personality or temperament. Infants' temperament is constitutionally based on individual differences in reactivity and self-regulation (Rothbart & Bates, 2006). Reactivity refers to an individual's response to changes in his/her internal or external environment and includes reactions on a physiological and behavioral level. Self-regulation refers to the regulation of these processes on a neural level, in terms of physiological arousal, and a behavioral level, in terms of social processes (Rothbart & Derryberry, 1981). These individual differences are commonly observed on three broad factors: Surgency/Extraversion, Negative Affectivity, and Orienting/Regulation (Gartstein & Rothbart, 2003). Surgency/Extraversion includes infants' enjoyment of social situations, positive emotions in social situations, and the desire to interact with others (Gartstein & Rothbart, 2003). Thus, infants' extraversion might promote the social function of imitation. The self-regulation factor is termed, depending on age group and measurement tool, as Orienting/Regulation (for infants up to 12 months, Gartstein & Rothbart, 2003) or Effortful Control (for toddlers between 18 and 36 months, Putnam et al., 2006), and describes infants' regulatory capacities related to their behavior and attention. Infants' attention capacities, such as the ability to direct and sustain attention and resist distractions, are conceived as part of the regulation domain (Putnam et al., 2006). Accordingly, attention capacities, as a prerequisite for learning, might promote the cognitive function of imitation.

Relation between imitation and temperament

Personality traits such as introversion and extraversion are related to imitative behavior from early in life (Fenstermacher & Saudino, 2006). Temperament, as a relatively stable characteristic in infants and toddlers (Putnam et al., 2008; Sieber & Zmyj, 2022), is also related to infant imitation (for an overview, see Heimann, 2022). Attention, as part of the regulation factor and a precursor of Effortful Control, is likewise associated with imitation early on (Heimann, 2002) and throughout the second year of life. Twelve-months-olds' attention measured via a temperament questionnaire and direct observation in the laboratory is related to infants' imitative behavior (Zmyj et al., 2017). Twenty-one-month-olds with a high attentional focus were less affected by environmental distractions in an imitation task compared to those with a low attentional focus (Dixon et al., 2006). Furthermore, Effortful Control was found to be associated with attentional capacities and long-term imitative recall in 21-month-olds (Dixon et al., 2012).

A link between social aspects of temperament and imitation has been found during but not at the beginning of the second year of life: At 12 months, imitation and extraversion were not yet correlated (Zmyj et al., 2017), while at 15 months, children with high levels of extraversion were more likely to imitate faithfully than children with low levels of extraversion (Hilbrink et al., 2013), and extraversion at 15 months was also found to predict imitative behavior at 21 months (Dixon et al., 2012). This previous research suggests that links between imitation and temperament do exist within the second year of life. Attention appears to be linked to imitation from 12 months (Zmyj et al., 2017) to the end of the second year of life (Dixon et al., 2006). In contrast, relations between extraversion and imitation were found to emerge from the middle of the second year of life (Hilbrink et al., 2013) but were not yet present at 12 months (Zmyj et al., 2017).

Nevertheless, several issues remain unresolved regarding the relation between imitation and temperament. First, to date, no longitudinal study has investigated the whole age range of the second year of life. For instance, Dixon and colleagues (2012) focused on 15- and 21-month-olds, while Hilbrink and colleagues (2013) focused on 12- and 15-month-olds, and the other studies discussed above had a cross-sectional design. To detect a change in the development of the two functions of imitation, longitudinal data throughout the entire second year of life are needed. Second, information about the psychometric properties (i.e., reliability and validity) of the imitation tests employed in the various studies is mostly lacking (e.g., Dixon et al., 2006; Dixon et al., 2012). Third, temperament was primarily measured only using a questionnaire (Dixon et al., 2006; Dixon et al., 2012; Hilbrink et al., 2013). Given that the research so far is mainly based on parent reports, additional observational measures of temperament are necessary. Fourth, the aforementioned studies largely relied on relatively small samples (e.g., $N = 32$, Dixon et al., 2012; $N = 39$, Dixon et al., 2006).

The present study

In the present study, we investigated the developmental course of the two functions of imitation over the second year of life (at 12, 18, and 24 months). Using a longitudinal design, we employed standardized imitation, temperament, and cognition tests. Imitation was assessed using the Frankfurt Imitation Tests for 12-, 18-, and 24-month-olds (FIT 12, Goertz et al., 2006; FIT 18, FIT 24, Goertz et al., 2008). The FITs are standardized imitation tests, can be analyzed objectively, and have shown good psychometric properties, with high test-retest reliability. They allow for the assessment of inter-individual differences within one age group and cover age-related increases between the different age groups by varying in their difficulty according to age. Temperament was assessed using the Infant Behavior Questionnaire Revised (IBQ-R, Gartstein & Rothbart, 2003) at 12 months and the Early Childhood Behavior Questionnaire (ECBQ, Putnam, et al., 2006) at 18 and 24 months, and through two tasks from the Laboratory Temperament Assessment Battery (Lab-TAB, Locomotor Version, Goldsmith & Rothbart, 1999), which entails 20 tasks that elicit emotions in a standardized environment to objectively test infants' temperament. For the present study, we chose the Task Orientation Test (TOT), which assesses children's interest in non-social objects, and particularly their persistence, by measuring their allocation of attention at each test phase, and the Puppet Game Test (PGT), which evaluates infants' social orientation at each test phase by measuring their response to social stimulation while interacting with hand puppets. The Cognitive Scale of the Bayley Scales of Infant and Toddler Development (BSID-III, Bayley, 2006) was applied at each test phase to control for children's cognitive developmental status when investigating the relation between imitation and temperament. To differentiate between the instrumental motivation to imitate and infants' cognitive abilities, we used the IBQ-R (Gartstein & Rothbart, 2003) and the Cognitive Scale of the BSID-III (Bayley, 2006). As mentioned above, we aimed to relate infants' characteristics to their imitative behavior in order to investigate the contribution of infants' motivation to their imitative behavior independently of their cognitive abilities. To the best of our knowledge, there is no existing scale that assesses infants' instrumental motivation to learn.

Based on Užgiris (1981), we first hypothesized that infants' social orientation measured using the IBQ-R and PGT would be unrelated to their imitative behavior at 12 months, but that a relation between social orientation and imitation would be apparent at 18 months and would be even stronger at 24 months. Second, we hypothesized that attentional aspects of infant temperament, as measured using the IBQ-R and ECBQ as well as the tasks from the TOT, would be associated with their imitative behavior throughout the entire second year of life, at 12, 18, and 24 months. Third, we hypothesized that these postulated associations would remain constant after controlling for infants' cognitive development (BSID-III). Fourth, we tested whether infants' social orientation at 12 months (IBQ-R subscale "Sociability"; PGT) predicts imitative behavior at 18 months and 24 months and whether their social orientation at 18 months predicts imitative behavior at 24 months.

Method

Participants

The participating families lived in one of two medium-sized cities in northwestern Germany. They were recruited from a database of families who had previously registered to participate in child development studies. The initial sample at the first measurement time point (T1) consisted of 157 infants and their parents (all Caucasian). Of these infants, 21 participants had to be excluded at T1. The majority of parents had either a university degree (mothers 60%, fathers 54%) or a university entrance-level diploma (mothers 18%, fathers 13%) as their highest educational attainment. Forty additional infants from the responding families had to be excluded at T2 and 25 additional toddlers had to be excluded at T3 (for detailed explantions, see Supplementary methods). Table 1 displays the final sample sizes at each age. Data collection took place between May 2018 and October 2019. Parents gave informed consent for their children to participate in the study. At each visit, the children received a small gift and parents received 10 Euros. Families who participated in all three test phases additionally received 15 Euros.

Table 1

Number of Children, Children's Sex, Children's Mean Age (Months; Days) with Standard Deviation in Parentheses (Days) and Age Range (Days) at 12, 18, and 24 Months

Measure	12-month-olds	18-month-olds	24-month-olds
N	136	102	107
Sex	74 female	55 female	58 female
Children's age *M* (*SD*)	11;26 (0;10)	17;27 (0;10)	23;26 (0;08)
Range (days)	344 – 386	530 – 572	708 – 748

To determine the appropriate sample size, an a priori power analysis was conducted using G*power 3.1, which revealed that a sample of 85 participants was sufficient to achieve 80% power and a medium effect size (Faul et al., 2007). To ensure that data were obtained from 85 participants at each assessment time point, we invited 160 children and their families to participate. The final sample for the longitudinal analyses from 12 to 18 months consisted of 94 children (53 girls). The children's mean age was 11 months, 23 days (*SD* = 8 days, range 344–381 days) at T1 and 17 months, 26 days (*SD* = 10 days; range 530–572 days) at T2. For the longitudinal analyses from 12 to 24 months, the sample consisted of 98 children (55 girls) with a mean age of 11 months, 23 days (*SD* = 8 days, range 347–381 days) at T1 and 23 months, 26 days (*SD* = 9 days; range 708-748 days) at T3. The final sample for the longitudinal analyses from 18 to 24 months consisted of 78 children (43 girls). The mean age was 17 months, 27 days (SD = 10 days; range 532-569 days) at T2 and 23 months, 25 days (SD = 9 days; range 708-748 days) at T3. The present study was approved by the local Ethics Committee of the Faculty of Psychology at Ruhr University Bochum.

Materials

The materials used in the imitation tests were as close as possible to the material used in the original tasks (for an overview, see Table 2). A detailed description of the materials used in the FIT 12, 18, and 24 is provided in the Supplementary Material. For the two tests of the Lab-TAB (Goldsmith & Rothbart, 1999), we used eight toy blocks of different colors and sizes in the TOT and two hand puppets in the PGT. During T1, the air conditioning in the laboratory did not work, and we therefore recorded the temperature in the laboratory using the USB data logger Elitech® RC-5 (Elitech UK, London, United Kingdom).

Table 2

Overview of Materials and Tasks of the Imitation Test

	Tin Can	Pig	Cup & Knife	Mouse	Drum			
FIT 12								
	Car	Goose, Tin Can	Mouse	Panda & Ring	Drum	Octopus		
FIT 18								
	Gondola	Boat & Box	Panda & Slide	Ball with Eyes	Turtle	Bunny	Box	Magnetic Plate
FIT 24								

Design and Procedure

Children were tested longitudinally at 12, 18, and 24 months. At each measurement time point, all children participated in all four tests (FIT 12/18/24, PGT, TOT, BSID-III) and all parents completed the temperament questionnaire (IBQ-R at T1, ECBQ at T2 and T3). The questionnaires were sent to the parents one week prior to the appointment and they were asked to bring the completed questionnaires to their appointment at the university. The same experimenter conducted all tests in the same laboratory at TU Dortmund University. Children were tested individually in the laboratory in the presence of their parents, and parents were instructed not to interact with their children. The session began with a warm-up phase, in which the experimenter played with the children until they felt comfortable, that is, he played with the child until the child smiled at him and gave him a toy back during play. Next, the study started with a baseline phase of the FIT. Children were seated in a highchair facing the experimenter at the opposite side of the table. For the demonstration phase, the children switched to another table in the room, where an eye tracker was located 60 cm in front of the child. The eye-tracking data are not reported because they address theoretically different questions. The experimenter demonstrated the target actions live, after which the children returned to the first table and the experimenter conducted the TOT followed by the PGT. After a break of approx. five minutes, the test phase of the FIT was conducted. The session ended with the cognitive subscale test of the Bayley Scales. The whole session lasted for approximately one hour.

Temperament Questionnaires

IBQ-R. At T1, we used the German version (Vonderlin et al., 2012) of the IBQ-R (Gartstein & Rothbart, 2003). The IBQ-R assesses early infant temperament with 191 items grouped into 14 scales. Parents were asked to indicate how often a specific behavior occurred in the past week or the past two weeks on a 7-point Likert scale ranging from 1 (never) to 7 (always). Whereas the English-language version suggests a three-factor solution (i.e., Surgency/Extraversion, Negative Affectivity, and Orienting/Regulation), the German-language version suggests a two-factor solution (i.e., Surgency/Extraversion and Negative Affectivity; Mink et al., 2012; Sieber & Zmyj, 2022). Therefore, we analyzed the IBQ-R data only on a scale level. Cronbach's αs for the 14 subscales of the German version range between .74 and .94 (Vonderlin et al., 2012). Additionally, we rationally derived another scale from the item pool of the IBQ-R, labeled "Sociability", for the following reasons: First, as we did not replicate the original Surgency/Extraversion factor, we were unable to use this factor as a predictor in our analyses. Second, the single scales loading on this factor (Approach, High-intensity pleasure, Smiling and laughter, Activity level, Vocal reactivity, and Perceptual sensitivity) assess not only children's social orientation and joy in social situations but also other facets of positive affect that are not associated with social situations (e.g., "How often during the last week did the baby repeat the same sounds over and over again", scale Vocal reactivity). Third, in contrast to the ECBQ, the IBQ-R does not contain a "Sociability" scale measuring "seeking and taking pleasure in interactions with others" (ECBQ, Putnam et al., 2006, p. 18). The derived Sociability scale consists of 19 items (for details, see Supplementary methods) and showed an adequate internal consistency, with a Cronbach's α of .77, indicating a homogenous scale.

ECBQ. At T2 and T3, we used the German translation (Kirchhoff et al., 2013) of the ECBQ (Putnam et al., 2006). The ECBQ assesses toddlers' temperament with 201 items grouped into 18 scales. Studies in Germany and in English-speaking countries revealed the same three-factor solution (i.e., Surgency/Extraversion, Negative Affectivity, and Effortful Control, Putnam et al., 2006; Sieber & Zmyj, 2022). Therefore, we analyzed questionnaire data at T2 and T3 on the scale and factor level. With the exception of the Activity level scale, which showed a Cronbach's α of .56, Cronbach's αs for the 18 subscales of the German translation ranged between .61 and .87 (Kirchhoff et al., 2013). Like in the IBQ-R, parents are asked to indicate how often a specific behavior occurred in the past two weeks on a 7-point Likert scale ranging from 1 (never) to 7 (always).

Imitation tests

FIT 12. At T1, we tested infants' imitation using the FIT 12 (Goertz et al., 2006). The test consisted of five tasks presented in a fixed order: Tin Can, Pig, Cup & Knife, Mouse, and Drum. In contrast to the original FIT, we implemented a baseline phase in which the infants received the objects without any prior demonstration of the target actions in order to assess whether the infants spontaneously showed the target actions (see also Zmyj et al., 2017). The experimenter placed the items on the table in front of the infant one after another, and from the moment the experimenter let go of each item, the infant had 30 seconds to act on it. Immediately thereafter, the demonstration phase started. The experimenter demonstrated the target actions four times without pausing between the five tasks. Infants sat on their parent's lap opposite the experimenter, who presented the target actions on a marked line on a wooden box on the table. Fifteen minutes after the demonstration phase had ended, the test phase started. Infants received the objects again one after another, and from the moment the experimenter placed an object on the table in front of the infant and let go of it, the infant again had 30 seconds to act on it. The one-hour duration of the session, which included several tasks (FIT, PGT, and TOT from the Lab-TAB, and the Cognitive Scale from the Bayley Scales), was challenging for the children, especially the 12-month-olds. Therefore, to keep the infants' attention during all tasks, we reduced the delay between the demonstration phase and the test phase from 30 minutes in the original version of the FIT (Goertz et al., 2006) to 15 minutes in the present study. According to Heimann and Meltzoff (1996), a delay of 10 minutes is sufficient to test deferred imitation if infants are suitably distracted during this delay with other tasks or toys. In our delay, the two tasks of the Lab-TAB distracted the infants with hand puppets and blocks. The results show that imitation rates in our study are similar to those reported in other studies using the FITs with a 30-minute delay (e.g., Kolling et al., 2010; Zmyj et al., 2017).” Another difference to the original FIT is that we awarded an extra point in the Pig task (see Zmyj et al., 2017) because we conducted a two-step action and put the hat back on the pig. The target actions on the objects consisted of one- and two-step actions with a total of eight target actions. For a detailed description of the target actions of all FITs, see the Supplementary methods.

FIT 18. At T2, we tested infants’ imitation using the FIT 18 (Goertz et al., 2008). The test consisted of six tasks presented in a fixed order: Car, Goose & Tin Can, Mouse, Panda & Ring, Drum, and Octopus. The procedure was the same as at T1. In contrast to the FIT 12,the target actions were demonstrated three times, in line with the original procedure. The target actions on the objects consisted of two- and three-step actions with a total of 12 target actions.

FIT 24. At T3, we tested infants’ imitation using the FIT 24 (Goertz et al., 2008). The test consisted of eight tasks presented in a fixed order: Gondola, Boat & Box, Panda & Slide, Ball with eyes, Turtle, Bunny, Box, and Magnetic Plate. The test followed the same procedure as at T1 and T2. The experimenter demonstrated the target actions twice. The target actions consisted of three- to six-step actions with a total of 29 target actions.

Task orientation test (Lab-TAB)

The TOT was derived from the locomotor version of the Lab-TAB (Goldsmith & Rothbart, 1999). The experimenter placed eight toy blocks on the table in front of the child and then left the room. The child then had 3 minutes to manipulate and play with the blocks. The parents were instructed to put the blocks back on the table if the child threw them off but to not otherwise intervene. After 3 minutes, the experimenter returned and the task ended.

Puppet game test (Lab-TAB)

The PGT was derived from the locomotor version of the Lab-TAB. The experimenter put on two hand puppets, named Sarah and Clara, and performed a scripted puppet show, translated into German for the purpose of the study, for one minute. During the puppet show, Sarah and Clara talk to each other and both tickle the child once. At the end of the show, they both tickle the infant simultaneously. After this, the experimenter placed both puppets on the desk in front of the child, and the child was able to play with the puppets for 30 seconds before the task ended.

Cognitive Scale (BSID-III)

We assessed children's cognitive development status using the Cognitive Scale of the BSID-III following the guidelines in the technical manual (Bayley, 2006). The Cognitive Scale provides standardized scores for the cognitive-developmental status (M = 100, SD = 15).

Coding

All sessions were video-recorded. Ten coders rated the videos independently using the software Interact® (version 14.3.1, Mangold Software & Consulting GmbH, Arnstorf, Germany).

Imitation tests

In the imitation tests, children received one point if they produced the respective target act. For a detailed description of each respective relevant target act of the imitation tests at T1, T2, and T3 coded with one point, see the Supplementary methods.

FIT 12. Infants received a score between 0 and 8 in the baseline phase and imitation phase (see Procedure section). Another independent observer coded sixty randomly chosen videos. The Intraclass Correlation Coefficient (ICC, average measure, absolute agreement, two-way mixed) was excellent for the baseline phase at T1 (*r* = .891, *F*(59,59) = 9.552, *p* < .001 and for the imitation phase at T1 (*r* = .968, *F*(59,59) = 32.428, *p* < .001).

FIT 18. Toddlers received a score between 0 and 12 in the baseline phase and in the imitation phase. Another independent observer additionally coded the videos of 65 randomly chosen children. The interrater reliability was high for the baseline phase (*r* = .827, *F*(64,64) = 6.845, *p* < .001, ICC) and good for the imitation phase (*r* = .760, *F*(61,61)= 4.888, *p* < .001, ICC).

FIT 24. Toddlers could receive a score between 0 and 29 in the baseline phase and in the imitation phase. Another independent observer coded fifty-six randomly chosen videos. The ICC was excellent for the baseline phase (*r* = 0.934, *F*(55,55) = 16.533, *p* < .001) and good for the imitation phase (*r* = 0.717, *F*(54,54) = 12.682, *p* < .001).

Task orientation test (Lab-TAB)

The same procedure for the TOT was used at all three measurement time points. We coded the *Duration of looking* towards the blocks and the *Duration of playing* with the blocks in seconds over the total test time of 3 minutes. We further coded the *Latency of the first look away* (in seconds, i.e., the time until the first glance away from the blocks). We additionally coded the child's *Baseline state* (1 = drowsy; 2 = alert/calm; 3 = alert/active; 4 = fussy; 5 = crying), depicting the child's state prior to the beginning of the playtime at T1, T2, and T3. For the coding of the child's intensity of facial interest (0 = no facial interest; 1 = low-intensity interest; 2 = strong facial interest), we used the standard coding as suggested by the manual (Goldsmith & Rothbart, 1999) and divided the 3 minutes of play into three 1-minute intervals, each subdivided into six episodes of 10 s. The intensity of facial interest was coded for each of these 18 episodes. Mean scores of the 18 episodes were calculated, resulting in a *Mean child's facial interest* at T1, T2, and T3. Independent second observers additionally coded the videos of 60 randomly chosen infants at T1, 61 randomly chosen infants at T2, and

53 randomly chosen toddlers at T3. Interrater reliabilities were all at least moderate at T1 (all *r*s > .661, all *p*s < .001) and good to excellent at T2 (all *r*s > .766, all *p*s < .001) and at T3 (all *r*s > .812, all *p*s < .001, ICC; see Supplementary methods for details).

Puppet game test (Lab-TAB)

In the PGT, we used the standard coding as suggested by the manual (Goldsmith & Rothbart, 1999). The procedure was the same for all three measurement time points. The play was divided into five episodes of different durations that covered different parts of the interaction. The first episode lasted from the beginning of the play until the first tickle, the second episode lasted from the first until the second tickle, the third episode lasted from the second until the third tickle, and the fourth episode started with the third tickle and lasted until the end of the story when the experimenter placed the puppets on the table. The last (fifth) episode consisted of the 30 s during which children were able to play with the puppets after they were placed on the table. For each of the five episodes, the observers coded the intensity of smiling (0 = no smile; 1 = small smile; 2 = medium smile; 3 = large smile), presence of laughter (0 = not present; 1 = present), positive vocalizations (0 = not present; 1 = present), positive motor activity (0 = not present; 1 = present), and engagement with toy (0 = indifferent; 1 = neutral, mild interest; 2 = fully engaged). We calculated the mean scores of the five episodes, resulting in a *Mean intensity of smiling*, *Mean presence of laughter*, *Mean positive vocalizations*, *Mean positive motor activity,* and *Mean engagement behavior with the toy* at T1, T2, and T3. Independent second observers additionally coded the videos of 60 randomly chosen infants at T1, 59 randomly chosen infants at T2, and 60 randomly chosen toddlers at T3. ICCs were all at least moderate at T1 (all *r*s > .613, all *p*s < .001) and T2 (all *r*s > .661, all *p*s < .001) and good to excellent at T3 (all *r*s > .782, all *p*s < .001; see Supplementary methods for detailed results).

Cognitive Scale (BSID-III)

The coding of the Cognitive Scale of the BSID-III at T1, T2, and T3 followed the guidelines in the technical manual (Bayley, 2006). Three independent second observers coded the videos of 60, 55, and 58 randomly chosen children at 12, 18, and 24 months, respectively. Interrater reliabilities for the BSID-III scores were excellent (all *r*s >.94; all *p*s < .001, ICC).

Analyses

Children's imitative performance at 12, 18, and 24 months was assessed using paired-sample *t*-tests comparing FIT baseline and FIT test phase scores. To conduct multiple linear regression analyses reliably, the data need to be suitably correlated with the dependent variable. Accordingly, we calculated Pearson correlations between the dependent variable FIT 12/18/24 test score and the attentional and social variables of temperament at the respective ages before running the hierarchical linear regressions.

Finally, hierarchical linear regression analyses were conducted from a cross-sectional perspective, with the FIT 12/18/24 test score as dependent variable and attentional and social aspects of temperament as predictors. Children's cognitive developmental status (BSID-III at T1, T2, and T3) and the baseline imitative behavior (FIT 12/18/24, baseline phase score) were included as control variables in each cross-sectional hierarchical regression analysis. At T1, temperature was added as the third control variable because the air conditioning in the lab failed (see Design & Procedure section).

Longitudinal relations were assessed using two hierarchical linear regression analyses. For longitudinal predictions from 12 to 18 months, the FIT 18 test score served as dependent variable and social aspects of temperament at 12 months (i.e., PGT of the Lab-TAB at T1 and "Sociability" scale of IBQ-R) served as predictors. For longitudinal predictions from 12 to 24 months, the FIT 24 test score was used as dependent variable and social aspects of temperament at T1 (i.e., PGT of the Lab-TAB at T1 and “Sociability” scale of IBQ-R) were used as predictors. Children’s cognitive developmental status at T1, FIT test score at T1, and FIT 18/24 baseline score were included as control variables in each hierarchical linear regression analysis for longitudinal relations. For longitudinal predictions from 18 to 24 months, the FIT 24 test score served as dependent variable and social aspects of temperament at T2 (i.e., PGT of the Lab-TAB at T2 and Sociability subscale of the ECBQ) served as predictors. The control variables were children’s cognitive developmental status at T2, FIT test score at T2, and FIT 24 baseline score.

Transparency and Openness

We report how we determined our sample size, all data exclusions, and all measures in the study**.** All data and analysis codes are available at https://osf.io/zwqty/?view_only=05a5c83109e04ca59256ae108b6cb2fe. Data were analyzed using IBM SPSS Statistics for Windows, version 28 (IBM Corp., 2021). This study’s design and its analysis were not pre-registered. The study was designed in 2016, when pre-registrations were not yet a standard.

Results

Test statistics at 12, 18, and 24 months

Imitation tests

Children produced fewer target acts in the baseline phase than in the test phase at all three time points (12-month-olds: $M_{\text{baseline}} = 2.66$, $SD = 1.26$ vs. $M_{\text{test}} = 4.47$, $SD = 1.79$, $t(135) = -13.61$, $p < .001$; 18-month-olds: $M_{\text{baseline}} = 2.39$, $SD = 1.06$ vs. $M_{\text{test}} = 5.38$, $SD = 1.82$, $t(101) = -14.93$, $p < .001$; 24-month-olds: $M_{\text{baseline}} = 2.70$, $SD = 1.90$ vs. $M_{\text{test}} = 11.32$, $SD = 4.17$, $t(106) = -23.08$, $p < .001$).

Temperament questionnaires

The descriptive data of the IBQ-R scales at T1 are provided in Table 3. The mean score of the rationally derived “Sociability” scale was *M* = 5.57 (*SD* = 0.60). In contrast to the IBQ-R, data from the ECBQ also allowed for an analysis on the factor level. The descriptive data of the ECBQ scales and factors at T2 and T3 are provided in Table 4.

Temperament observations (Lab-TAB)

Task orientation tests. The median of the latency of first look away was *Mdn* = 3.46 s at T1, *Mdn* = 2.67 s at T2, and *Mdn* = 3.67 s at T3. Descriptive statistics (*M*, *SD*) of mean duration of looking, mean duration of playing, mean facial interest, and children’s baseline state scores at 12, 18, and 24 months are shown in Table 5.

Puppet game tests. Mean scores and standard deviations of variables coded in the PGT at 12, 18, and 24 months are depicted in Table 6.

Cognitive Scale (BSID-III)

Children’s mean cognitive developmental status scores assessed using the Cognitive Scale of the BSID-III were *M* = 78.64 (*SD* = 14.86) at 12 months, *M* = 86.95 (*SD* = 17.49) at 18 months, and *M* = 89.62 (*SD* = 14.53) at 24 months.

Table 3

Mean Scores and Standard Deviations of IBQ-R scales at 12 months

IBQ-R Scale	*M*	*SD*
Activity Level	3.88	0.83
Distress to Limitations	4.22	0.81
Fear	2.69	0.94
Duration of Orienting	3.03	0.98
Smiling and Laughter	4.65	0.84
High-Intensity Pleasure	6.03	0.64
Low-Intensity Pleasure	4.85	1.02
Soothability	5.34	0.81

Falling Reactivity	5.04	1.02
Cuddliness	5.34	0.81
Perceptual Sensitivity	4.46	1.14
Sadness	3.22	0.83
Approach	5.78	0.69
Vocal Reactivity	4.90	0.99

Table 4

Mean Scores and Standard Deviations of ECBQ scales at 18 and 24 months

ECBQ Scales and Factors	18 months		24 months	
	M	*SD*	*M*	*SD*
Activity Level	4.73	0.81	4.63	0.77
Attentional Focusing	4.20	0.93	4.47	0.84
Attentional Shifting	4.46	0.58	4.54	0.69
Cuddliness	5.08	0.80	5.29	0.71
Discomfort	2.31	0.88	2.63	0.83
Fear	2.12	0.69	2.34	0.82
Frustration	3.56	0.87	3.68	0.83
High-Intensity Pleasure	5.10	0.79	5.05	0.83
Impulsivity	5.02	0.72	5.06	0.68
Inhibitory Control	3.52	0.97	3.82	0.96
Low-Intensity Pleasure	4.97	0.81	5.06	0.71
Motor Activation	2.37	0.86	2.16	0.70
Perceptual Sensitivity	4.44	1.03	4.59	0.96
Positive Anticipation	4.72	0.98	5.25	0.83
Sadness	2.75	0.84	2.80	0.83
Shyness	3.19	0.86	3.35	0.93
Sociability	5.27	0.93	5.43	0.93
Soothability	5.90	0.76	5.79	0.69
Extraversion	4.97	0.55	5.08	0.54
Negative Affectivity	2.85	0.50	2.74	0.49

Effortful Control	4.45	0.54	4.63	0.49

Note. Factors are marked in bold.

Table 5

Mean Scores and Standard Deviations of Task Orientation Test Variables

Variable	12 months	18 months	24 months
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Mean duration of looking [a]	136.42 (24.12)	136.89 (27.55)	141.41 (31.63)
Mean duration of playing [a]	132.41 (35.95)	120.11 (41.81)	132.92 (38.60)
Mean intensity of interest [b]	1.62 (0.29)	1.56 (0.32)	1.64 (0.37)
Baseline state [c]	2.30 (0.59)	2.26 (0.61)	2.25 (0.64)

Note. [a] *M* and *SD* are presented in seconds. [b] Scores between 0 and 2 are possible. [c] Scores between 1 and 5 are possible, with higher scores indicating greater upset.

Table 6

Mean Scores and Standard Deviations of Puppet Game Test Variables

Variable	12 months	18 months	24 months
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Mean intensity of smiling [a]	1.76 (0.98)	1.87 (0.98)	1.99 (1.02)
Mean presence of laughter [b]	.02 (0.07)	.04 (0.12)	.07 (0.17)
Mean positive vocalizations [b]	.23 (0.26)	.27 (0.30)	.17 (0.27)
Mean positive motor activity [b]	.54 (0.32)	.60 (0.30)	.58 (0.35)
Mean engagement with toy [c]	1.86 (0.25)	1.91 (0.30)	1.91 (0.28)

Note. [a] Scores between 0 and 3 are possible. [b] Scores between 0 and 1 are possible. [c] Scores between 0 and 2 are possible.

Relation between children's temperament and their imitative behavior

12 months

Based on the theoretical assumptions, we first calculated Pearson correlations between the FIT 12 test score and attentional and social aspects of temperament as possible predictors of imitative behavior (see Supplementary results). We selected the following variables, with a $p < .15$, to be included in regression analyses as possible predictors: Duration of playing

(TOT), Duration of looking (TOT), Duration of orienting (IBQ-R), Sociability (IBQ-R), Smiling and laughter (IBQ-R), High-intensity pleasure (IBQ-R), and Approach (IBQ-R). Before running the regression analyses, we calculated Pearson correlations between the possible predictors from the same test procedure (i.e., TOT scores, IBQ-R subscales; see Supplementary results) to control for multicollinearity between the predictors. Correlations with a *p*-value above .80 indicate multicollinearity (Field, 2018). The IBQ-R subscales Sociability and High-intensity pleasure were strongly correlated, $r = .82$, $p < .001$. Thus, we excluded High-intensity pleasure as a predictor, as two predictors that are correlated even if they miss the threshold of multicollinearity ($p > .80$) might bias the results of a regression analysis. Both Smiling and laughter ($r = .79$, $p < .001$) and Approach ($r = .57$, $p < .001$) were significantly correlated with sociability. Therefore, we excluded both scales from the regression analysis as predictors. Duration of looking (TOT) and Duration of playing (TOT) were highly correlated ($r = .34$, $p < .001$). We chose Duration of playing as a predictor because it assesses infants' persistence in this specific task and therefore seems to be a better indicator of infants' instrumental motivation than the simple looking time during this task. Thus, Duration of looking was excluded as a possible predictor. We conducted theory-driven hierarchical linear regression analyses (inclusion method) with the chosen variables as possible predictors and the FIT 12 test score as dependent variable. In the first step, the control variables FIT 12 baseline score, Cognitive Scale score (BSID-III), and temperature were added. In the second step, we added the attentional aspects of temperament (Duration of playing from the TOT, Duration of orienting from the IBQ-R) as predictors. In the third step, we added Sociability (IBQ-R) as a social aspect of temperament. The final model (adjusted $R^2 = .45$, $F(6,128) = 19.43$, $p < .001$) is presented in Table 7. The FIT 12 test score was significantly predicted by the control variables FIT 12 baseline score, Cognitive Scale (BSID-III), and temperature, and by Sociability measured using the IBQ-R.

18 months

Again, we first calculated Pearson correlations (see Supplementary results) between the FIT 18 test score and attentional aspects of temperament, as well as social aspects of temperament, based on the theoretical assumptions. The following variables with a $p < .15$ were included as possible predictors: Baseline state (TOT), Mean positive motor activity (PGT), and High-intensity pleasure (ECBQ). The correlation between the FIT 18 test score and Sociability (ECBQ) missed the threshold for significance ($r = .13$, $p = .206$). For theoretical reasons, we decided to include Sociability as a possible predictor. No correlations with a $p < .15$ between the ECBQ measures assessing attentional aspects (subscales Attentional focusing and Attentional shifting, factor Effortful Control) or the five other TOTs and the FIT 18 test score occurred. No intercorrelations above .80 were found between the possible predictors from the same test procedures (see Supplementary results). We applied a hierarchical linear regression analysis (inclusion method) with the FIT 18 test score as dependent variable and the aforementioned attentional and social aspects of temperament as predictors. In the first step, we added the FIT baseline score at 18 months and the Cognitive Scale score (BSID-III) at 18 months as control variables. In the second step, the baseline score of the TOT as an attentional variable was added. In the third step, the above-mentioned social variables were added. In the final model (adjusted $R^2 = .11$, $F(6,93) = 2.94$, $p = .011$, see Table 8), the FIT 18 test score was significantly predicted by the TOT baseline state, Positive motor activity (PGT), and High-intensity pleasure (ECBQ). At 18 months, the control variables did not significantly predict the imitation test score.

24 months

Pearson correlations between the FIT 24 test score and attentional and social aspects of temperament revealed the following possible predictors of imitative behavior with a $p < .15$ Latency to look away (TOT), Effortful Control (ECBQ), Mean engagement behavior with toy

(PGT), Sociability (ECBQ), and Extraversion (ECBQ). There were no intercorrelations above .80 between the relevant possible predictors of the same test procedures. As the Sociability subscale (ECBQ) loads on the factor Extraversion (ECBQ), these two possible predictors were highly correlated ($r = .56$, $p < .001$). From a theoretical perspective, both predictors are relevant regarding infants' social orientation. Nevertheless, they should not be added to one model due to their high correlation. Thus, we conducted two hierarchical linear regression analyses (inclusion method) with the FIT 24 test score as dependent variable, FIT 24 Baseline scores and Cognitive Scale score (BSID-III) at 24 months as control variables, and the abovementioned variables as possible predictors. In the first step, we added the control variables, in the second step, we added the attentional variables (TOT: Latency to look away; ECBQ: Effortful Control), and in the third step, we added the social temperament variables. The third step differed between the two regression analyses: In the first analysis, we added Extraversion (ECBQ) and Mean engagement behavior with toy (PGT, adjusted $R^2 = .24$, $F(6,94) = 619$, $p < .001$, AIC = 270.37) and in the second analysis, we added Sociability (ECBQ) and Mean engagement behavior with toy (PGT) (adjusted $R^2 = .21$, $F(6,94) = 5.48$, $p < .001$, AIC = 273.68). Comparing both models' adjusted R^2s and AICs with the first model, the inclusion of Extraversion as a predictor yielded a higher adjusted R^2 and a smaller AIC, indicating that Extraversion is the better predictor of the FIT 24 test score. Accordingly, the final model (see Table 9) included Extraversion (ECBQ) and Mean engagement behavior with toy as social temperament predictors. The FIT 24 test score was significantly predicted by the FIT 24 baseline score, cognitive developmental status, and Extraversion (ECBQ). The attentional variables included in the model, and Engagement with toy (PGT), did not make a significant unique contribution to the model.

Longitudinal analyses

To further investigate the longitudinal relation between early temperament and later imitative behavior, we conducted hierarchical linear regressions. First, we analyzed the predictive power from temperament at T1 to imitative behavior at T2. Pearson correlations were calculated between PGT scores at T1 as well as IBQ-R subscales assessing social aspects of temperament ("Sociability", High-intensity pleasure, Approach, Smiling and laughter) and the FIT 18 test scores. No significant association emerged between early social aspects of temperament and imitative behavior at 18 months, and the correlations did not reach the threshold of $p < .15$ (all rs < .106, all ps > .304, $N = 94$). Even the relation between the FIT 18 test score and the control variables FIT 12 test score ($r = .17$, $p = .097$), Bayley scales score at T1 ($r = .16$, $p = .117$), and FIT 18 baseline score ($r = .09$, $p = .438$) failed to reach significance. Given that the possible predictors and the control variables were not related to the dependent variable, we did not conduct a hierarchical linear regression. Social aspects of temperament at T1 did not predict imitative behavior at T2.

Before running the hierarchical linear regression analysis, we calculated Pearson correlations between social temperament at T1 (PGT test scores and IBQ-R subscales "Sociability", High-intensity pleasure, Approach, Smiling and laughter) and the FIT 24 test score. No significant correlations emerged (all rs < .18, all ps > .077, $N = 98$). IBQ-R Smiling and laughter at 12 months, with a p-value < .15, was selected for inclusion in the regression analyses as a possible predictor. The hierarchical linear regression (inclusion method) was performed with the FIT 24 test score as dependent variable, Smiling and laughter (IBQ-R, T1) as predictor, and the FIT 12 test score, cognitive developmental status at T1, and FIT 24 baseline score as control variables. As shown in Table 10, only the control variables cognitive developmental status at 12 months and FIT 24 baseline score predicted imitative behavior in the FIT 24 (adjusted $R^2 = .24$, $F(3, 94) = 11.40$, $p < .001$). Adding 12-month-olds' extraversion in the second step did not significantly increase the adjusted R^2 value.

Before running the hierarchical linear regression to analyze longitudinal relations between T2 and T3, we calculated Pearson correlations between social aspects of temperament at T2 (PGT test scores and ECBQ subscales Sociability, High-intensity pleasure, Positive anticipation, and ECBQ factor Extraversion), including the following variables with a p-value < .15 as possible predictors: High-intensity pleasure (ECBQ), Positive anticipation (ECBQ), Sociability (ECBQ), Extraversion (ECBQ), and Mean presence of laughter (PGT). There were no intercorrelations above .80 between the possible predictors. As the factor Extraversion (ECBQ) mainly consists of the three subscales mentioned above, high correlations were found between Extraversion (ECBQ) and all three subscales: High-intensity pleasure ($r = .71$, $p < .001$), Positive anticipation ($r = .61$, $p < .001$), and Sociability ($r = .55$, $p < .001$). Again, from a theoretical perspective, all predictors regarding infants' social orientation are relevant. Nevertheless, due to their high correlations, these scales should not be added to one model with the factors they constitute. Thus, we conducted two hierarchical linear regression analyses (inclusion method) with the FIT 24 test score as dependent variable, FIT 24 baseline scores, FIT 18 test score and Cognitive Scale score (BSID-III) at 18 months as control variables, and the aforementioned variables as possible predictors. In the first step, we added the control variables. The second step differed between the two regression analyses. In the first analysis, we added Extraversion (ECBQ) and Mean presence of laughter (PGT, adjusted $R2 = .31$, $F(5,75) = 7.86$, $p < .001$, AIC = 187.13) and in the second analysis, we added High-intensity pleasure (ECBQ), Positive anticipation (ECBQ), Sociability (ECBQ), and Mean presence of laughter (PGT, adjusted $R^2 = .27$, $F(7, 68) = 4.86$, $p < .001$, AIC = 194.18). Comparing both models' adjusted R^2 and AIC with the first model, the inclusion of Extraversion as a predictor revealed a higher adjusted R^2 and a smaller AIC, indicating that Extraversion is the better predictor of the FIT 24 test score. Accordingly, the final model (see Table 11) included Extraversion (ECBQ) and Mean presence of laughter (PGT) as possible predictors. The control variables cognitive developmental status at 18 months and FIT 24 baseline score as well as Extraversion (ECBQ) and PGT mean presence of laughter predicted the imitative behavior in the FIT 24.

Table 7

Hierarchical Regression Analysis Predicting the Imitation Score at 12 months

Predictor	*B*	*SE B*	*β*	*p*	R^2	ΔR^2
Step 1					.44	.42***
Constant	-5.96	2.88				
FIT 12 baseline score	0.66	0.09	.47***	<.001		
Cognitive Scale (BSID-III)	0.05	0.01	.38***	<.001		
Temperature	0.19	0.10	.12	.07		
Step 2					.46	.44
Constant	-4.86	2.91				
FIT 12 baseline	0.63	0.10	.44***	<.001		
Cognitive Scale (BSID-III)	0.05	0.01	.39***	<.001		
Temperature	0.20	0.10	.13	.06		
Duration of orienting (IBQ-R)	-0.14	0.12	-.08	.25		
Duration of playing (TOT)	-0.01	0.00	-.11	.09		
Step 3					.48	.45*
Constant	-8.23	3.23				
FIT 12 baseline	0.61	0.09	.43***	<.001		
Cognitive Scale (BSID-III)	0.05	0.01	.38***	<.001		
Temperature	0.23	0.10	.15*	.03		
Duration of orienting (IBQ-R)	-0.22	0.13	-.12	.08		
Duration of playing (TOT)	-0.00	0.00	-.08	.21		
Sociability (IBQ-R)	0.46	0.21	.16*	.03		

Note. * $p < .05$. *** $p < .001$.

Table 8

Hierarchical Regression Analysis Predicting the Imitation Score at 18 months

Predictor	*B*	*SE B*	*β*	*p*	R^2	ΔR^2
Step 1					.03	.01
Constant	3.79	0.98				
FIT 18 baseline	0.12	0.17	.07	.48		
Cognitive Scale (BSID-III)	0.02	0.01	.15	.15		

Step 2					.07	.04
Constant	5.08	1.17				
FIT 18 baseline	0.08	0.17	.05	.62		
Cognitive Scale (BSID-III)	0.02	0.01	.16	.11		
Baseline state (TOT)	-0.60	0.31	-.19	.06		
Step 3					.16	.11*
Constant	5.60	1.62				
FIT 18 baseline	0.04	0.17	.03	.79		
Cognitive Scale (BSID-III)	0.01	0.01	.13	.18		
Baseline state (TOT)	-0.64	0.30	-.21*	.04		
Motor activity (PGT)	1.18	0.59	.20*	.05		
High pleasure (ECBQ)	-0.55	0.23	-.24*	.02		
Sociability (ECBQ)	0.39	0.21	.19	.07		

Note. * $p < .05$.

Table 9

Hierarchical Regression Analysis Predicting the Imitation Score at 24 months

Predictor	*B*	*SE B*	*β*	*p*	R^2	ΔR^2
Step 1					.20	.18
Constant	4.05	2.48				
FIT 24 baseline	0.81	0.20	.37***	<.001		
Cognitive Scale (BSID-III)	0.06	0.03	.20*	.04		
Step 2					.23	.20
Constant	10.12	4.40				
FIT 24 baseline	0.84	0.20	.38***	<.001		
Cognitive Scale (BSID-III)	0.05	0.03	.18*	.05		
Effortful control (ECBQ)	-1.31	0.78	-.15	.10		
Latency to look away (TOT)	0.02	0.02	.10	.28		
Step 3					.28	.24*
Constant	-2.15	6.33				
FIT 24 baseline	0.75	0.20	.34***	<.001		
Cognitive Scale (BSID-III)	0.06	0.03	.22*	.02		
Effortful control (ECBQ)	-1.18	0.77	-.14	.13		
Latency to look away (TOT)	0.03	0.02	.12	.19		
Extraversion (ECBQ)	1.49	0.70	.19*	.04		
Engagement with toy (PGT)	1.78	1.57	.11	.26		

Note. * $p < .05$. *** $p < .001$.

Table 10

Hierarchical Regression Analysis Predicting the Imitation Score Longitudinally from 12 to 24 months

Predictor	*B*	*SE B*	*β*	*p*	R^2	ΔR^2
Step 1					.27	.24***
Constant	2.20	2.11				
FIT 24 baseline	0.83	0.20	.38***	<.001		
FIT 12 test score	0.26	0.24	.11	.28		
Cognitive Scale (BSID-III, T1)	0.07	0.03	.25*	.02		
Step 2					.27	.24
Constant	0.72	2.89				
FIT 24 baseline	0.80	0.37	.37***	<.001		
FIT 12 test score	0.23	0.10	.10	.34		
Cognitive Scale (BSID-III, T1)	0.71	0.03	.25*	.02		
Smiling and laughter (IBQ-R, T1)	0.34	0.46	.07	.46		

Note. * $p < .05$. *** $p < .001$.

Table 11

Hierarchical Regression Analysis Predicting the Imitation Score Longitudinally from 18 to 24 months

Predictor	*B*	*SE B*	*β*	*p*	R^2	ΔR^2
Step 1					.21	.18***
Constant	5.56	2.36				
FIT 24 baseline	0.78	0.27	.37***	<.001		
FIT 18 test score	-0.17	0.23	-.08	.46		
Cognitive Scale (BSID-III, T2)	0.05	0.03	.23*	.04		
Step 2					.36	.31***
Constant	-4.98	4.05				
FIT 24 baseline	0.61	0.21	.29**	.005		
FIT 18 test score	-0.21	0.21	-.10	.32		
Cognitive Scale (BSID-III, T2)	0.06	0.02	.27*	.013		
Extraversion (ECBQ, T2)	2.17	0.75	.28*	.005		
Mean presence of laughter (PGT, T2)	-9.55	3.45	-.28*	.007		

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

The aim of the present study was to investigate the developmental course of the two functions of imitation throughout the second year of life by testing infants' imitative behavior and temperament at 12, 18, and 24 months. First, we found that imitative behavior was related to social aspects of temperament throughout the entire second year of life. In contrast, we did not find consistent evidence for an association between attention and imitation in this age range. These results suggest that within the second year of life, no change from an instrumental to a social motivation for imitation occurs. Instead, imitation serves a social function from early on, which increases in strength over time. Furthermore, these postulated associations were predominantly based on parent reports of infant temperament, whereas the behavioral measures did not show the same patterns. Finally, longitudinal analyses revealed no significant association between early social orientation and later imitative behavior.

Our first hypothesis was that infants' social orientation would be unrelated to their imitative behavior at 12 months, but that such an association would be apparent at 18 months and would be even stronger at 24 months. However, infants' social orientation was related to their imitative behavior throughout their second year. The association between Extraversion and imitation later in the second year of life corresponds to previous findings. For instance, Hilbrink et al. (2013) showed that 15-month-olds with higher levels of Extraversion were more exact imitators than 15-month-olds with lower levels of extraversion, and Dixon et al. (2012) reported an association between Extraversion and imitation in 21-month-olds. At first glance, our finding of such an association at 12 months seems to contradict the findings of Zmyj et al. (2017), who reported no relation between the social aspects of temperament and imitative behavior in 12-month-olds. However, upon closer inspection, this discrepancy might be explained by a methodological difference: In the study by Zmyj et al. (2017), imitative behavior was not correlated with the IBQ-R subscales (Approach, Smiling and laughter, Vocal reactivity) and the PGT (Lab-TAB), a finding which we replicated in the present study. These IBQ-R subscales assess infants' social orientation in social situations and facets of positive affect that are unrelated to social motivation. The present study revealed that the rationally derived Sociability scale, which assesses seeking and taking pleasure in interactions with others, was related to imitation at 12 months. Thus, our analyses suggest that infants' social motivation to interact with others already affects their imitative behavior at 12 months. This finding adds to a growing body of literature emphasizing the importance of children's social motivation in imitation and of "view[ing] imitation through the social lens" (Over & Carpenter, 2013, p. 9).

The second hypothesis was that attentional aspects of temperament would be associated with imitative behavior throughout the entire second year of life. We found no relation between imitation and attentional aspects of temperament measured through questionnaires and observation at 12, 18, and 24 months. Attention measured using questionnaires was not related to imitative behavior at 12 months (Duration of orienting, IBQ-R) or at 18 and 24 months (Attentional focusing, Attentional shifting, Effortful Control, ECBQ). In contrast to Zmyj et al. (2017), we did not find an association between Duration of orienting and imitative performance at 12 months. These divergent findings correspond to previous research, which likewise revealed a mixed picture regarding the relation between attention and imitation (see Heimann et al., 2022). For instance, one study found that Effortful control, which encompasses attentional processes, was not related to imitation among 15-month-olds (Dixon et al., 2012), in line with the present findings. Furthermore, infants' looking time at a model, which is often used as a measure of overt attention while watching the demonstrations of the target acts (e.g., Óturai et al., 2013), was not predictive of their imitative behavior at 12 months (Kolling et al., 2014) and at 18 months (Óturai et al., 2013) in studies using eye tracking. Sonne et al. (2016) found a negative correlation between looking time and memory performance in 20-month-olds. The discrepant findings between the present study and the study by Zmyj et al. (2017) might be explained by the fact that we added infants' cognitive developmental status as a control variable, whereas Zmyj et al. (2017) did not assess cognitive development.

The third hypothesis was that the relation between temperament and imitation would remain constant after controlling for infants' cognitive development. The results revealed that at 12 and 24 months, cognitive development was a predictor of imitative behavior, but no attentional aspect emerged as a predictor. At 18 months, cognitive development did not predict imitation; instead, a behavioral measure of attention (baseline state in the TOT) was a significant predictor. Contrary to our hypothesis, the relation between attention, especially when measured via questionnaires, and imitation disappeared when controlling for infants' cognitive development. Thus, attentional processes measured via parent report might not be suitable to detect the cognitive function of imitation without additionally considering the infants' cognitive abilities. Seemingly, the cognitive dimensions of temperament measured using the IBQ-R do not cover all aspects of infants' instrumental motivation. Nevertheless, the use of both scales partially enabled us to differentiate between individual characteristics and cognitive abilities.

Measuring attention in children is complicated by the fact that different methods, that is, ratings of attention and performance-based measures, may tap into different levels and parts of attentional processes (Acar et al., 2019). An advantage of the present study is that we used a multi-method approach to assess infants' attention, which is especially relevant given that parent reports and laboratory tasks of attention are not correlated (Acar et al., 2019). Acar et al. (2019) even found that parent ratings did not load on a latent factor of focused attention, while secondary caregiver ratings, observer ratings, and behavioral tasks did, indicating that parental ratings do not assess the same construct of attention as the other measures. Nevertheless, neither of the attention measures in the present study, from different domains, predicted imitative behavior at all ages. Duration of Orienting (IBQ-R) and Attentional Focusing (ECBQ), which were employed in our study as parental ratings of attention, assess whether parents rate their infants as attending to one activity for several minutes or longer. In view of the short demonstrations of the target acts in the demonstration phases, longer ratings of sustained attention to objects may not be the relevant part of attention for the performance in imitation tasks. The TOT, as a behavioral task, did not predict imitative behavior, with the exception of the baseline state at 18 months. Furthermore, we used only one behavioral measure of infants' attention. As discussed previously, looking time, an indicator of attention in behavioral tasks such as the TOT from the Lab-TAB and measured via eye tracking, is not predictive of imitative behavior (Kolling et al., 2014; Óturai et al., 2013; Sonne et al., 2016). As behavioral tasks and parent reports measure different parts of attention that did not predict imitative behavior, another method to detect infants' perception of the actions during demonstration might yield more conclusive findings regarding the relation between attention and imitation in infants. Looking time at the whole screen when an action is demonstrated on a computer (e.g. "Overall AoI", Sonne et al., 2016) is not informative with respect to where, precisely, the child is looking and at what point during the demonstration (action, goal, demonstrator). Thus, investigating infants' anticipatory looking behavior during the demonstration phase using eye tracking to assess action understanding might be more suitable to tap into the relation between attentional processes and imitative behavior.

The fourth hypothesis was that infants' early social orientation (at 12 months) would predict their imitative behavior at 18 and 24 months. The exploratory analyses did not reveal a longitudinal association between early social motivation and later imitative behavior. To the best of our knowledge, only one previous study has reported longitudinal relations between early social orientation and later imitative behavior (Dixon et al., 2012). We found a longitudinal association between social orientation at 18 months and imitative behavior at 24 months, insofar as the ECBQ factor Extraversion and the Mean presence of laughter in the PGT at 18 months predicted imitative behavior at 24 months. Contrary to our expectation, however, the Mean amount of laughter was inversely related to imitative behavior. Our findings of longitudinal associations from the middle of the second year of life to the end of the second year align with the findings of Dixon et al. (2012), who reported that Extraversion at 15 months predicted long-term recall at 21 months. However, these longitudinal relations must be interpreted with caution, as the latter authors applied two different imitation tasks but only found longitudinal associations in one task. Moreover, the association was affected by the inclusion of distractors. Furthermore, the psychometric properties of these imitation tasks are unknown and the study was based on a relatively small sample ($N = 32$). The results of our systematic study with a larger sample size and age-appropriate imitation tests with good psychometric properties suggest that early social motivation is not related to later imitative behavior. Thus, early social orientation might not be related to imitative behavior from the beginning of the second year of life to the end of the second year. However, according to our findings, this relation is present towards the end of the second year. Due to the in part inversely related predictors and the aforementioned problems in the study by Dixon et al. (2012), longitudinal relations need to be investigated in further studies.

A major advantage of the present study is that we used a multi-method approach to assess temperament. The postulated relations discussed above are based on parental ratings of temperament, whereas imitative behavior was not related to behavioral measures of temperament, with the exception of baseline state behavior in the TOT and positive motor activity in the PGT, which predicted imitative behavior at 18 months. Eighteen-month-olds who were calm prior to the blocks task showed higher imitation scores, indicating that the ability to concentrate and attend to a new task promotes imitation. Nevertheless, the discrepancy between parental and behavioral measures of temperament are unsurprising given that only moderate convergence has been found between parental report and observational laboratory tasks of temperament (e.g., Majdandzic & van den Boom, 2007). Indeed, with regard to attention, it is questionable whether parental report and observational tasks measure the same latent construct of focused attention (Acar et al., 2019). Thus, one explanation for our findings is that the different measures might depict different parts of infant temperament, although it is still important to use a multi-method approach to study temperament in order to cover the whole construct of infant temperament (Vroman et al., 2014). Furthermore, we applied only one behavioral task for each dimension relevant to our study (PGT to assess social orientation, TOT to assess attention). To gain more insight into the association between temperament and imitation, future research should therefore employ more than one task per temperament domain in order to explore the links between behavioral assessments and imitation in greater depth.

The imitation tests employed in the present study may be criticized because the total number of items and target actions, as well as the number of demonstrations, differed between the three time points. However, these differences do not impact the comparability of the data across the time points. In longitudinal studies covering periods of rapid development, it is necessary, from a test-theoretical perspective, to adapt the tests across the different time points (Kolling et al., 2010). Empirical findings from longitudinal studies using the FITs support the use of the tests to assess imitative behavior from a longitudinal perspective (Kolling et al., 2010; Óturai et al., 2018). When comparing longitudinal studies using the FITs to longitudinal studies using other imitative procedures, it is apparent that these latter studies also changed the target objects across tests and time points (Heimann et al., 2006; Nielsen & Dissanayake, 2004). Thus, as a series of imitation tests that are age-adapted in terms of the degree of difficulty, the FITs are adequate to assess imitative behavior longitudinally within the second year of life, even if they differ slightly across time points.

It is noteworthy that the results at 18 months differ slightly from those at 12 and 24 months: While imitative behavior at 12 and 24 months was predicted by the control variables imitative baseline behavior and cognitive development, as well as social orientation measured via questionnaire (sociability at T1, Extraversion at T3), at 18 months, the imitation test performance was not related to baseline behavior in the FIT 18 or to cognitive developmental status. Instead, imitative behavior was predicted by behavioral variables, which was not the case at 12 and 24 months. Furthermore, the amount of explained variance was considerably lower at 18 months ($\Delta R^2 = .11$) compared to 12 months ($\Delta R^2 = .45$) and 24 months ($\Delta R^2 = .24$). Moreover, at 18 months, a further 40 infants had to be excluded from the analyses, 24 of them due to crying. Some developmental theorists regard 18 months as an age that encompasses a qualitative shift in cognitive and socioemotional development (e.g., Perner, 1991; Bischof-Köhler, 2000). Accordingly, this structural change in development might obscure the relation between temperament and imitation.

Contrary to expectation, the mean score on the Cognitive Scale (BSID-III) at 12 months was more than one standard deviation below the norm of this scale. This finding might be attributable to fatigue, as the session lasted for approximately one hour for all ages, and the Cognitive Scale was applied last. Additionally, it was relatively warm in the laboratory at T1.

Importantly, to the best of our knowledge, the present study is the first to systematically investigate the developmental course of the two functions of imitation within the second year of life using standardized tests of imitation and temperament. The results of our study did not reveal a change from the dominance of the cognitive function at the beginning of the second year of life to the dominance of the social function at the end of the second year of life, as suggested by Užgiris (1981). A recent experimental study by Gellén & Buttelmann (2019) supported the idea of a change in the underlying motivations to imitate within this age range. The authors tested selective imitation in 14-, 18-, 24-, and 36-month-olds, and found that selective imitation decreased from 14 months onwards, whereas exact imitation increased. Simultaneously, beyond the age of 14 months, children showed an increasing tendency to gaze at the model while imitating the target actions. The authors interpreted these findings as evidence of a change in the underlying motivation to imitate. However, as Gellén and Buttelmann (2019) only employed the imitation task, and did not apply tests of the underlying motivation, their interpretation remains speculative. Increased looking time at the model might indicate increased social orientation. However, in the absence of other behavioral information, such as whether the infants were willing to engage in contact with the demonstrator and initiated contact, the reasons why older infants displayed longer looking time at the demonstrator remain unclear. Looking time was coded as "they looked at the model either while imitating the target action or within the following 2 s after breaking contact between the target body part and the apparatus" (Gellén & Buttelmann, 2019, p. 155). Another possible explanation for longer looking times might lie in performance, insofar as toddlers wanted to ensure that they solved the task correctly. As the authors stated, gaze behavior has not been previously used to measure social motivation. Thus, as the looking duration towards the model was the only measure of social motivation, this finding cannot be conclusively interpreted.

Further support for the notion that the social function develops after the instrumental function within the second year of life has been provided by Nielsen and colleagues (e.g., Nielsen, 2006), who found that 12-month-olds predominantly imitated the outcome of an action whereas 18- and 24-month-olds imitated both the action and the outcome, and 18- and 24-month-olds showed higher imitation rates when the model acted socially. In line with the latter finding, the vast majority of research on the onset of the social function of imitation has investigated the social influence on copying behavior by varying the social context of the situation (e.g., communicative vs. non-communicative cues, Kupán et al., 2017; model acting socially vs. aloof, Nielsen, 2006). By contrast, our study was the first to systematically consider children's individual characteristics concerning their imitative behavior rather than changing the social context of the imitative situation. We aimed to investigate the relation between children's individual characteristics and their imitative behavior in order to capture their general motivation to engage in imitation without external influences. Therefore, we excluded the performance aspect of imitative behavior by controlling for cognitive development and the influence of situational characteristics by choosing an instrumental task without varying the social context. As such, we assessed children's common motivation to engage in imitative behavior via temperament.

Furthermore, the task characteristics of the FITs used in our study differ slightly from the imitation tasks used by Nielsen (2006) and Óturai et al. (2018). Whereas the latter tasks were designed to test whether infants and toddlers imitate action steps that are not necessary to achieve a goal, we assessed instrumental imitative behavior. In line with Nielsen (2006), we found a relation between social orientation and imitation at 18 and 24 months (see also Óturai et al., 2018). Thus, the results provide a coherent picture that social orientation and imitation are related at least from 18 months on. Accordingly, while we do not reject the idea that children might pay more attention to social cues with increasing age, our results suggest that the social function of imitation in infants is present early on. Although older children might be more receptive to social cues, younger children's social orientation already predicts their imitative behavior at 12 months. In this vein, Over and Carpenter (2012) argued that the dominance of one function over the other depends on the context and that imitation itself is a profoundly social process.

Finally, we examined the two functions of imitation by investigating the relation between characteristics and imitative behavior within the second year of life. In the current study, infants' temperament was used to mirror their social and cognitive motivation to engage in imitative behavior. We consider our study as the first attempt to systematically investigate the development of the functions of imitation without experimentally manipulating social or instrumental cues. However, in view of the difficulties discussed above, especially in terms of assessing infants' attention, the sole consideration of individual differences might not be sufficient to reflect both functions.

To conclude, the results of our study suggest that the social function of imitation matters from early in life. Our findings suggest that children's characteristics are related to their imitative behavior. As every imitative situation is embedded in a social context, future studies of imitation should consider children's motivation. The lack of consistent evidence for an influence of cognitive capacities on imitation throughout the entire second year of life may indicate the need for future research to measure instrumental motivation independently of children's cognitive development.

Acknowledgements

The German Research Foundation (DFG, ZM 54/6-1) funded this research. We thank our student assistants Maximilian Schauerte, Gero Noßke, Sandra Doods, Svenja Haas, Lena Lohkämper, Sina Hamester, and two additional student assistants for coding the videotapes. Finally, we thank all children and parents for participating in this study.

References

Acar, I. H., Frohn, S., Prokasky, A., Molfese, V. J., & Bates, J. E. (2019). Examining the associations between performance based and ratings of focused attention in toddlers: Are we measuring the same constructs? *Infant and Child Development*, *28*(1), e2116. https://doi.org/10.1002/icd.2116

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior & Development*, *19*(2), 159–170. http://dx.doi.org/10.1016/S0163-6383(96)90015-6

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). Psychological Cooperation. https://doi.org/10.1037/t14978-000

Bischof-Köhler, D. (2000). Empathy, prosocial behavior and security of attachment in two years olds. *Psychologie in Erziehung und Unterricht*, *47*, 142-158.

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child development*, *78*(3), 806-824. https://doi.org/10.1111/j.1467-8624.2007.01034.x

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, *84*(2), 422-428. https://doi.org/10.1111/j.1467-8624.2012.01860.x

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2), 315-330. https://doi.org/10.1016/S0163-6383(98)90009-1

Carpenter, M., & Call, J. (2009). Comparing the imitative skills of children and nonhuman apes. *Revue de Primatologie*, *1*(6). https://doi.org/10.4000/primatologie.263

Dixon, W. E., Jr, Lawman, H. G., Johnson, E. B. H., May, S., Patton, L. A., Lowe, A. K., & Snyder, C. M. (2012). Effects of exogenous and endogenous distracters on immediate and long-term recall in toddlers. *Infancy*, *17*(5), 525-557. https://doi.org/10.1111/j.1532-7078.2011.00090.x

Dixon Jr, W. E., Salley, B. J., & Clements, A. D. (2006). Temperament, distraction, and learning in toddlerhood. *Infant Behavior and Development*, *29*(3), 342-357. https://doi.org/10.1016/j.infbeh.2006.01.002

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fenstermacher, S. K., & Saudino, K. J. (2006). Understanding individual differences in young children's imitative behavior. *Developmental Review*, *26*(3), 346-364. https://doi.org/10.1016/j.dr.2006.05.001

Field, A. (2018). *Discovering Statistics Using IBM SPSS statistics* (5th ed.). Sage Publications.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behavior and Development*, *26*(1), 64-86. https://doi.org/10.1016/S0163-6383(02)00169-8

Gellén, K., & Buttelmann, D. (2019). Rational imitation declines within the second year of life: Changes in the function of imitation. *Journal of Experimental Child Psychology*, *185*, 148-163. https://doi.org/10.1016/j.jecp.2019.04.019

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*, 755. https://doi.org/10.1038/415755a

Goertz, C., Knopf, M., Kolling, T., Frahsek, S., & Kressley, R. (2006). Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12): Erprobung einer Aufgabenserie zur Erfassung des deklarativen Gedächtnisses Einjähriger. *Zeitschrift für*

Entwicklungspsychologie und Pädagogische Psychologie, 38(2), 88–96. https://doi.org/10.1026/0049-8637.38.2.88

Goertz, C., Kolling, T., Frahsek, S., & Knopf, M. (2008). Die Frankfurter Imitationstests für 18 und 24 Monate alte Kinder: Entwicklung altersangepasster Gedächtnistests. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *40*(3), 152-160. https://doi.org/10.1026/0049-8637.40.3.152

Goldsmith, H. H., & Rothbart, M. K. (1999). *Laboratory Temperament Assessment Battery (Lab-TAB): Locomotor version 3.1.* Personality Development Laboratory, Department of Psychology, University of Wisconsin, Madison, WI.

Heimann, M. (2002). Notes on individual differences and the assumed elusiveness of neonatal imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The Imitative Mind: Development, Evolution and Brain Bases* (pp. 74–84). Cambridge University Press.

Heimannn, M. (2022). *Imitation from Infancy Through Early Childhood*. Springer Cham. https://doi.org/10.1007/978-3-031-08899-5

Heimann, M., & Meltzoff, A. N. (1996). Deferred imitation in 9-and 14-month-old infants: A longitudinal study of a Swedish sample. *British Journal of Developmental Psychology*, *14*(1), 55-64. https://doi.org/10.1111/j.2044-835X.1996.tb00693.x

Heimann, M., Strid, K., Smith, L., Tjus, T., Erik Ulvund, S., & Meltzoff, A. N. (2006). Exploring the relation between memory, gestural communication, and the emergence of language in infancy: A longitudinal study. *Infant and Child Development*, *15*(3), 233-249. https://doi.org/10.1002/icd.462

Hilbrink, E. E., Sakkalou, E., Ellis-Davies, K., Fowler, N. C., & Gattis, M. (2013). Selective and faithful imitation at 12 and 15 months. *Developmental Science*, *16*(6), 828-840. https://doi.org/10.1111/desc.12070

IBM Corp. (2021). *IBM SPSS Statistics for Windows* (Version 28.0) [Computer software]. IBM Corp.

Jagiello, R., Heyes, C., & Whitehouse, H. (2022). Tradition and invention: The bifocal stance theory of cultural evolution. *Behavioral and Brain Sciences*, *45*, e249. https://doi.org/10.1017/S0140525X22000383

Kirchhoff, C., Gartstein, M., Iverson, S., Reed-Smith, H., Brekke, A., & Cress, E. (2013). *Temperament and behavioral difficulties across cultures: A comparison between German and US toddlers*. Paper presented at the Occasional Temperament Conference, Salt Lake City, UT, USA.

Kolling, T., Goertz, C., Stefanie, F., & Knopf, M. (2010). Memory development throughout the second year: Overall developmental pattern, individual differences, and developmental trajectories. *Infant Behavior and Development*, *33*(2), 159-167. https://doi.org/10.1016/j.infbeh.2009.12.007

Kolling, T., Óturai, G., & Knopf, M. (2014). Is selective attention the basis for selective imitation in infants? An eye-tracking study of deferred imitation with 12-month-olds. *Journal of Experimental Child Psychology*, *124*, 18-35. https://doi.org/10.1016/j.jecp.2014.01.016

Kupán, K., Király, I., Kupán, K., Miklósi, Á., & Topál, J. (2017). Interacting effect of two social factors on 18-month-old infants' imitative behavior: Communicative cues and demonstrator presence. *Journal of Experimental Child Psychology*, *161*, 186-194. https://doi.org/10.1016/j.jecp.2017.03.019

Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences*, *19*(11), 688-699. https://doi.org/10.1016/j.tics.2015.08.005

Majdandžić, M., & Van Den Boom, D. C. (2007). Multimethod longitudinal assessment of temperament in early childhood. *Journal of Personality*, *75*(1), 121-168. https://doi.org/10.1111/j.1467-6494.2006.00435.x

Mink, D., Henning, A., & Aschersleben, G. (2013). Visual habituation tasks: The role of infant temperament. *Infant Behavior and Development*, *36*(3), 377-390. https://doi.org/10.1016/j.infbeh.2013.03.005

Nielsen, M. (2006). Copying actions and copying outcomes: social learning through the second year. *Developmental Psychology*, *42*(3), 555-565. https://doi.org/10.1037/0012-1649.42.3.555

Nielsen, M. (2009). The imitative behaviour of children and chimpanzees: A window on the transmission of cultural traditions. *Revue de Primatologie*, *1*(5). https://doi.org/10.4000/primatologie.254

Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior and Development*, *27*(3), 342-365. https://doi.org/10.1016/j.infbeh.2003.12.006

Óturai, G., Kolling, T., & Knopf, M. (2013). Relations between 18-month-olds' gaze pattern and target action performance: A deferred imitation study with eye tracking. *Infant Behavior and Development*, *36*(4), 736-748. https://doi.org/10.1016/j.infbeh.2013.08.005

Óturai, G., Kolling, T., & Knopf, M. (2018). Developmental trend towards exact imitation in the second year of life: Evidence from a longitudinal study. *International Journal of Behavioral Development*, *42*(4), 388-395. https://doi.org/10.1177/0165025417713727

Over, H., & Carpenter, M. (2012). Putting the social into social learning: explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, 126(2), 182-192. https://doi.org/10.1037/a0024555

Over, H., & Carpenter, M. (2013). The social side of imitation. *Child Development Perspectives*, *7*(1), 6-11. https://doi.org/10.1111/cdep.12006

Paulus, M. (2011). Imitation in infancy: Conceptual considerations. *Theory & Psychology*, *21*(6), 849-856. https://doi.org/10.1177/0959354310395990

Perner, J. (1991). *Understanding the representational mind*. MIT Press.

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development*, *29*(3), 386-401. https://doi.org/10.1016/j.infbeh.2006.01.004

Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and Child Development: An International Journal of Research and Practice, 17*(4), 387–405. https://doi.org/10.1002/icd.582

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of Child Psychology: Social, Emotional, and Personality Development* (6th ed., Vol.3, pp. 99–166). John Wiley & Sons, Inc..

Rothbart, M. K. & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds.), *Advances in Developmental Psychology* (Vol. 1, pp. 37–86). Erlbaum.

Seehagen, S., & Herbert, J. S. (2011). Infant imitation from televised peer and adult models. *Infancy*, *16*(2), 113-136. https://doi.org/10.1111/j.1532-7078.2010.00045.x

Sieber, F. (2023, November 14). The Two Functions of Imitation in the Second Year of Life. Retrieved from osf.io/zwqty

Sieber, F., & Zmyj, N. (2022). Stability and structure of infant and toddler temperament in two longitudinal studies in Germany. *Infant Behavior and Development*, *67*, 101714. https://doi.org/10.1016/j.infbeh.2022.101714

Sonne, T., Kingo, O. S., & Krøjgaard, P. (2016). Empty Looks or Paying Attention? Exploring Infants' Visual Behavior during Encoding of an Elicited Imitation Task. *Infancy*, *21*(6), 728-750. https://doi.org/10.1111/infa.12141

Užgiris, I. C. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development*, *4*(1), 1-12. https://doi.org/10.1080/00221325.1981.10534136

Vonderlin, E., Ropeter, A., & Pauen, S. (2012). Assessment of temperament with the Infant Behavior Questionnaire Revised (IBQ-R)-the psychometric properties of a German version. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, *40*(5), 307-314. https://doi.org/10.1024/1422-4917/a000187

Vroman, L. N., Lo, S. L., & Durbin, C. E. (2014). Structure and convergent validity of children's temperament traits as assessed by experimenter ratings of child behavior. *Journal of Research in Personality*, *52*, 6-12. https://doi.org/10.1016/j.jrp.2014.06.002

Zmyj, N., Schölmerich, A., & Daum, M. M. (2017). The relationship between attention and deferred imitation in 12-month-old infants. *Infant Behavior and Development*, *48*, 175-183. https://doi.org/10.1016/j.infbeh.2017.02.001

Supplementary Materials

Supplementary Methods

Exclusion Criteria at Each Measurement Time Point

Table S1

Number of Participants Excluded at Each Time Point and Reasons for Exclusion

Exclusion criterion	T1	T2	T3
Infants' crying	9	24	14
Experimenter error	2	2	0
Absence from laboratory session	1	0	6
More than 20% missing questionnaire values	9	13	5
Technical problems	0	1	0

Materials of the Deferred Imitation Tests

In the deferred imitation test at T1 (FIT 12, Goertz et al., 2006), the following material was used: two tin cans (height 8 cm, diameter 5 cm), one filled with rice in the demonstration phase and an empty one in the baseline and test phase; a pink stuffed toy pig (height 16 cm) wearing a brown hat attached by Velcro; a red plastic cup (height 8 cm, diameter 8.5 cm) with a wooden toy knife (length 13 cm); a wooden mouse (height 4 cm, length 6 cm, width 4 cm); a red buzzer (height 2 cm, length 13.5 cm, width 13.5 cm) that is activated using a wooden stick (length 18 cm) attached to the side of the buzzer. The buzzer and the stick differed from the original material because the original items were no longer commercially available, but allowed for the same target action (i.e. making a sound when the stick is used to activate the buzzer).

The materials used in the deferred imitation test at T2 were again as close as possible to the original material of the FIT 18 (Goertz et al., 2008): a yellow stuffed car (height 12 cm, length 13.5 cm) that can be used as a glove; a white stuffed goose (height 15 cm) with a tin can (height 8 cm, length 5 cm, width 5 cm) onto which the goose can be clipped; a wooden mouse (height 4 cm, length 6 cm, width 4 cm); a black and white stuffed panda (height 8 cm, length 4 cm) and a green ring (diameter 8 cm); a red buzzer (height 2 cm, length 13.5 cm, width 13.5 cm) that is activated with a wooden stick (length 18 cm) attached to the side of the buzzer; a yellow duck (height 6.5 cm, length 5 cm, width 4.5 cm) and a yellow octopus (diameter 8.5 cm) onto which the duck can be placed.

In the deferred imitation test at T3, we used the following material as close as possible to the FIT 24 (Goertz et al., 2008): a white plastic gondola (height 6 cm, length 14.5 cm, width 8 cm) with a wooden spoon (length 30 cm) and a black manikin (height 11.5 cm); a yellow and blue boat (height 7.5 length 15 cm, width 7 cm), with a blue tube (height 10 cm) in a blue sheath hanging on it, and the tube containing a manikin (height 5 cm); a black and white stuffed panda (height 8 cm, length 4 cm) with a wooden board (height 20.5 cm, length 10 cm) and a wooden stand (height 12 cm, length 15.5 cm, width 10 cm); a red soft ball (diameter 4 cm) with a slot and plastic eyes (height 1.5 cm, length 4 cm); a wooden turtle with rollers (height 10 cm, length 16 cm, width 9 cm) and a blue ball (diameter 8 cm) that is cut in half and clipped onto a green and yellow striped cone (height 12 cm, diameter 6.5 cm); a bunny (height 19 cm, length 16 cm, width 6 cm), with three Velcro strips onto which a round yellow pillow (diameter 10 cm), a square green pillow (height, length 11 cm), and a triangular rose pillow (base 11 cm, legs 13 cm) can be attached, with a black pillow ball (diameter 6 cm), a pink ribbon (height 5 cm, length 11 cm), and a white flower-shaped pillow (diameter 9 cm) as distractors; a wooden box (height 13 cm, length, width 13 cm) with 26 floral stickers and a hook on one side onto which a ring of cardboard (diameter 10 cm) fits, and the box can be opened on top with a white handle and has a drawer with a little bird inside it; a white magnetic plate (diameter 25 cm) with three magnetic buttons (black, red, yellow, diameter 3.5 cm) and a magnetic croissant (height 4.5 cm, length 8 cm), with two magnetic plastic sausages (height 2 cm, length 6 cm) as distractors.

“Sociability” Scale of the IBQ-R

Table S2

Overview of Items Included in the Rationally Derived Sociability Scale Including their Original Scale Assignment

Item number	Item	Original scale
104	How often during the week did your baby seem excited when you or other adults acted in an excited manner around him/her?	Approach
172	When familiar relatives/friends visited, how often did the baby get excited?	Approach
173R	When familiar relatives/friends visited, how often did the baby seem indifferent?	Approach
58	How often during the last week did the baby smile or laugh when tickled?	High Pleasure
65	How often during the last week did your baby enjoy being tickled by you or someone else in your family?	High Pleasure
66	How often during the last week did your baby enjoy being involved in rambunctious play?	High Pleasure
67	How often during the last week did your baby enjoy watching while you, or another adult, playfully made faces?	High Pleasure
77	When tossed around playfully how often did the baby smile?	High Pleasure
78	When tossed around playfully how often did the baby laugh?	High Pleasure
79	During a peekaboo game, how often did the baby smile?	High Pleasure
80	During a peekaboo game, how often did the baby laugh?	High Pleasure
81	How often did your baby enjoy bouncing up and down while on your lap?	High Pleasure
165	When in a crowd of people, how often did the baby seem to enjoy him/herself?	High Pleasure
34	When being dress or undressed during the last week, how often did the baby smile or laugh?	Smiling and Laughter
36	When put into the bath water, how often did the baby smile?	Smiling and Laughter
37	When put into the bath water, how often did the baby laugh?	Smiling and Laughter
40	When face was washed, how often did the baby smile or laugh?	Smiling and Laughter
43	When face was washed, how often did the baby smile?	Smiling and Laughter
149	When you returned from having been away and the baby was awake, how often did s/he smile or laugh?	Smiling and Laughter

Target Actions of Deferred Imitation Tests

FIT 12

We presented the following target actions in the FIT 12 at T1, while the infant sat on the caregiver's lap facing the experimenter. The experimenter stood at the opposite side of the table and presented the target actions on a wooden box. In the *Tin Can task*, he shook the tin can up and down three times, resulting in a rattling sound. In the *Pig task*, the experimenter removed the hat from the pig's head and then put it back on. In the *Cup & Knife task*, the cup was turned upside down on the box and the knife was placed next to it. The experimenter turned the cup into an upright position, inserted the knife and knocked four times against the opposite sides of the cup. In the *Mouse task*, the experimenter pressed the top of the mouse down. In the *Drum task*, the experimenter took the stick that was attached to the side of the apparatus and used it to push the button twice, resulting in a buzzer sound. The sound only occurred in the demonstration phase and was switched off in the baseline and test phase.

FIT 18

We presented the following target actions in the FIT 18 at T2, while the infant sat on the caregiver's lap facing the experimenter. The experimenter stood at the opposite side of the table and presented the target actions on a wooden box. In the *Car task*, he put his hand into the car that was placed on the box and waved with it twice. In the *Goose & Tin Can task*, the goose and the tin can were placed next to each other on the box. The experimenter clipped the tin can onto the goose's belly and placed it on the box. In the *Mouse task*, the experimenter pressed the top of the mouse down. In the *Panda & Ring task*, the experimenter turned the panda upside down, bounced the panda with its head in front on the box making a headstand, and placed it back upright on the box. He sat the panda into the ring and slid it from left to right. The *Drum task* was identical to the FIT 12 *Drum task*. In the *Octopus task*, the experimenter placed the duck on the octopus and turned the duck on the octopus around so that it was facing the child and back again.

FIT 24

We presented the following target actions in the FIT 24 at T3, while the infant sat on the caregiver's lap facing the experimenter. The experimenter stood at the opposite side of the table and presented the target actions on a wooden box. In the *Gondola task*, the experimenter placed the manikin in the gondola. He leaned the spoon against the manikin and then slid the gondola back and forth across the table. In the *Boat & Box task*, the experimenter removed the tube from the sheath, opened it, and took the manikin off the tube. He bent the legs of the manikin into a sitting position and placed it into the boat. In the *Panda & Slide task*, the experimenter leaned the board against the stand, bounced the panda on the socket of the stand and then slid the panda down the board. In the *Ball with Eyes task*, the experimenter placed the eyes in the slots of the ball and bounced the ball up and down. In the *Turtle task*, he clipped the half-cut ball on the cone and put them both on the back of the turtle. He then let the turtle fly up and down once. In the *Bunny task*, first, the experimenter picked up the bunny and attached the yellow pillow onto the rabbit's front, second, he attached the green pillow onto the rabbit's head, and third, he attached the rose pillow onto the rabbit's back. In the *Box task*, the experimenter hung the ring onto the hook on the box and turned the ring around. He turned the box around and put the ring back on the wooden box. Then, he opened the drawer and the bird appeared. In the *Magnetic Plate task*, the experimenter turned the plate around and attached the magnetic buttons in the following order from top down: the red button, the yellow button, the black button, and the croissant. He lifted the plate and rotated it back and forth.

Coding of Deferred Imitation Tests

FIT 12

In the FIT 12, infants were given one or two points if they produced the respective target acts depending on the task. For the Tin can task, infants could receive one point if they quickly moved the can in one direction and moved it back in the same trajectory. Banging the can on the table was not counted as a produced target act. For the Pig task, infants could receive two points. One point was given if the hat was removed from the pig or if the infant at least tried to do so. Another point was given if the hat was put back onto the pig's head. This second point was not in the original version of the FIT 12 (Goertz et al., 2006). For the Cup & Knife task, infants could receive two points. They received one point if they put the cup back in its upright position with its opening at the top. They received another point if they put the knife into the cup. In the Mouse task, infants received one point if they pressed the top of the mouse down. The mouse had to be placed on the table when the infants produced this action. Taking the mouse into one hand and squeezing it together was not counted as target act. In the Drum task, infants could receive two points. If infants removed the stick from the side of the apparatus, they received one point. If they touched the button with the stick, they received another point.

FIT 18

In the FIT 18, infants were given one to three points, depending on task, if they produced the respective target acts. For the Car task, children could receive two points. One point was given if they put single fingers or their whole hand into the car. Another point was given if they intentionally waved the car with their fingers/ hand inside the car.

In the Goose and tin can task, infants could receive two points. One point was given if the goose was clipped onto the tin can, even if the tin can did not stay there. Accidental tripping Accidentally knocking the goose and the tin can over was not coded as a target act. Another point was given if the goose clipped onto the tin can was put on the table. In the Mouse task, infants received one point if they pressed the top of the mouse down. The mouse had to be placed on the table when the infants produced this action. Taking the mouse into one hand and squeezing it together was not counted as a target act. For the Panda and ring task, infants could receive three points. One point was given if they lifted the panda up and bounced it with its head in front on the table. Dropping the panda on the table was not counted as a target at. Another point was given if the panda was placed into the ring with its bottom downwards. To be counted as a target act, at least half of the panda had to be inside the ring. A third point was given if the ring with the panda inside was slid over the table, even if the panda fell out during the task. Merely sliding the panda over the table without the ring was also coded as a target act. In the Drum task, infants could receive two points. If infants removed the stick from the side of the apparatus, they received one point. If they touched the button with the stick, they received another point. For the Octopus task, children could receive two points. One point was given if children placed the duck on the octopus even if the duck did not stick. They received another point if they turned the duck sitting on the octopus in one direction or forward and backward.

FIT 24

In the FIT 24, toddlers were given one to six points, depending on the task, if they produced the respective target acts. For the Gondola task, toddlers could receive three points. They received one point if they placed the blue manikin in the gondola. They received another point if they tried to lean the spoon on the arm of the manikin. This was only counted as a target act if the first step occurred and the manikin was placed in the gondola. Children received a third point if they slid the gondola back and forth on the table. Pushing the gondola only in one direction was not coded as a target act. The first two target acts did not necessarily have to be produced to receive the third point. In the Boat & Box task, children could receive five points. They received one point if they removed the tube from the sheath or at least tried to do so. If the tube accidentally fell out of the sheath, it was coded as a missing value. Children received another point if they tried to open the tube. They received a third point if they took the manikin out of the tube. If the manikin accidentally fell out of the tube, this was coded as a missing value. Bending the legs of the manikin into a sitting position was coded as another point. Children received a fifth point if they tried to place the sitting manikin into the boat. In the Panda & Slide task, toddlers received one point if they tried to place the long board against the stand, producing a slide. They received another point if they bounced the panda on the socket of the stand prior to producing the third target act. They received a third point if they slid the panda down the board. For the Ball with Eyes task, children could receive three points. They received one point if they found the slot in the ball by targeted touching of the ball or by turning it around and searching for it. They received another point if they tried to place the eyes into the slot. They eyes did not necessarily have to be stuck into the slot to receive the point. Children received a third point if they bounced the ball up and down on the table. Bouncing the ball only once was not coded as a target act. The first two target acts did not necessarily have to be produced to receive the third point. For the Turtle task, children could receive three points. They received one point if they clipped the half-cut ball and the cone together. Merely stacking them on top of each other was not coded as a target act. They could receive another point if they placed the objects, which had been clipped together, on the turtle. Children could receive a third point if they lifted the turtle up. The first two target acts did not necessarily have to be produced to receive the third point. In the Bunny task, children could receive three points. They received one point if they tried to attach the yellow pillow to the front of the bunny. Attaching the yellow pillow to the top or the back of the bunny was also coded as a target act. They received another point if they tried to attach the green pillow to the top of the bunny. Attaching the green pillow to the back or front of the bunny also counted as a target act. They received a third point if they tried to attach the rose pillow to the back or any other position on the bunny. For all three target acts, the pillows did not need to be stuck firmly on the bunny and were coded as target acts if placed approximately correctly. Placing distractors on the bunny was noted, including their position. In the Box task, children could receive three points. One point was given if they tried to hang the ring on the hook. Hanging it on a part of the box other than the hook was not coded as a target act. Children received another point if they tried to turn the ring around. A third point was given if they tried to open the drawer, either by pulling on the string or by using other strategies. The first two target acts did not necessarily have to be produced to receive the third point. In the Magnetic plate task, children could receive six points. One point was given if they turned the plate around. Another point was given if they intentionally placed the red button anywhere on the plate. A third point was given if they intentionally placed the yellow button anywhere on the plate. Another point was given if they intentionally placed the black button on the plate and a further point was given for intentionally placing the croissant anywhere on the plate. A final point was given for lifting and turning the plate around. This was coded as a target act even if the first five target acts had not been produced. Placing distractors (i.e., two plastic sausages) on the plate was noted.

Intraclass correlation coefficients (ICCs) of Lab-TAB test scores

Table S3

Results of ICC Calculations of the Task Orientation Test Scores at 12, 18, and 24 Months

Task orientation test variable	*N*	*r*	*p*
12 months			
Duration of looking	60	.850	<.001
Duration of playing	60	.930	<.001
Latency of first look away	60	.861	<.001
Mean intensity of facial interest	60	.661	<.001
Baseline state	60	.823	<.001
18 months			
Duration of looking	61	.946	<.001
Duration of playing	61	.995	<.001

Latency of first look away	61	.988	<.001
Mean intensity of facial interest	61	.897	<.001
Baseline state	61	.766	<.001
24 months			
Duration of looking	53	.982	<.001
Duration of playing	53	.996	<.001
Latency of first look away	53	.960	<.001
Mean intensity of facial interest	53	.919	<.001
Baseline state	53	.876	<.001

Note. ICC using average measure, absolute agreement, two-way mixed.

Table S4

Results of ICC Calculations of the Puppet Game Test Scores at 12, 18, and 24 Months

Puppet game test variable	*N*	*r*	*p*
12 months			
Mean intensity of smiling	60	.939	<.001
Mean presence of laughter	60	.832	<.001
Mean positive vocalizations	60	.946	<.001
Mean positive motor activity	60	.839	<.001
Mean toy engagement behavior	60	.613	<.001
18 months			
Mean intensity of smiling	59	.946	<.001
Mean presence of laughter	59	.874	<.001
Mean positive vocalizations	59	.942	<.001
Mean positive motor activity	59	.832	<.001
Mean toy engagement behavior	59	.661	<.001
24 months			
Mean intensity of smiling	60	.976	<.001
Mean presence of laughter	60	.866	<.001
Mean positive vocalizations	60	.813	<.001
Mean positive motor activity	60	.647	<.001
Mean toy engagement behavior	60	.782	<.001

Note. ICC using average measure, absolute agreement, two-way mixed.

Supplementary Results

Correlations at 12 Months

Table S5

Correlations of FIT 12 Test Score and Attentional and Social Aspects of Temperament and Intercorrelations of Variables from the Same Test Procedure

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. FIT 12 test score	-															
2. Duration of looking (TOT)[a]	-.14	-														
3. Duration of playing (TOT)[a]	-.17*	.34***	-													
4. Latency of first look away (TOT)[a]	-.01	.26**	.15	-												
5. Mean intensity of facial interest (TOT)[a]	-.09	.46**	.23**	.02	-											
6. Baseline state (TOT)[a]	-.03	-.28**	-.16	.02	-.34**	-										
7. Duration of Orienting (IBQ-R)[b]	-.16						-									
8. Mean intensity of smiling (PGT)[b]	.10							-								
9. Mean presence of laughter (PGT)[b]	.04							.20*	-							
10. Mean positive vocalizations (PGT)[b]	.05							.50***	.22*	-						
11. Mean positive motor activity (PGT)[b]	.12							.61***	.20*	.51***	-					
12. Mean toy engagement behavior (PGT)[b]	.03							.47***	.08	.22*	.45***	-				
13. High-intensity pleasure (IBQ-R)[b]	.17*												-			
14. Smiling and laughter (IBQ-R)[b]	.15												.40***	-		
15. Approach (IBQ-R)[b]	.16												.49***	.39***	-	
16. “Sociability” (IBQ-R)[b]	.17*												.82***	.79***	.57***	-

Note. [a] Variables referring to attentional aspects of temperament. [b] Variables referring to social aspects of temperament.
*** $p < .001$. ** $p < .01$. * $p < .05$.

Correlations at 18 Months

Table S6

Correlations of FIT 18 Test Score and Attentional and Social Aspects of Temperament and Intercorrelations of Variables from the Same Test Procedure

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. FIT 18 test score	-																		
2. Duration of looking (TOT)[a]	-.05	-																	
3. Duration of playing (TOT)[a]	.07	.77***	-																
4. Latency of first look away (TOT)[a]	.02	.07	.01	-															
5. Mean intensity of facial interest (TOT)[a]	-.00	.51***	.41***	-.02	-														
6. Baseline state (TOT)[a]	-.21*	-.00	-.05	.23*	-.18	-													
7. Attentional focusing (ECBQ)[a]	-.05						-												
8. Attentional shifting (ECBQ)[a]	-.00						.50***	-											
9. Effortful control (ECBQ)[a]	.02						.69***	.68***	-										
10. Mean intensity of smiling (PGT)[b]	.14									-									
11. Mean presence of laughter (PGT)[b]	-.10									.30**	-								
12. Mean positive vocalizations (PGT)[b]	-.06									.52***	.52***	-							
13. Mean positive motor activity (PGT)[b]	.18									.63***	.20*	.53***	-						
14. Mean toy engagement behavior (PGT)[b]	.09									.40***	.09	.23*	.29**	-					
15. Sociability (ECBQ)[b]	.13														-				
16. High-intensity pleasure (ECBQ)[b]	-.15														.33**	-			
17. Positive anticipation (ECBQ)[b]	.10														.26**	.22*	-		
18. Impulsivity (ECBQ)[b]	-.04														.07	.27**	.21*	-	
19. Extraversion (ECBQ)[b]	.02														.65***	.69***	.66***	.50***	-

Note. [a] Variables referring to attentional aspects of temperament. [b] Variables referring to social aspects of temperament.
*** $p < .001$. ** $p < .01$. * $p < .05$.

Correlations at 24 Months

Table S7

Correlations of FIT 24 Test Score and Attentional and Social Aspects of Temperament and Intercorrelations of Variables from the Same Test Procedure

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. FIT 24 test score	-																		
2. Duration of looking (TOT)[a]	.14	-																	
3. Duration of playing (TOT)[a]	.12	.82***	-																
4. Latency of first look away (TOT)[a]	.15	.19	.14	-															
5. Mean intensity of facial interest (TOT)[a]	.12	.81***	.73***	.12	-														
6. Baseline state (TOT)[a]	-.04	-.04	-.07	-.01	-.07	-													
7. Attentional focusing (ECBQ)[a]	-.08						-												
8. Attentional shifting (ECBQ)[a]	-.10						.24*	-											
9. Effortful control (ECBQ)[a]	-.14						.58***	.68***	-										
10. Mean intensity of smiling (PGT)[b]										-									
11. Mean presence of laughter (PGT)[b]	-.02									.36***	-								
12. Mean positive vocalizations (PGT)[b]	.07									.44***	.66***	-							
13. Mean positive motor activity (PGT)[b]	.17									.71***	.43***	.51***	-						
14. Mean toy engagement behavior (PGT)[b]	.22*									.53***	.13	.19	.35***	-					
15. Sociability (ECBQ)[b]	.16														-				
16. High-intensity pleasure (ECBQ)[b]	.11														.13	-			
17. Positive anticipation (ECBQ)[b]	.06														.14	.30**	-		
18. Impulsivity (ECBQ)[b]	.14														.20*	.40***	.38***	-	
19. Extraversion (ECBQ)[b]	.19														.56***	.68***	.66***	.69***	-

Note. [a] Variables referring to attentional aspects of temperament. [b] Variables referring to social aspects of temperament.
*** $p < .001$. ** $p < .01$. * $p < .05$.

5 Article 3 – Sieber, F., Czarnomski, J., Daum, M.M., & Zmyj, N. (2024). The relation between goal-predictive gaze behavior and imitation – A live eye-tracking study in 12-months-olds. *PsyArXiv Preprints.*

The relation between goal-predictive gaze behavior and imitation – A live eye-tracking study in 12-month-olds

Franziska Sieber[1,] Jan Czarnomski[1], Moritz M. Daum[2] & Norbert Zmyj[1]
[1]TU Dortmund University, Dortmund, Germany
[2]University of Zurich, Zurich, Switzerland

Published on *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/4mfys

Author Note

Franziska Sieber https://orcid.org/0009-0003-0019-6613

Moritz M. Daum https://orcid.org/0000-0002-4032-4574

Norbert Zmyj https://orcid.org/0000-0003-2603-0729

The German Research Foundation (DFG, ZM 54/6-1) funded this research. We thank our student assistants Hannah Linder, Maximilian Schauerte, and Gero Noßke for coding the videotapes. Finally, we thank all children and parents for participating in this study.

The data necessary to reproduce the analyses presented here are publicly accessible at https://osf.io/7rhzk/?view_only=8ac6570a6cf146819410c5803f4faa24. The authors have no conflict of interest to declare.

Correspondence concerning this article should be addressed to Franziska Sieber, Institute of Psychology, TU Dortmund University, Emil-Figge-Straße 50, 44227 Dortmund, Germany. *E-mail address*: franziska.sieber@tu-dortmund.de

Abstract

Children learn from others by imitating observed behavior. According to some theorists, in order to imitate an agent's action, infants need to identify the agent's action goal. To test this assumption, infants' predictive gaze shifts were measured while they observed a live model's action, and related to their imitation of this action, controlling for cognitive developmental status. In total, 104 German twelve-month-olds (57 female) observed an established imitation task (FIT 12, Goertz et al., 2006). The results revealed that infants anticipated action goals presented at a realistic speed. Furthermore, infants' imitative behavior was related to their goal-predictive gaze shifts. This association was partially explained by cognitive-developmental status, which should be considered as an important factor in the development of imitation.

Keywords: imitation, action prediction, goal prediction, live eye tracking, cognitive development, infants

The relation between goal-predictive gaze behavior and imitation – A live eye-tracking study in 12-month-olds

Imitation is central to early human development, as it allows for the transmission of cultural practices and knowledge (Tomasello et al., 2005). The development of imitation and factors influencing imitative behavior, such as situational, model, or task characteristics, have been widely studied over the past decades (e.g., Barr et al., 1996; Wood et al., 2013; Zmyj & Seehagen, 2013). However, the mechanisms underlying imitation and its development still require further investigation. Defining imitation as encompassing an understanding of the model's intention implies that in order to imitate, infants need to identify and understand the action goals of others (e.g., Carpenter & Call, 2009). One possibility to assess goal understanding in infants is to measure their goal-predictive gaze behavior, which should be related to imitation if goal understanding is indeed one of the mechanisms underlying imitative behavior. The present study therefore aimed to investigate the relation between imitative behavior and goal prediction by testing 12-month-olds in a live eye-tracking setting. This setting further enabled the investigation of goal prediction in a more naturalistic context, with a live model presenting actions at a realistic speed.

Imitation

Studies investigating imitation are initially faced with the challenge of defining imitation itself. Depending on the main research question, most definitions can be assigned to one of two main categories: Broader, behavior-based definitions conceive imitative behavior as copying any observed behavior (e.g., Barr et al., 1996; Paulus, 2011), whereas narrow, intention-based definitions specify imitative behavior as copying the means of an action in order to achieve a specific goal (e.g., Carpenter & Call, 2009; Tomasello et al., 2005). Depending on the definition employed, different potentially relevant mechanisms underlying imitation may be suggested. According to behavior-based definitions, attention to the action is the first part of the imitative process (Bandura, 1971). However, studies with infants in the first two years of life have shown that the duration of looking at an observed action is not correlated with the imitative behavior of this action (Kolling et al., 2014; Óturai et al., 2013; Sonne et al., 2016).

Intention-based definitions emphasize the role of means and goals in imitation (e.g., Gergley et al., 2002; Tomasello et al., 2005). Imitative behaviors include those behaviors in which an infant reproduces the goal of the model and the specific means that have led to this goal. Actions are construed as movements performed by an agent to elicit a desired goal. Accordingly, actions are represented in terms of their goals (Prinz, 1997), and infants must therefore identify the goal of an action in order to be able to imitate it (Tomasello et al., 2005). Thus, according to these approaches, imitation expresses what is typically called "infants' action prediction".

Goal prediction

The question of how young children perceive the behavior of others is central to developmental psychology (Hunnius & Bekkering, 2014), and infants' action prediction plays a key role in this field of research. One prominent way to assess infants' action prediction is to capture their goal-predictive gaze behavior during an ongoing action using eye tracking (Falck-Ytter et al., 2006). Goal prediction refers to the ability to anticipate the goal of an observed action before the action is completed, and goal-predictive gaze shifts have been used to measure action understanding both in adults (e.g., Flanagan & Johansson, 2003) and in infants (e.g., Falck-Ytter et al., 2006; Melzer et al., 2012). Infants demonstrate goal-predictive gazes from early in life, as measured through anticipatory looking at screen-based events. For instance, even six-month-olds anticipate familiar actions (i.e., bringing a cup to their mouth or a phone to their ear; Hunnius & Bekkering, 2010).

Relation between imitation and goal prediction

The two conceptual accounts of intention-based imitation and goal prediction can be conceived as depicting infants' action understanding. However, to the best of our knowledge, only one previous study has investigated the relation between imitation and goal-predictive gaze behavior. Gampe and colleagues (2016) presented 12-, 18-, 24-, and 30-month-olds with two multi-step actions on a computer screen, one familiar action (i.e., hammering) and one unfamiliar action (i.e., pulling) for this specific context (i.e., inserting a building block in a box). The children's gaze behavior was analyzed during the demonstration. After the demonstration, the children received the objects and their imitative behavior was measured. The faster infants anticipated the goal in the hammering condition, the more likely they were to imitate the observed action, while this was not the case in the pulling condition. These results suggest that action prediction is related to imitation of familiar actions but not unfamiliar actions.

Two criticisms can be raised regarding the methodology used by Gampe et al. (2016). First, the demonstrated actions differed from a lifelike presentation in several ways. The actions were presented on a digital screen, which could have underestimated young children's performance due to the video deficit effect, according to which infants show a lower ability to transfer learning from videos compared to learning from equivalent real-life experiences (Anderson & Pempek, 2005). Analogously to other related studies, the demonstration video showed only the main elements of the action and no distractors (e.g., Falck-Ytter et al., 2006; Hunnius & Bekkering, 2010; Melzer et al., 2012) and was presented at a relatively slow speed (i.e., 1000-1960 ms per action step, see also Ambrosini et al., 2013). Action duration plays an important role in infants' goal prediction, with research showing that slower action demonstrations result in more goal-predictive gaze shifts in 12-month-olds (Daum et al., 2016). Nevertheless, slower action demonstrations do not reflect the everyday learning environment of young children. Monroy and colleagues (2021) were the first to investigate action prediction in a real-life setting during infant-parent interactions, and reported that nine-month-olds anticipated their parents' goal-directed movements. Second, Gampe et al. (2016) did not control for children's general cognitive development, for example using the Bayley Scales of Infant and Toddler Development III (BSID-III, Bayley, 2006). Accordingly, imitation and anticipatory looking might result from changes in cognitive development rather than the prediction of actions.

The present study

The aim of the present study was twofold. First, we aimed to investigate infants' action prediction in a more naturalistic context, with a live model presenting the action. Second, we aimed to investigate the relation between goal prediction and imitation of goal-directed actions while controlling for infants' cognitive developmental status. Therefore, creating a novel approach, we investigated 12-month-olds' anticipatory looking during action observation in a real-life setting and examined its association with their imitation of the observed action, and additionally employed the Cognitive Scale of the BSID-III (Bayley, 2006).

First, we hypothesized that infants would show goal-predictive gaze behavior in live settings using an action with realistic speed. Second, we hypothesized that infants who imitated a goal-directed action would show faster goal-predictive gaze shifts during the demonstration of this action than infants who did not imitate.

Method

Participants

The participating families were recruited as part of a longitudinal study, from a database of families who had previously signed up to participate in child development studies. The families lived in one of two medium-sized cities in northwestern Germany. The final sample comprised 104 twelve-month-olds (57 girls; *M* = 360 days, *SD* = 9; range 344-386). Fifty-two additional children had to be excluded because they did not complete the test phase of the FIT (*n* = 2), had less than two valid eye-tracking trials (*n* = 15), did not complete the test phase of the FIT and had less than two valid eye-tracking trials (*n* = 2), or due to technical problems during calibration of the eye-tracking system (*n* = 33). All of the children were White. The majority of the parents had either a university degree (mothers 61%, fathers 51%) or a university entrance-level diploma (mothers 18%, fathers 16%) as their highest educational attainment. Parents provided informed consent for their children to participate in the study. Children received a small gift and parents received 10 Euros after the session. The present study was approved by the local Ethics Committee of the Faculty of Psychology at Ruhr University Bochum. The research was conducted in accordance with APA ethical standards in the treatment of the study sample.

An a priori G*power 3.1 analysis was run to determine the appropriate sample size for the longitudinal study, which showed that a sample of 85 participants was sufficient to achieve 80% power and a medium effect size (Faul et al., 2007). To ensure that we get data of 85 participants at each test point, we invited 160 children and their families.

Apparatus and stimuli

Gaze behavior was recorded using an EyeLink® 1000 Plus eye tracker (SR Research, Ottawa, Canada). During calibration, the eye tracker was mounted on a 17-inch screen with a resolution of 1280 x 1024 pixels. The sampling rate was 500 Hz. During the eye-tracking session, infants sat on their caregiver's lap, facing the experimenter at the opposite side of the table, approximately 60 cm away from the eye tracker. The calibration screen was placed on the table on a wooden box (27 x 15 x 36 cm; see Figure 1). When the screen was removed, another wooden box (27 x 10 x 36 cm) was placed on top of the first (see Figure 2).

The material used in the imitation task was as close as possible to the material used in the original task from the FIT 12 (Goertz et al., 2006). Infants were presented with a pink stuffed toy pig (height 16 cm) wearing a brown hat attached with Velcro.

Experimental setup and procedure

To test infants' imitative behavior, we used the Frankfurt Imitation Test for 12-month-olds (FIT 12, Goertz et al., 2006). The FIT 12 entails five tasks presented in a fixed order: Tin can, Pig, Cup & knife, Mouse, and Drum. The infants were presented with all five tasks. For the present study, we only used the Pig task because it is the only task that enabled us to measure anticipatory looks during action observation. The other tasks either do not entail target actions that capture goal prediction (Tin can, Cup & knife, Mouse) or cannot be analyzed due to spatial overlap of the action steps (Drum; see https://osf.io/7rhzk/?view_only=8ac6570a6cf146819410c5803f4faa24 for one FIT demonstration phase containing all tasks).

For the purpose of the present study, we adapted the original version of the Pig task in two ways: First, we implemented a baseline phase to assess whether the infants spontaneously showed the target behavior (Zmyj et al., 2017). Second, we awarded an extra point, because we conducted a two-step action compared to the original single step, putting the hat back on the pig in order to assess goal anticipation (see also Óturai et al., 2012; Zmyj et al., 2017).

After the child and their caregiver arrived at the laboratory, the experimental session began with a short warm-up phase, in which the experimenter played with the infants until they felt comfortable, that is, he played with the child until the child smiled at him and gave him a toy back during play. Next, the study started with the baseline phase of the FIT. Infants were seated in a highchair facing the experimenter at the opposite side of the table. The experimenter gave the infant one object after another for 30 s each, without any prior demonstration of the target action. For the subsequent demonstration phase, the infants were moved to another table in the same room, where the eye tracker was located. Eye tracking started with a 5-point calibration procedure, during which a small yellow duck expanded and contracted on the screen. The calibration screen was placed on a wooden box, and the eye tracker was placed in front of this box (see Figure 1). After calibration, the experimenter removed the screen and positioned another wooden box onto the first wooden box. Subsequently, the experimenter started the demonstration phase, in which he demonstrated the action live on the wooden boxes on a marked line, so that the demonstration height corresponded to the center of the calibration (see Figure 2). The experimenter removed the hat from the pig's head and then put the hat back on. This sequence was presented four times. The total duration of one action demonstration, from taking off the hat to putting it back on, was approximately 3,500 ms. The duration of the target action "putting the hat back on", from the start of the movement to reaching the pig's head, was approximately 900ms (see https://osf.io/7rhzk/?view_only=8ac6570a6cf146819410c5803f4faa24 for an example of one demonstration phase of the FIT).

After the demonstration phase, infants returned to the first table. Due to time constraints, we reduced the delay between the demonstration phase and the test phase from 30 min in the original version of the FIT (Goertz et al., 2006) to 15 min. During the 15-min delay, two tasks assessing infants' temperament were conducted (as part of the longitudinal project) and a break of 5 min was included. The temperament data are not reported here because they address theoretically different questions. After the delay, the FIT test phase started. Infants received the objects again and could play with them for 30 s each. The session ended with an assessment of infants' cognitive developmental status using the Cognitive Scale of the BSID-III, following the guidelines in the manual (Bayley, 2006). The Cognitive

Scale provides standardized scores for cognitive-developmental status (M = 100, SD = 15). The whole session lasted for approximately 60 min and was video-recorded.

Figure 1

Experimental setup during calibration

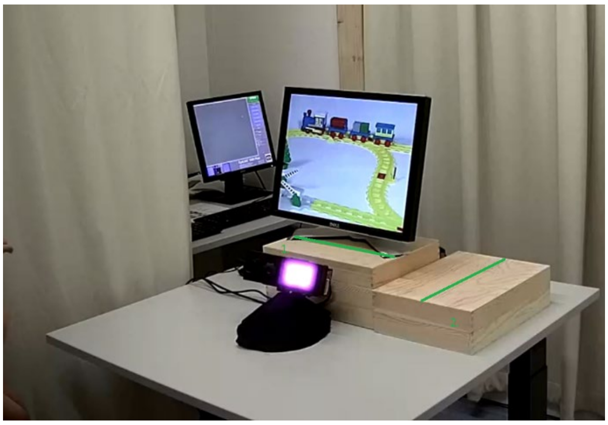
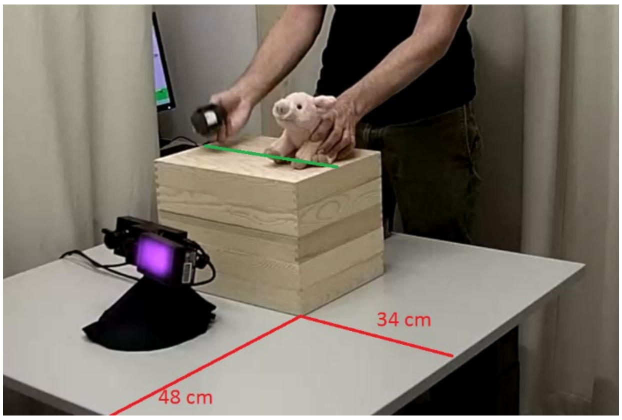


Figure 2

Experimental setup during demonstration



Data Processing

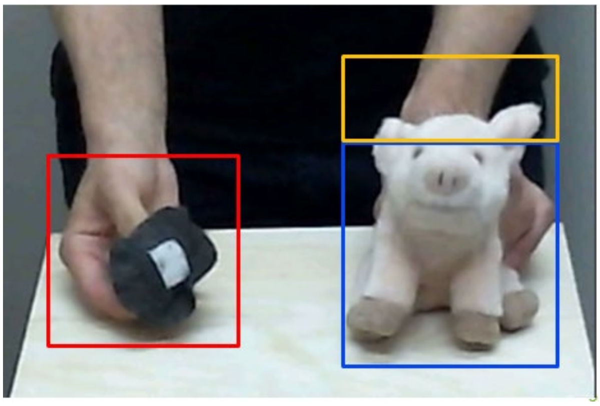
Eye-tracking data

At the time when the present study was conducted, there was no available ready-to-use set-up for live eye tracking. Thus, the videos were initially synchronized offline by hand in order to align the eye-tracking data with the video files using ffmpeg (https://ffmpeg.org). To analyze predictive gaze shifts, we defined three areas of interest (AoIs) covering the hat (dynamic Hat AoI), the pig's head as the intended target (static Goal AoI), and the pig's body (static Pig AoI) for each trial (see Figure 3). The dynamic Hat AoI was added frame by frame to match the gaze trace with the moving hat. Given that the action was demonstrated live and was not screen-based, the AoIs were defined manually for each child in each demonstration trial. Therefore, the size varied minimally between the trials (see Supplementary Material for definition criteria of the AoIs). We included the Pig AoI because infants show a preference for eyes in faces as the most salient cue (Wagner et al., 2013). To avoid falsely including gazes on the eyes or the nose of the pig as goal-predictive gaze shifts, we added this third AoI covering the face of the pig separated from the relevant Goal AoI. Data were only included in the analysis if participants fulfilled the following criteria for at least two trials: Infants had to gaze at the dynamic Hat AoI for at least 100 ms before shifting their gaze to the Goal AoI, and positive gaze-arrival times had to be under 2000 ms to be included as a valid trial, with gaze-arrival times over 2000ms excluded as outliers. We did not employ the conservative criterion that gaze shifts only count as predictive if the gaze is shifted after the hat starts to move (e.g., Falck-Ytter et al., 2006). Instead, we chose a more liberal criterion, including gaze shifts even if they appeared before the start of the movement. Unlike many other studies on goal-predictive gaze shifts, infants could see the experimenter's motionless hand after he had removed the hat and could therefore anticipate the action goal before the movement had started (for a similar approach, see Henrichs et al., 2012; Henrichs et al., 2014).

We also included the first trial in the analysis (see also Cannon et al., 2012), because infants had seen the action goal (the hat on the pig's head) twice before, in the baseline phase and at the beginning of the demonstration.

Figure 3

Exemplary definition of the AoIs for one trial



Note. Red = dynamic Hat AoI; yellow = static Goal AoI; blue = static Pig AoI.

Gaze-arrival times were calculated by subtracting the time when infants first fixated on the Goal AoI from the time when the hat entered the Goal AoI for each trial. Thus, positive gaze-arrival times represent goal-predictive gaze shifts whereas negative gaze-arrival times represent reactive gaze shifts. Mean gaze-arrival times were calculated for each infant using an average across all valid trials.

Two observers coded the videos independently using the software Data Viewer (SR Research, Ottawa, Canada). Another independent observer additionally coded 41 randomly chosen videos (37%). Interrater reliability for the mean gaze-arrival time was excellent, at $r = .893$, $F(40,40) = 11.010$, $p < .001$, ICC (2, 1, absolute agreement).

Imitation

Infants' imitative behavior was coded from the videos. Infants were rated as "imitators" if they put the hat back on the pig's head in the test phase or if the hat made contact with the pig's head, indicating that they attempted to do so. They were rated as "non-imitators" if they did not reproduce the target action. Another independent observer additionally coded the behavior of 60 randomly chosen infants. The interrater reliability for the imitative behavior was excellent, at $r = .924$, $F(59,59) = 13.632$, $p < .001$, ICC (2, 1, absolute agreement).

Cognitive Scale (BSID-III)

The coding of the Cognitive Scale of the BSID-III followed the guidelines in the technical manual (Bayley, 2006). An independent second observer additionally coded the behavior of 59 randomly chosen infants. Interrater reliability for the BSID-III scores was excellent, $r > .958$, $F(58,58) = 24.832$, $p < .001$, ICC (2, 1, absolute agreement).

Results

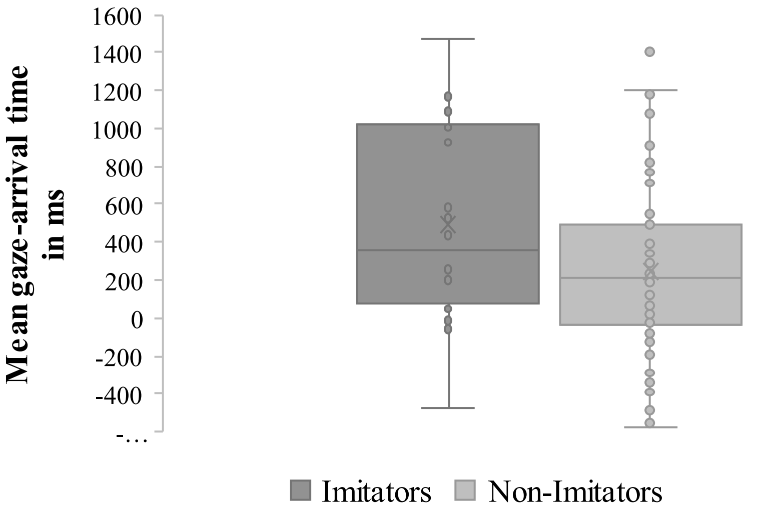
In the baseline phase of the FIT, no infant showed the target behavior. In the test phase of the FIT, 18 infants (17%) imitated the target action. The mean score on the Cognitive Scale (BSID-III) was 78.5 ($SD = 14.5$). Infants' mean gaze-arrival time was $M =$ 288 ms ($SD = 464$). The one-sample *t*-test against the threshold of 0 ms revealed a significant effect, $t(103) = 6.313$, $p < .001$, indicating that, on average, infants showed goal-predictive gaze shifts.

We first tested whether imitators were generally more attentive to the stimulus presentation than non-imitators by comparing the total looking time on all three AoIs averaged across all four trials of imitators and non-imitators. An independent samples *t*-test revealed no significant difference in the mean total looking time between imitators ($M = 2679$ ms, $SD = 531$) and non-imitators, $M = 2468$ ms, $SD = 1049$; $t(102) = -.827$, $p = .410$).

We then compared the mean gaze-arrival times of imitators ($M = 493$, $SD = 530$) and non-imitators ($M = 245$, $SD = 441$; see Figure 4). An independent samples *t*-test revealed that imitators showed significantly faster goal-predictive gazes than did non-imitators, $t(102) =$ -2.098, $p = .038$.

Figure 4

Box plots representing mean gaze-arrival times of imitators and non-imitators



Note. Circles represent the mean gaze-arrival time of every infant. Crosses represent the mean gaze-arrival times of the imitators and non-imitators.

Next, we conducted a hierarchical linear regression (inclusion method) with the FIT 12 test score as dependent variable. In the first step, we added infants' cognitive developmental status (BSID-III score) as a control variable. In the second step, we added infants' mean gaze-arrival time as a predictor. The final model, adjusted R^2 = .069, F(2, 100) = 4.764, p = .011, is depicted in Table 1. The cognitive-developmental status predicted the imitation score, whereas the mean gaze-arrival time did not uniquely contribute to the model.

Table 1

Hierarchical Regression Analysis Predicting the Imitation Score

Predictor	B	$SE\ B$	β	p	R^2	ΔR^2
					.068	.059**
Step 1						
Constant	-.368	.262				
Cognitive Scale (BSID-III)	.007	.003	.262**	.008		

					.087	.069
Step 2						
Constant	-.321	.204				
Cognitive Scale (BSID-III)	.006	.003	.223*	.027		
Mean gaze-arrival time	.00	.00	.142	.157		

Note. * $p < .05$. ** $p < .01$.

Discussion

The present study aimed to investigate the relation between imitative behavior and goal prediction in 12-month-old infants. In contrast to previous studies, which often presented the actions to be imitated on a computer monitor, we used a more naturalistic setting, in which a live model presented the actions at a realistic speed. Furthermore, we controlled for the infants' cognitive development.

Overall, the 12-month-old infants showed goal-predictive gaze shifts of the observed action. Moreover, we found a relation between goal prediction and imitation, insofar as infants who imitated the action showed faster goal-predictive gaze shifts than infants who did not imitate the action. However, goal-predictive gaze shifts did not predict imitative behavior when controlling for infants' cognitive development.

While previous research on goal prediction is largely based on screen-based eye tracking, our study is among the first to investigate goal-predictive gaze shifts in a more naturalistic context, with a live model presenting a specific goal-directed action at real-time speed. It adds to a growing body of literature addressing the need to investigate action prediction in real-world scenarios (Krol & Jellema, 2022; Monroy et al., 2021). In line with Monroy et al. (2021), who investigated action prediction in free-flowing real-time infant-parent interactions with six familiar toys in nine-month-olds, in the present study, 12-month-olds predicted the action goals in our setting. Our novel approach to study goal prediction in infants combines Monroy et al.'s (2021) approach with the previously established approach of controlled, screen-based laboratory paradigms (e.g., Falck-Ytter et al., 2006) by using live eye tracking in a controlled setting. This allowed us to investigate infants' perception of specific actions presented at a realistic speed that approaches real-life behavior.

In our setting, several potential distractors (e.g., the experimenter's face; presence of the caregiver) were present. Our results suggest that infants are able to allocate their attention to the relevant target action at real-time speed, even when these distractors are present. As such, the present findings indicate that real-time interactions in everyday life constitute infants' learning environment, which they use to build an understanding of social interactions and the behavior of others.

Our second main finding is the relation between goal prediction and imitative behavior. Infants who imitated the Pig task showed faster goal-predictive gaze shifts than infants who did not imitate this action. Previous studies found that infants' ability to perform a specific goal-directed action correlates with understanding the action goal when another person performs this action (e.g., Cannon et al., 2012; Melzer et al., 2012). The present study shows that this relation between action perception and action production (see Prinz, 1997) is also present in infants' imitative behavior. This finding is in line with theories on infant imitation that highlight the role of infants' sensitivity to the agent's action goal, which assume that infants need to identify an agent's internal goal (Tomasello et al., 2005), or at least the external goal of an ongoing action (Gergely et al., 2002), in order to imitate this goal-directed action.

Goal prediction is related not only to imitative behavior but also to other social-cognitive processes. For instance, a correlation study reported a link between goal prediction and theory of mind in two-year-olds (Krogh-Jespersen et al., 2015), and the aforementioned study by Monroy et al. (2021) found a relation between goal prediction and child-led joint attention in nine-month-olds. Moreover, another study found an association between goal-predictive gaze shifts and the creation of an internal feedforward model in six-month-olds

(Gredebäck et al., 2018). This further suggests that infants' goal prediction abilities are related to how they build an understanding of the behavior of others, reason about others' beliefs, and learn from social interaction partners early in life.

Although imitators showed faster goal-predictive gaze shifts than non-imitators, this difference was partially explained by infants' cognitive developmental status. This pattern of findings might be attributable to our study design, as we used only one imitation task of relatively high difficulty, which might have partially obscured the relation between goal prediction and imitation. Future studies might therefore use more than one task, with medium difficulty, to assess this relation. Furthermore, as our study is among the first to assess infants' cognitive development with respect to the relation between goal prediction and other social-cognitive processes, future research should consider the role of infants' cognitive development in action prediction.

The present study has several limitations. First, less than 20% of the infants imitated the target action, and we analyzed only one (out of five) target actions. Thus, future studies need to replicate these findings with a more extensive set of actions in order to assess action prediction and its association with action imitation more reliably. Another limitation pertaining to our methodology is that even if the infants looked at the experimenter's face, we were unable to include this behavior in our analysis. Faces, especially eyes, are potential sources of information for infants, and therefore particularly interesting (Wagner et al., 2013). We did not analyze infants' gaze behavior towards the experimenter's face in our paradigm because the video section focused on the presentation of the target action. A future step would therefore be to include the whole scene into the paradigm to allow for conclusions about the impact of additional sources of information on infants' goal prediction.

This study provides a novel approach to study infants' perception of others' behavior by measuring predictive gaze shifts in a combined live eye-tracking and imitation paradigm.

The results suggest that similar to paradigms using computer monitors for stimulus presentation, infants also predict others' action goals in a more naturalistic setting that resembles the speed of activities in daily life. Moreover, the results highlight the importance of predicting and imitating others' behavior, and support the notion that goal prediction is one of the key processes involved in imitation. Future studies should consider infants' cognitive developmental status to gain a better understanding of the role of infants' cognitive development in their goal prediction.

References

Ambrosini, E., Reddy, V., De Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PloS one*, *8*(7), e67916. https://doi.org/10.1371/journal.pone.0067916

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist*, *48*(5), 505-522. https://doi.org/10.1177/0002764204271506

Bandura, A. (1971). *Social learning theory*. General Learning Press. https://doi.org/10.1016/B978-0-12-813251-7.00057-2

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development*, *19*(2), 159-170. https://doi.org/10.1016/S0163-6383(96)90015-6

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). Psychological Cooperation. https://doi.org/10.1037/t14978-000

Cannon, E. N., Woodward, A. L., Gredebäck, G., von Hofsten, C., & Turek, C. (2012). Action production influences 12-month-old infants' attention to others' actions. *Developmental Science*, *15*(1), 35-42. https://doi.org/10.1111/j.1467-7687.2011.01095.x

Carpenter, M., & Call, J. (2009). Comparing the imitative skills of children and nonhuman apes. *Revue de primatologie*, *1*(6). https://doi.org/10.4000/primatologie.263

Daum, M. M., Gampe, A., Wronski, C., & Attig, M. (2016). Effects of movement distance, duration, velocity, and type on action prediction in 12-month-olds. *Infant Behavior and Development*, *43*, 75-84. https://doi.org/10.1016/j.infbeh.2016.03.002

Falck-Ytter, T., Gredebäck, G. & Von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769-771. https://doi.org/10.1038/nature01861

Gampe, A., Prinz, W. & Daum, M. M. (2016). Measuring action understanding: Relations between goal prediction and imitation. *British Journal of Development Psychology*, *34*(1), 53–65. https://doi.org/10.1111/bjdp.12125

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*(6873), 755-755. https://doi.org/10.1038/415755a

Goertz, C., Knopf, M., Kolling, T., Frahsek, S., & Kressley, R. (2006). Der Frankfurter Imitations-Test für 12 Monate alte Kinder (FIT 12): Erprobung einer Aufgabenserie zur Erfassung des deklarativen Gedächtnisses Einjähriger. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*(2), 88–96. https://doi.org/10.1026/0049-8637.38.2.88

Gredebäck, G., Lindskog, M., Juvrud, J. C., Green, D., & Marciszko, C. (2018). Action prediction allows hypothesis testing via internal forward models at 6 months of age. *Frontiers in Psychology*, *9*, 338825. https://doi.org/10.3389/fpsyg.2018.00290

Henrichs, I., Elsner, C., Elsner, B., & Gredebäck, G. (2012). Goal salience affects infants' goal-directed gaze shifts. *Frontiers in Psychology*, *3*, 391. https://doi.org/10.3389/fpsyg.2012.00391

Henrichs, I., Elsner, C., Elsner, B., Wilkinson, N., & Gredebäck, G. (2014). Goal certainty modulates infants' goal-directed gaze shifts. *Developmental Psychology*, *50*(1), 100 - 107. https://doi.org/10.1037/a0032664

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*(2), 446 –454. https://doi.org/10.1037/a0016543

Hunnius, S., & Bekkering, H. (2014). What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1644), 20130490. https://doi.org/10.1098/rstb.2013.0490

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*(1), 341. https://doi.org/10.1038/ncomms1342

Kolling, T., Óturai, G., & Knopf, M. (2014). Is selective attention the basis for selective imitation in infants? An eye-tracking study of deferred imitation with 12-month-olds. *Journal of Experimental Child Psychology*, *124*, 18-35. https://doi.org/10.1016/j.jecp.2014.01.016

Krogh-Jespersen, S., Liberman, Z., & Woodward, A. L. (2015). Think fast! The relationship between goal prediction speed and social competence in infants. *Developmental Science*, *18*(5), 815-823. https://doi.org/10.1111/desc.12249

Krol, M. A., & Jellema, T. (2022). Sensorimotor anticipation of others' actions in real-world and video settings: Modulation by level of engagement? *Social Neuroscience*, *17*(3), 293–304. https://doi.org/10.1080/17470919.2022.2083229

Melzer, A., Prinz, W., & Daum, M. M. (2012). Production and perception of contralateral reaching: A close link by 12 months of age. *Infant Behavior and Development*, *35*(3), 570-579. https://doi.org/10.1016/j.infbeh.2012.05.003

Monroy, C., Chen, C. H., Houston, D., & Yu, C. (2021). Action prediction during real-time parent-infant interactions. *Developmental Science*, *24*(3), e13042. https://doi.org/10.1111/desc.13042

Óturai, G., Kolling, T., Hall, L. R., & Knopf, M. (2012). The role of object functions for deferred imitation–Do infants selectively retain and forget target actions? *Infant Behavior and Development*, *35*(2), 195-204. https://doi.org/10.1016/j.infbeh.2012.01.004

Óturai, G., Kolling, T., & Knopf, M. (2013). Relations between 18-month-olds' gaze pattern and target action performance: A deferred imitation study with eye tracking. *Infant Behavior and Development*, *36*(4), 736-748. https://doi.org/10.1016/j.infbeh.2013.08.005

Paulus, M. (2011). Imitation in infancy: Conceptual considerations. *Theory & Psychology*, *21*(6), 849-856. https://doi.org/10.1177/0959354310395990

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*(2), 129-154. https://doi.org/10.1080/713752551

Sonne, T., Kingo, O. S., & Krøjgaard, P. (2016). Empty Looks or Paying Attention? Exploring Infants' Visual Behavior during Encoding of an Elicited Imitation Task. *Infancy*, *21*(6), 728-750. https://doi.org/10.1111/infa.12141

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675-691. https://doi.org/10.1017/S0140525X05000129

Wagner, J. B., Luyster, R. J., Yim, J. Y., Tager-Flusberg, H., & Nelson, C. A. (2013). The role of early visual attention in social development. *International Journal of Behavioral Development*, *37*(2), 118-124. https://doi.org/10.1177/0165025412468064

Wood, L. A., Kendal, R. L., & Flynn, E. G. (2013). Whom do children copy? Model-based biases in social learning. *Developmental Review*, *33*(4), 341-356. https://doi.org/10.1016/j.dr.2013.08.002

Zmyj, N., Schölmerich, A., & Daum, M. M. (2017). The relationship between attention and deferred imitation in 12-month-old infants. *Infant Behavior and Development*, *48*, 175-183. https://doi.org/10.1016/j.infbeh.2017.02.001

Zmyj, N., & Seehagen, S. (2013). The role of a model's age for young children's imitation: A research review. *Infant and Child Development*, *22*(6), 622-641. https://doi.org/10.1002/icd.1811

Supplementary Materials

Definition criteria of the AoIs

Hat AoI

The dynamic hat AoI must always cover the hat. In the first frame, its top edge is above the ball of the hand. The right edge is next to the hat. The left edge is next to the knuckle of the index finger and the bottom edge is under the hat. Consequently, the first dynamic AoI covers the hand and the hat. When the hand starts to move, the hat always has to be covered by the AoI and stays in its center. The last dynamic AoI is set next to the goal AoI and the pig AoI, when the hat enters the goal AoI, because the AoIs must not overlap.

Goal AoI

The static goal AoI covers the area on the head of the pig, where the hat is placed. The bottom edge is set directly above the eyes, so that it does not include the eyes. The left and right edge is set next to the respective ears. They are included in the AoI. The top edge covers a part of the demonstrators' forearm. It has to be big enough that it covers the hat when it is placed on the head of the pig.

Pig AoI

The static pig AoI is placed directly below the goal AoI. The top edge covers the pig's eyes. The bottom edge is placed under its feet. The left and right edges are placed so that the AoI covers the entire body of the pig, starting below the eyes.

6 General Discussion

The major aim of the current dissertation was to systematically examine the relation between imitative behavior and temperament during the second year of life, exploring whether the motivation to imitate shifts from an instrumental to a social purpose between 12 and 24 months of age. Given the important role of attention in imitative learning (Bandura, 1977) and inconclusive evidence from looking time studies, I also aimed to investigate infants' gaze direction, more specifically their goal-predictive gaze shifts, during action demonstration and its relation to imitative behavior using a novel methodological approach in a live eye-tracking setting. To investigate this, a longitudinal study employing standardized assessments of imitation, temperament, and cognition at 12, 18, and 24 months was conducted. Eye-tracking was further implemented during the action demonstration in the imitation tests.

In Article 1, we began by examining the factor structure and longitudinal stability of the two infant temperament questionnaires utilized in the study: IBQ-R (Gartstein & Rothbart, 2003) and ECBQ (Putnam et al., 2006). As outlined in section 1.2.2 (Measuring Temperament), results regarding the underlying factor structure of the IBQ-R remain inconsistent, and especially in Germany, the three-factor solution is questionable. To determine the suitability of scale and factor level data for the main research question on the developmental path of the two functions of imitation, we first validated the factor structure in two longitudinal samples from 9 to 12 months and from 12 to 24 months. In line with previous studies using the German version of the IBQ-R (Mink et al., 2013; Vonderlin et al., 2012), a two-factor solution was confirmed in both samples, as we failed to replicate the self-regulation factor Orienting/ Regulation. On the other hand, the three-factor structure of the ECBQ was confirmed. Stability of temperament occurred within the questionnaires (IBQ-R; ECBQ) and between the two questionnaires.

There are two possible explanations for the absence of the regulation factor in the German samples (see Mink et al., 2013; Vonderlin et al., 2012). First, it remains uncertain whether the regulation factor is an autonomous factor within the first year of life, that can be measured independently of the two affectivity factors. Self-regulatory capacities, which modulate reactivity, begin to emerge toward the end of the first year of life and during the transition to the second year (Rothbart, 1981). In particular, the maturation of neural networks of attention, which begins late in the first year of life, leads to an increase in the child's ability to regulate emotional and behavioral responses (Rothbart & Ahadi, 1994). Consistent with this, other studies confirming the three-factor solution, have found notable double loadings of the regulation factor scales on the other two factors, Surgency/ Extraversion and Negative Affectivity (e.g., Dragan et al., 2011; Gartstein et al., 2016). Second, cultural differences may account for the divergent findings regarding the underlying factor structure. As outlined in section 1.3 (Imitation and Temperament) using the example of the 5-HTTPLR allele in individualistic vs. collectivistic cultures, individual differences can be attributed to genes and environment. Culture encompasses the attitudes, values, goals, and practices, that guide parents in raising their children in alignment with the values of their cultural group. It also plays a crucial role in harmonizing children's individual characteristics, such as temperament, with cultural expectations (Gartstein et al., 2024). Thus, cultural expectations and norms are one factor likely to influence temperament. Evidence from studies comparing cross-cultural samples and single-sample studies suggests the importance of culture in the development of temperament (Putnam et al., 2024). As cultures differ in what constitutes desirable behavior, some temperament traits are fostered while others tend to be discouraged (Chen et al., 2012). Even among Western cultures that are often perceived as similar, notable cultural differences exist in values, parenting practices, and social expectations: while American mothers highlight positive emotionality with a friendly and easy attitude in parenting, German mothers focus on raising their children as being empathetic and helpful (Kirchhoff, 2015). These differences may result in different structures of temperament and may explain the divergent results between studies with samples from different countries (e.g., Desmarais et al., 2019; Gartstein et al., 2006; Sung et al., 2015).

A recent study by Schmidt and colleagues (2024) provided at least partial evidence for the three-factor structure in a German sample. In line with our findings and previous studies (Mink et al., 2013), factor analyses suggested a two-factor solution with Negative Affectivity and Surgency/ Extraversion at 6 months. In contrast, they were able to replicate the third factor, Orienting/ Regulation at 12 months. Nevertheless, the evidence is not yet conclusive, as the results are accompanied by certain limitations. In line with our finding, the model fit of the original model (Gartstein & Rothbart, 2003) in the confirmatory factor analysis (CFA) was poor. The exploratory factor analysis (EFA), conducted due to the poor fit of the expected structure, identified three factors. However, a closer examination suggests that the regulation factor lacks a clear and cohesive structure. Only three (Attentional Focusing, Low-Intensity Pleasure, Cuddliness) of the four scales usually associated with this factor, loaded positively on it. Two of these scales, Cuddliness and Low-Intensity Pleasure, also showed negative double loadings on Negative Affectivity. In addition, Vocal Reactivity and Smiling and Laughter, which are usually associated with Surgency/ Extraversion, instead loaded positively on the regulation factor. High-Intensity Pleasure and Perceptual Sensitivity, both associated with Surgency/ Extraversion, also showed double loadings on the regulation factor. Therefore, it remains uncertain whether the third factor that emerged in the EFA truly represents only regulatory components or if it also captures other aspects of positive affect.

In summary, we found no support for the three-factor solution underlying infant temperament as measured by the IBQ-R (Gartstein & Rothbart, 2003). As the results of Schmidt and colleagues (2024) can only be interpreted with certain limitations and do not provide a coherent picture of the third factor, the use of the German IBQ-R is recommended only at scale level.

Article 2 provided longitudinal data on the relation between imitation and social and cognitive aspects of temperament within the second year of life. Imitative behavior was related to social aspects of temperament throughout the entire second year of life. We found no consistent evidence for the relation between attentional aspects of temperament and imitation at 12, 18, and 24 months while controlling for infants' cognitive development. The reported relations are mostly based on questionnaire data, whereas behavioral data did not show the same patterns. Longitudinal analysis did not reveal a connection between early social temperament at 12 months and later imitative behaviors. These results indicate that there is no shift from an instrumental to a social motivation within the second year of life, but that imitation serves a social function from an early age.

The reported relations between imitation and social aspects of temperament are partly at odds with previous findings. While the association between imitation and social orientation later in the second year of life at 18 and 24 months is consistent with previous research (Dixon et al., 2012; Hilbrink et al., 2013), the relation at 12 months contradicts the findings of Zmyj and colleagues (2017), who reported no association between social aspects measured by the IBQ-R and imitative behavior. On closer inspection, this inconsistent result is not all that divergent and may be explained by a methodological difference. Zmyj et al. (2017) correlated the IBQ-R subscales related to the extraversion factor with imitative behavior and found no relation. Our results replicate previous findings, showing no relation between these subscales and imitation. Given that these subscales assess some aspects of positive affect unrelated to infants' social motivation, we decided to rationally derive a sociability subscale from the item pool of the IBQ-R that assesses pleasure in interacting with others. This sociability scale was found to be related to imitation, suggesting that infants' social motivation to engage with others is already linked to their imitative behavior at 12 months.

Contrary our expectations, we found no relation between imitation and attentional aspects of temperament measured through questionnaire and observation at any. In contrast, Zmyj et al. (2017) found a relation between attention as measured by IBQ-R and observation in the laboratory. These contradictory findings may be explained by the fact that we assessed infants' cognitive development, whereas Zmyj and colleagues (2017) did not control for this. The implications of this difference will be discussed below. Nevertheless, this contradictory finding is consistent with the inconsistent picture of the relation between attention and imitation from previous research (Heimann, 2022b). While Zmyj and colleagues (2017) found a relation between attention as part of infant temperament in 12-month-olds, Dixon and colleagues (2012) reported no relation between the temperament factor Effortful Control and imitation in 15-month-olds, in line with our findings. These mixed findings illustrate the difficulty of capturing attention. Different methods of capturing attention, i.e., ratings of attention and performance-based measures, may tap into different parts and levels of attentional processes (Acar et al., 2019). Parent-report ratings of attention showed no correlation with teacher or observer ratings, nor with performance-based measures. While teacher and observer ratings, along with laboratory attention tasks, clustered into a latent factor of focused attention, this factor was not identified in parent ratings (Acar et al., 2019), suggesting that these measures do not assess the same construct of attention. Following these considerations and given the importance of attention for observational learning (Bandura, 1977), our results do not suggest that imitation and attention are not related throughout the second year of life, but rather that we measured parts of attention that do not predict imitative behavior. Given that attention measured through looking time was also not related to imitative behavior in 12-month-olds (Kolling et al., 2014), 18-month-olds (Óturai et al., 2013), and 20-month-olds (Sonne et al., 2016), another method that captures infants' attention at a different level may be more informative about the relation between imitation and attention. Studies examining infants' gaze direction during action demonstrations may provide a more accurate understanding of the relationship between attentional processes and imitation, as they enable precise tracking of where the child is looking at specific moments during the demonstration (see Article 3).

However, the lack of relation between attentional aspects of temperament measured by questionnaire and imitation does not imply, that imitation does not serve a cognitive function. Indeed, not only do aspects of attentional temperament appear to be unsuitable for capturing the attentional processes relevant to imitative behavior, but they also appear to be unsuitable for capturing instrumental motivation. A major advantage of our study design was that we also assessed infants' cognitive development as a control variable. After controlling for infants' cognitive development, there was no relation between attention and imitation at 12 and 24 months. Instead, cognitive development at these ages predicted imitative behavior. These results suggest that attentional processes assessed through questionnaires may not capture all aspects of instrumental motivation, and therefore, may not be effective in detecting cognitive functions without taking infants' cognitive abilities into account. Fenstermacher and Saudino (2016) showed that imitative behavior, attentional aspects of temperament and cognitive developmental level assessed via Bayley are genetically linked. More specifically, the genetic factors linking imitation and task orientation are the same as those linking imitation and cognitive development. This may further explain the difficulties in relating attentional aspects of temperament to imitative behavior without considering the cognitive developmental level. In line with that, Ingersoll and Meyers (2011) found an association between attention following and two imitation tasks in children with autism, that was no longer observable after controlling for developmental level, replicating our findings.

Accordingly, the association may be better accounted for by developmental level and the role of attention in imitation needs to be investigated differently, as outlined above. However, the use of both cognition and temperament tests allowed us to distinguish, at least in part, between individual traits and cognitive abilities. Furthermore, the results highlight the importance of infants' cognitive development for imitative behavior. This will be discussed in section 6.1 (Relevance).

To the best of my knowledge, our study is the first to systematically explore the developmental trajectory of the two functions of imitation during the second year of life using a longitudinal design and standardized assessments. A recent experimental study by Gellén and Buttelmann (2019) supports the idea that the two functions of imitation change within the second year of life. They applied two rational imitation paradigms cross-sectionally in 14-, 18-, 24-, and 36-month-olds to investigate the developmental path of rational imitative behavior. In rational imitation paradigms, participants in the current imitation situation do not face the same constraints as the model when acting on an apparatus (e.g., Gergley et al., 2002). Depending on the constraints of the model, infants engage in selective imitation (e.g., Gergely et al., 2002; Paulus et al., 2011, Beisert et al., 2012). Gellén and Buttelmann (2019) found a significant decrease in selective imitation from 14 months onwards, while exact imitation increased. They also observed an increase in gaze direction toward the model during imitation after 14 months, which was interpreted as evidence of a shift in the underlying motivations for imitation. However, there are several aspects that limit the authors' interpretation as they only applied the rational imitation tasks and no further test of the underlying motivation. Gaze behavior has to date never been applied to test why infants imitate and to draw conclusions about the social motivation (Gellén & Buttelmann, 2019). In the absence of other behavioral information, such as whether infants were likely to initiate contact, another explanation than increased social orientation for increased looking times is possible. Increased looking times may indicate that the infants wanted to make sure that they were solving the task correctly. Thus, the reason for increased looking times in older infants remains unclear. Since looking time is the sole measure of social orientation, interpreting these results as a shift in infants' motivation is speculative. Additionally, a key limitation of their interpretation is the cross-sectional study design.

Importantly, we used an instrumental task to assess imitation and considered infants' individual characteristics rather than experimentally changing the context of the situation. The previously discussed studies by Gellén and Buttelmann (2019) and by Nielsen (2006; see section 1.3, Imitation and Temperament), that were seen as evidence for the change in the two functions, experimentally manipulated the model's constraints (Gellén & Buttelmann, 2019) or the social cues provided by the model (Nielsen, 2006). This is consistent with many studies that have attempted to investigate infants' motivation to imitate by experimentally manipulating the task (e.g., adding irrelevant actions, Hilbrink et al., 2013), the model (e.g., a model acting socially vs. aloof, Nielsen, 2006), or the context (e.g., live vs. video demonstration, Nielsen et al., 2008; social context by varying the situation prior to the imitation test, Yu & Kushnir, 2014). Differences in imitative behavior might therefore not only be attributed to the underlying motivation but be caused by external factors. To capture infants' general motivation to imitate without external influences, we considered infants' individual characteristics in relation to their imitative behavior. We minimized the influence of situational factors by selecting an instrumental task without varying the social setting and excluded the performance aspect by controlling for infants' cognitive development. This approach enabled us to focus on examining children's intrinsic motivation to imitate, as reflected in their temperament. In line with Nielsen (2006), our results provide an unambiguous picture that imitation and social orientation are at least related from 18 months onwards. Therefore, I do not reject the idea that infants may become more sensitive to social cues as they get older. Manipulating these social cues and the characteristics of the situation could result in younger infants showing lower levels of social motivation because they are not as sensitive to the different cues as older infants. However, excluding external influences and social cues and instead considering individual differences, our results suggest that even at 12-months old infants’ imitative behavior is predicted by their social orientation. The idea of an early emerging social function of imitation is further supported by Thiele and colleagues (2021) who found an increased preference for social interactions already in 13-month-olds.

In summary, infants' individual characteristics, particularly their social orientation, are closely linked to their imitative behavior. Every act of imitation occurs within a social context, highlighting that imitation is fundamentally a social process (Over, 2020). Therefore, infants’ individual social orientation must be taken into account in order to determine their motivation to imitate in a given situation. Regardless of the context, social motivation plays a role as early as 12 months of age, arguing against a change in the two functions of imitation in the second year of life.

Article 3 provided a new methodological approach to study infants’ gaze behavior, in particular their goal-predictive gaze shifts, during action demonstration using a live eye-tracking setting, and its relation to imitative behavior. The results revealed that infants are able to anticipate actions of others presented at naturalistic speed. These goal-predictive gaze shifts were related to infants’ imitative behavior. However, when controlling for infants’ cognitive development, the relation disappeared.

The results confirm previous findings on infants' ability to predict others' goals (e.g., Cannon et al., 2012; Gampe et al., 2016; Falck-Ytter et al., 2006), while extending these findings by demonstrating their ability to predict goals in a more naturalistic setting where actions are presented at realistic speeds. In line with this, Monroy and colleagues (2021) found that nine-month-olds were able to predict actions during free-flowing real-time infant-parent interactions. By combining their approach with the established screen-based paradigms of action prediction (e.g., Cannon et al., 2012), we provided a novel approach that overcomes the limitations of these established approaches, while providing a controlled setting to investigate specific research questions. This allows us to examine infants' action perception of specific actions at realistic speeds that resemble real-life interactions and represent infants' learning environment, and thus investigate how infants gain an understanding of others' behaviors. This adds to a growing body of research highlighting the need to investigate action prediction in real-life scenarios (Gredebäck & Falck-Ytter, 2015; Krol & Jellema, 2022; Monroy et al., 2021). In addition to offering a novel approach for more realistically studying goal prediction, this method is also significant for examining imitation behavior, as it provides a means to investigate attention during action demonstrations. This is particularly relevant, as many imitation studies attribute infants' imitation performance to how they allocate their visual attention to the demonstration, emphasizing the crucial role of attention in imitation (e.g., Beisert et al., 2012; Heimann, 2022b; Paulus, 2011), although direct evidence from live imitation paradigms is mostly lacking due to the difficulties of using eye-tracking in live settings.

Furthermore, the results are consistent with a finding on the relation between goal prediction and imitation for a familiar action (i.e., the hammering action, Gampe et al., 2016) and extend it to unfamiliar actions. As the target action 'putting the hat back on the pig's head' was not shown spontaneously in the baseline phase, the action can be considered novel. This finding adds to a number of studies demonstrating a relation between action perception and action production (e.g., Ambrosini et al., 2013; Cannon et al., 2012; Daum et al., 2011; Melzer et al., 2012). This is consistent with the common coding approach (Prinz, 1997), which assumes that action perception and action production are closely linked, as they share a common representational domain.

This finding adds to the ongoing debate regarding the mechanisms underlying imitative behavior. Infants' ability to predict action goals before witnessing the completion of an action implies that they are not merely copying movements but engaging in a more advanced form of social learning, where they infer the underlying intention behind the observed action. This supports the assumption that goal understanding is central to imitation, a view that has been proposed by several theorists (e.g., Tomasello et al., 2005). Accordingly, in order to imitate an action, infants must identify the action goal (Tomasello et al., 2005).

However, the relation observed between imitation and goal prediction should be interpreted with caution. Imitators did indeed show faster goal-predictive gaze shifts than non-imitators. However, this difference was explained by the infants' cognitive development. Although methodological issues, which will be discussed in section 6.2 (Limitations and Future Research), may have partially obscured the relation between goal prediction and imitation, the results on infants' gaze direction also suggest that infants' cognitive development needs to be taken into account in order to understand their imitative behavior, as outlined in the discussion of Article 2. Therefore, I will now address the relevance and implications of the findings from the longitudinal study in the following section.

6.1 Relevance

Infants’ ability to imitate plays a crucial role in their social-cognitive development. Imitation is for example related to language acquisition (Charman et al., 2000), understanding of others’ minds (Meltzoff, 1995), and forming and maintaining relationships (Over & Carpenter, 2013). Understanding how imitative abilities develop within infancy is therefore an important aspect of understanding how these related cognitive and social abilities develop. However, drawing a coherent picture of the development of imitative abilities is difficult (Damm et al., 2011; Jones, 2009). Imitation studies differ in terms of their theoretical orientation, their focus on cognitive (declarative memory capacity), socio-cognitive (understanding the intentions of others) or social mechanisms (contact and communication with the model), as well as differences between imitation tasks in terms of their content, their complexity and the context in which they are presented (see Damm et al., 2011). As a result, imitative behavior exhibits considerable variation due to a range of influencing factors, making it challenging to form a consistent understanding. The current dissertation has identified two constructs that contribute to infants' imitative behavior independently of the differences and across the different aspects outlined above without experimentally manipulating the task, model or context.

First, understanding imitative behavior requires taking into account infants' individual characteristics. To fully grasp whether imitative behavior emerges, it is crucial to understand how children utilize imitation to achieve their social or instrumental goals. Consistent with our results, Yu and Kushnir (2020) found a relation between infants' social motivation, as assessed by their temperamental social orientation, and imitative behavior across tasks. Individual-level differences suggest that even when faced with the same task in the same context, children of the same age differ in their tendency to imitate (Yu & Kushnir, 2020). Accordingly, children may differ in what they focus on during imitative tasks. While one child may perceive the same task as an opportunity for social interaction, another may focus on the action itself and view the situation as an opportunity to learn. This variability in imitative behavior across different imitation tasks highlights the significant influence of infants' individual characteristics on their tendency to imitate (Over, 2020). Therefore, I argue that in order to fully understand children's imitative behavior in a given situation, it is necessary to consider their multiple possible motivations, which vary as a function of their individual characteristics. Accordingly, disregarding individual characteristics and instead attributing differences in imitative behavior to task, context or model characteristics would be insufficient to understand whether a child imitates a behavior or not. Accordingly, the current dissertation supports various proposals to consider individual characteristics into account to understand infants' imitative behavior (Heimann, 2022b; Fenstermacher & Saudino, 2006; 2016; McCall et al., 1977; Yu & Kushnir, 2020).

A second implication that can be drawn from the results of the longitudinal project concerns the relevance of infants' cognitive developmental level for their imitative behavior. In both Article 2 and 3, infants' cognitive development predicted imitative behavior. In Article 3, the observed relation between imitation and goal prediction disappeared after controlling for developmental level. This is consistent with other studies that have found a relation between imitation and cognitive skills in children with autism (e.g., motor ability, Rogers et al., 2003; attention following, Ingersoll & Meyers, 2011), which was no longer observed after controlling for infants' cognitive development. Further, Rogers and colleagues (2003) found that imitative abilities were highly related to overall developmental status in typically developing children, children with autism, and children with other developmental disorders. Evidence for a direct relation between imitation performance and developmental level was provided by Fenstermacher and Saudino (2016). Their findings revealed overlapping genetic influences, with many of the same genetic factors that underpin children's mental development also affecting imitation performance. Accordingly, infants' cognitive development serves as an indicator of imitative behavior, supporting theoretical approaches to cognitive development such as Bandura's Social Cognitive Theory (1989), which proposes that cognitive capacities to attend to, process and remember observed actions are necessary for imitative behavior. At the same time, this underscores the importance of controlling for infants' cognitive developmental status when examining potential connections between imitation and other cognitive developmental skills, in order to accurately determine the specific contribution of each skill to imitative behavior.

In sum, the current dissertation has provided relevant input for understanding the development of imitation in the second year of life. Infants' individual characteristics and their cognitive development level must be considered as determinants of infants' imitative behavior.

6.2 Limitations and Future Research

The study design has many methodological strengths, including the longitudinal design to investigate a developmental process, the use of standardized imitation and temperament tests, the use of a multi-method approach to assess temperament, the ability to control for infants' cognitive development, and the ability to analyze infants' gaze behavior during live action demonstrations. In light of the discussion of the results, some limitations and open questions still remain, paving the way for future research.

The aim was to determine infants' motivation to imitate independently of their abilities and situational cues. While infants' social motivation to imitate is described by their social orientation, which is assessed as temperament (see also Hilbrink et al., 2013; Yu & Kushnir, 2020), it is questionable whether infants' instrumental motivation is fully captured by the attentional aspects of temperament. Since, to my knowledge, no scale specifically captures infants' instrumental motivation, assessing the attentional aspects of temperament has at least enabled the differentiation between individual characteristics and abilities. However, the results suggest that attentional aspects of temperament do not predict imitative behavior without considering their cognitive abilities. Thus, they may not be suitable for detecting the instrumental motivation of imitation. Therefore, future studies should develop a measure that captures infants' instrumental motivation. However, developing a measure of instrumental motivation may be challenging. Comparable measures exist for adults and are often used to examine academic achievement, such as second-language acquisition (e.g., Yu & Downing, 2012; Mehrpour & Vojdani, 2012) or mathematics achievement (e.g., Liu et al., 2020). These measures rely on self-report, including various items assessing instrumental motivation. It will not be possible to transfer these self-report scales to young children.

As outline above, the novel approach of analyzing infants’ gaze behavior during action demonstration provides important implications for understanding infants’ perception of others and their actions. However, there are a certain number of methodological limitations that limit the interpretation of the results at this time. First, we only included one target action in the analysis as the other tasks of the FIT 12 (Goertz et al., 2006) did not allow for the investigation of goal-predictive gaze shifts. This task has been shown to be more difficult, as less than 20% of the infants imitated the target action. This could also be due to the fact that imitating the target action (putting the hat back on) would have required taking a first action step (removing the hat from the head). Accordingly, imitating the target action involves relatively high cognitive demands. This may explain why the relation between goal prediction and imitation disappeared when infants’ cognitive development was controlled for. However, the question whether infants’ visual attention is related to their imitative behavior remains open. Future studies need to replicate our findings with a range of other actions of varying difficulty to better understand the relation between goal prediction and imitation and the relevance of cognitive development to this.

Furthermore, the data are limited to the presentation of the target action and do not encompass the full scene, particularly the experimenter's face. As a result, we were unable to analyze infants' gaze behavior towards the experimenter. The influence of the experimenter’s gaze on infants’ visual attention and their imitation behavior remains unclear. Including the experimenter in the paradigm would be particularly relevant given that infants show a preference for faces (e.g., Gluckman & Johnson, 2013) and that attention is increasingly directed to faces during infancy (e.g., Frank et al., 2014). People's eyes reveal information about where their attention is focused (Frischen et al., 2007) and are therefore potential sources of information for infants (Wagner et al., 2013). Koch and colleagues (2018) investigated infants' visual attention during successive demonstrations of target actions. Twelve- and 16-month-olds' visual attention changed across demonstrations, as they showed increased attention to the experimenter's face relative to the action with each presentation. Accordingly, the experimenter's gaze, to which infants pay increased attention to over the course of multiple action demonstrations, may influence infants' imitative behavior and their goal-predictive gaze shifts. Thus, future studies should experimentally investigate the influence of the experimenter's gaze direction on infants' imitative behavior and their goal-predictive gaze shifts. Furthermore, a future step would be to include the entire scene in the paradigm, which would allow conclusions to be drawn about how additional sources of information (i.e., the experimenter's gaze) affect infants' goal prediction. Another critical point is that the application is susceptible to errors due to the complex construction of the setting, which is reflected in relatively high drop-out rates of the eye-tracking data. In conclusion, this can be seen as a first approach to establishing an applicable framework for a live eye-tracking setting, which needs to be extended in future studies.

As outlined above, the relation between imitation and goal prediction, both of which reflect infants' action understanding, is still unclear from the results of this study. Thus, the question of whether infants' action understanding is relevant to imitation, and thus reveals an underlying mechanism of imitative behavior, as suggested by various accounts of imitation (e.g., Tomasello et al., 2005), remains open. As imitative behavior does not have a single underlying mechanism, but rather represents the interaction of multiple mechanisms and factors (Jones, 2007), future studies should further explore the role of infants' action understanding in imitation.

7 Conclusion

The current dissertation systematically investigated the relation between imitation and temperament in the second year of life to determine whether infants' motivation to imitate changes from an instrumental to a social motivation at this age. To further investigate the role of attention during action demonstration in imitative behavior, infants' goal-predictive gaze shifts and their relation to imitation were examined. In conclusion, no evidence was found for a shift in the underlying motivations for imitation, as the social function of imitation is present from an early age. By identifying infants' individual characteristics and cognitive developmental level as key determinants of imitative behavior, this dissertation contributes significantly to understanding the variations in imitative behavior. The investigation of infants' visual attention using a novel approach in a live eye-tracking setting has important implications for future research on infants' goal-predictive gaze shifts, which are, at least in part, more ecologically valid.

References

Acar, I. H., Frohn, S., Prokasky, A., Molfese, V. J., & Bates, J. E. (2019). Examining the associations between performance based and ratings of focused attention in toddlers: Are we measuring the same constructs? *Infant and Child Development*, *28*(1), Article e2116. https://doi.org/10.1002/icd.2116

Ambrosini, E., Reddy, V., Looper, A. de, Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PLOS ONE*, *8*(7), Article e67916. https://doi.org/10.1371/journal.pone.0067916

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist*, *48*(5), 505–522. https://doi.org/10.1177/0002764204271506

Anisfeld, M. (2005). No compelling evidence to dispute Piaget's timetable of the development of representational imitation in infancy. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation.* MIT Press. https://doi.org/10.7551/mitpress/5331.003.0006

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*(9), 1175–1184. https://doi.org/10.1037/0003-066X.44.9.1175

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development*, *19*(2), 159–170. https://doi.org/10.1016/S0163-6383(96)90015-6

Barr, R., & Hayne, H. (2003). It's not what you know, it's who you know: Older siblings facilitate imitation during infancy. *International Journal of Early Years Education*, *11*(1), 7–21. https://doi.org/10.1080/0966976032000066055

Beisert, M., Zmyj, N., Liepelt, R., Jung, F., Prinz, W., & Daum, M. M. (2012). Rethinking 'rational imitation' in 14-month-old infants: A perceptual distraction approach. *PLOS ONE*, *7*(3), Article e32563. https://doi.org/10.1371/journal.pone.0032563

Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummett, B., & Williams, R. (2009). Vulnerability genes or plasticity genes? *Molecular Psychiatry, 14*(8), 746–754. https://doi.org/10.1038/mp.2009.44

Bornstein, M. H., Putnick, D. L., Gartstein, M. A., Hahn, C.-S., Auestad, N., & O'Connor, D. L. (2015). Infant temperament: Stability by age, gender, birth order, term status, and socioeconomic status. *Child Development*, *86*(3), 844–863. https://doi.org/10.1111/cdev.12367

Bosquet Enlow, M., White, M. T., Hails, K., Cabrera, I., & Wright, R. J. (2016). The Infant Behavior Questionnaire-Revised: Factor structure in a culturally and sociodemographically diverse sample in the United States. *Infant Behavior and Development*, *43*, 24–35. https://doi.org/10.1016/j.infbeh.2016.04.001

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child Development*, *78*(3), 806–824. https://doi.org/10.1111/j.1467-8624.2007.01034.x

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. Wiley-Interscience.

Buss, A. H., & Plomin, R. (1984). *Temperament: Early developing personality traits*. Erlbaum. https://doi.org/10.4324/9781315745701

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences*, *21*(5), 667–721. https://doi.org/10.1017/S0140525X98001745

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, *15*(2), 292–298. https://doi.org/10.1111/j.1467-7687.2011.01127.x

Cannon, E. N., Woodward, A. L., Gredebäck, G., von Hofsten, C., & Turek, C. (2012). Action production influences 12-month-old infants' attention to others' actions. *Developmental Science*, *15*(1), 35–42. https://doi.org/10.1111/j.1467-7687.2011.01095.x

Carpenter, M. (2006). Instrumental, social, and shared goals and intentions in imitation. In S. J. Rogers & J. H. G. Willams (Eds.), *Imitation and the development of the social mind: Autsim and typical development* (pp. 48–70). The Guilford Press.

Carpenter, M., & Call, J. (2009). Comparing the imitative skills of children and nonhuman apes. *Revue De Primatologie.* Advance online publication. https://doi.org/10.4000/primatologie.263

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, *8*(1), F13–F20. https://doi.org/10.1111/j.1467-7687.2004.00385.x

Casalin, S., Luyten, P., Vliegen, N., & Meurs, P. (2012). The structure and stability of temperament from infancy to toddlerhood: A one-year prospective study. *Infant Behavior and Development*, *35*(1), 94–108. https://doi.org/10.1016/j.infbeh.2011.08.004

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, *15*(4), 481–498. https://doi.org/10.1016/S0885-2014(01)00037-5

Chen, X., Yang, F., & Fu, R. (2012). Culture and temperament. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 462–478). The Guilford Press.

Collie, R., & Hayne, H. (1999). Deferred imitation by 6-and 9-month-old infants: More evidence for declarative memory. *Developmental Psychobiology*, *35*(2), 83–90. https://doi.org/10.1002/(SICI)1098-2302(199909)35:2<83::AID-DEV1>3.0.CO;2-S

Costa, R., & Figueiredo, B. (2018). Infant behaviour questionnaire–revised version: a psychometric study in a Portuguese sample. *Journal of Reproductive and Infant Psychology*, *36*(2), 207–218. https://doi.org/10.1080/02646838.2018.1436752

Damm, F., Petermann, F., & Petermann, U. (2011). Imitationsfähigkeit von Kleinkindern in den ersten beiden Lebensjahren. *Psychologische Rundschau, 62*(2), 93–100. https://doi.org/10.1026/0033-3042/a000072

Daum, M. M., Prinz, W., & Aschersleben, G. (2011). Perception and production of object-related grasping in 6-month-olds. *Journal of Experimental Child Psychology*, *108*(4), 810–818. https://doi.org/10.1016/j.jecp.2010.10.003

Davis, J., Redshaw, J., Suddendorf, T., Nielsen, M., Kennedy-Costantini, S., Oostenbroek, J., & Slaughter, V. (2021). Does neonatal imitation exist? Insights from a meta-analysis of 336 effect sizes. *Perspectives on Psychological Science*, *16*(6), 1373–1397. https://doi.org/10.1177/1745691620959834

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science, 335*(6072), 1114–1118. https://doi.org/10.1126/science.1213969

Desmarais, E., Majdandžić, M., Gartstein, M. A., Bridgett, D. J., & French, B. F. (2019). Cross-cultural differences in temperament: Comparing paternal ratings of US and Dutch infants. *European Journal of Developmental Psychology*, *16*(2), 137–151. https://doi.org/10.1080/17405629.2017.1356713

De Pauw, S. S. W., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry and Human Development, 41*(3), 313–329. https://doi.org/10.1007/s10578-009-0171-8

Dixon Jr., W. E., Lawman, H. G., Johnson, E. B. H., May, S., Patton, L. A., Lowe, A. K., & Snyder, C. M. (2012). Effects of exogenous and endogenous distracters on immediate and long-term recall in toddlers. *Infancy*, *17*(5), 525–557. https://doi.org/10.1111/j.1532-7078.2011.00090.x

Dixon Jr., W. E., Salley, B. J., & Clements, A. D. (2006). Temperament, distraction, and learning in toddlerhood. *Infant Behavior and Development*, *29*(3), 342–357. https://doi.org/10.1016/j.infbeh.2006.01.002

Dragan, W. Ł., Kmita, G., & Fronczyk, K. (2011). Psychometric properties of the Polish adaptation of the Infant Behavior Questionnaire—Revised (IBQ-R). *International Journal of Behavioral Development*, *35*(6), 542–549. https://doi.org/10.1177/0165025411422181

Elsner, B. (2014). Theorien zu Imitation und Handlungsverständnis. In L. Ahnert (Ed.), *Theorien in der Entwicklungspsychologie* (pp. 310–329). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34805-1_12

Fagard, J., & Lockman, J. J. (2010). Change in imitation for object manipulation between 10 and 12 months of age. *Developmental Psychobiology*, *52*(1), 90–99. https://doi.org/10.1002/dev.20416

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729

Fenstermacher, S. K., & Saudino, K. J. (2006). Understanding individual differences in young children’s imitative behavior. *Developmental Review*, *26*(3), 346–364. https://doi.org/10.1016/j.dr.2006.05.001

Fenstermacher, S. K., & Saudino, K. J. (2007). Toddler see, toddler do? Genetic and environmental influences on laboratory-assessed elicited imitation. *Behavior Genetics*, *37*(5), 639–647. https://doi.org/10.1007/s10519-007-9160-5

Fenstermacher, S. K., & Saudino, K. J. (2016). Exploring links among imitation, mental development, and temperament. *Infancy*, *21*(5), 536–559. https://doi.org/10.1111/infa.12132

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces during early infancy. *Journal of Experimental Child Psychology*, *118*, 13–26. https://doi.org/10.1016/j.jecp.2013.08.012

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Fu, X., & Pérez-Edgar, K. (2015). Theories of temperament development. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., pp. 191–198). Elsevier Inc. https://doi.org/10.1016/B978-0-08-097086-8.23032-8

Gagne, J. R., van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological Assessment*, *23*(2), 337–353. https://doi.org/10.1037/a0021746

Gampe, A., Prinz, W., & Daum, M. M. (2016). Measuring action understanding: Relations between goal prediction and imitation. *British Journal of Developmental Psychology*, *34*(1), 53–65. https://doi.org/10.1111/bjdp.12125

Ganglmayer, K., Attig, M., Daum, M. M., & Paulus, M. (2019). Infants' perception of goal-directed actions: A multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behavior and Development*, *57*, Article 101340. https://doi.org/10.1016/j.infbeh.2019.101340

Gartstein, M. A., Bogale, W., & Meehan, C. L. (2016). Adaptation of the Infant Behavior Questionnaire-Revised for use in Ethiopia: Expanding cross-cultural investigation of temperament development. *Infant Behavior and Development*, *45*, 51–63. https://doi.org/10.1016/j.infbeh.2016.09.004

Gartstein, M. A., Gonzalez, C., Carranza, J. A., Ahadi, S. A., Ye, R., Rothbart, M. K., & Yang, S. W. (2006). Studying cross-cultural differences in the development of infant temperament: People's Republic of China, the United States of America, and Spain. *Child Psychiatry and Human Development*, *37*(2), 145–161. https://doi.org/10.1007/s10578-006-0025-6

Gartstein, M. A., Kirchhoff, C. M., & Lowe, M. E. (2024). Individual differences in temperament: A developmental perspective. In J. D. Osofsky, H. E. Fitzgerald, M. Keren, & K. Puura (Eds.), *WAIMH Handbook of infant and early childhood mental health: Biopsychosocial factors* (Vol. 1, pp. 31–48). Springer International Publishing. https://doi.org/10.1007/978-3-031-48627-2_3

Gartstein, M. A., Prokasky, A., Bell, M. A., Calkins, S., Bridgett, D. J., Braungart-Rieker, J., Leerkes, E., Cheatham, C. L., Eiden, R. D., Mize, K. D., Jones, N. A., Mireault, G., & Seamon, E. (2017). Latent profile and cluster analysis of infant temperament: Comparisons across person-centered approaches. *Developmental Psychology*, *53*(10), 1811–1825. https://doi.org/10.1037/dev0000382

Gartstein, M. A., Putnick, D., Kwak, K., Hahn, C.-S., & Bornstein, M. H. (2015). Stability of temperament in South Korean infants from 6 to 12 to 18 months: Moderation by age, gender, and birth order. *Infant Behavior and Development*, *40*, 103–107. https://doi.org/10.1016/j.infbeh.2015.05.002

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior and Development*, *26*(1), 64–86. https://doi.org/10.1016/S0163-6383(02)00169-8

Gellén, K., & Buttelmann, D. (2019). Rational imitation declines within the second year of life: Changes in the function of imitation. *Journal of Experimental Child Psychology*, *185*, 148–163. https://doi.org/10.1016/j.jecp.2019.04.019

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*(6873), 755. https://doi.org/10.1038/415755a

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, *4*, Article 527. https://doi.org/10.3389/fpsyg.2013.00527

Goertz, C., Kolling, T., Frahsek, S., & Knopf, M. (2008). Die Frankfurter Imitationstests für 18 Monate alte Kinder (FIT 18) und für 24 Monate alte Kinder (FIT 24): Entwicklung altersangepasster Instrumente für die Verzögerte Imitation [The Frankfurt Imitation Test for 18-and 24-month-old infants]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *40*, 152–160. https://doi.org/10.1026/0049-8637.40.3.152

Goldsmith, H. H., Lemery, K. S., Buss, K. A., & Campos, J. J. (1999). Genetic analyses of focal aspects of infant temperament. *Developmental Psychology*, *35*(4), 972–985. https://doi.org/10.1037/0012-1649.35.4.972

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., Hinde, R. A., & McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child Development*, *58*(2), 505–529. https://doi.org/10.2307/1130527

Goldsmith, H. H., Buss, K. A., & Lemery, K. S. (1997). Toddler and childhood temperament: Expanded content, stronger genetic evidence, new evidence for the importance of environment. *Developmental Psychology*, *33*(6), 891–905. https://doi.org/10.1037/0012-1649.33.6.891

Goldsmith, H. H., & Campos, J. J. (1982). Toward a theory of infant temperament. In R. N. Emde & R. J. Harmon (Eds.), *The development of attachment and affiliative systems* (pp. 161–193). Springer US. https://doi.org/10.1007/978-1-4684-4076-8_13

Goldsmith, H. H., & Gagne, J. R. (2012). Behavioral assessment of temperament. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 209–228). The Guilford Press.

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2021). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, *17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Gredebäck, G., & Falck-Ytter, T. (2015). Eye movements during action observation. *Perspectives on Psychological Science*, *10*(5), 591–598. https://doi.org/10.1177/1745691615589103

Hayne, H., & Herbert, J. (2004). Verbal cues facilitate memory retrieval during infancy. *Journal of Experimental Child Psychology*, *89*(2), 127–139. https://doi.org/10.1016/j.jecp.2004.06.002

Heimann, M. (2002). Notes on individual differences and the assumed elusiveness of neonatal imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 74–84). Cambridge University Press. https://doi.org/10.1017/CBO9780511489969.005

Heimann, M. (2022a). A fresh look on neonatal imitation. In M. Heimann (Ed.), *Imitation from infancy through early childhood: Typical and atypical development* (pp. 9–28). Springer International Publishing. https://doi.org/10.1007/978-3-031-08899-5_2

Heimann, M. (2022b). Imitation and temperament in infancy. In M. Heimann (Ed.), *Imitation from infancy through early childhood: Typical and atypical development* (pp. 65–78). Springer International Publishing. https://doi.org/10.1007/978-3-031-08899-5_5

Henderson, H. A., & Wachs, T. D. (2007). Temperament theory and the study of cognition–emotion interactions across development. *Developmental Review*, *27*(3), 396–427. https://doi.org/10.1016/j.dr.2007.06.004

Hilbrink, E. E., Sakkalou, E., Ellis-Davies, K., Fowler, N. C., & Gattis, M. (2013). Selective and faithful imitation at 12 and 15 months. *Developmental Science*, *16*(6), 828–840. https://doi.org/10.1111/desc.12070

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*(2), 446–454. https://doi.org/10.1037/a0016543

Hunnius, S., & Bekkering, H. (2014). What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1644), Article 20130490. https://doi.org/10.1098/rstb.2013.0490

Ingersoll, B. (2008). The social role of imitation in autism. *Infants and Young Children*, *21*(2), 107–119. https://doi.org/10.1097/01.IYC.0000314482.24087.14

Ingersoll, B., & Meyer, K. (2011). Examination of correlates of different imitative functions in young children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(3), 1078–1085. https://doi.org/10.1016/j.rasd.2010.12.001

Jones, S. S. (2007). Imitation in infancy: The development of mimicry. *Psychological Science*, *18*(7), 593–599. https://doi.org/10.1111/j.1467-9280.2007.01945.x

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2325–2335. https://doi.org/10.1098/rstb.2009.0045

Kagan, J., & Fox, N. A. (2006). Biology, culture, and temperamental biases. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (6th ed., Vol. 3, pp. 167-225). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0304

Kiel, N., Bruckdorfer, R., Petermann, F., & Reinelt, T. (2018). Temperament in early childhood and developmental externalizing disorders: clinical implications. *Zeitschrift Für Psychiatrie, Psychologie Und Psychotherapie*, *66*(3), 177–186. https://doi.org/10.1024/1661-4747/a000354

Kirchhoff, M. C. (2015). *Temperament and behavioral difficulties across cultures: A comparison between German and U.S. toddlers applying the concept of the 'developmental niche'* (Publication No. 10043063). [Doctoral dissertation, Washington State University]. ProQuest Dissertations and Theses Global.

Koch, F.-S., Sundqvist, A., Herbert, J., Tjus, T., & Heimann, M. (2018). Changes in infant visual attention when observing repeated actions. *Infant Behavior and Development*, *50*, 189–197. https://doi.org/10.1016/j.infbeh.2018.01.003

Kolling, T., Óturai, G., & Knopf, M. (2014). Is selective attention the basis for selective imitation in infants? An eye-tracking study of deferred imitation with 12-month-olds. *Journal of Experimental Child Psychology*, *124*, 18–35. https://doi.org/10.1016/j.jecp.2014.01.016

Krol, M. A., & Jellema, T. (2022). Sensorimotor anticipation of others' actions in real-world and video settings: Modulation by level of engagement? *Social Neuroscience*, *17*(3), 293–304. https://doi.org/10.1080/17470919.2022.2083229

Kupán, K., Király, I., Kupán, K., Krekó, K., Miklósi, Á., & Topál, J. (2017). Interacting effect of two social factors on 18-month-old infants' imitative behavior: Communicative cues and demonstrator presence. *Journal of Experimental Child Psychology*, *161*, 186–194. https://doi.org/10.1016/j.jecp.2017.03.019

Liu, Y., Hau, K.-T., & Zheng, X. (2020). Does instrumental motivation help students with low intrinsic motivation? Comparison between Western and Confucian students. *International Journal of Psychology*, *55*(2), 182–191. https://doi.org/10.1002/ijop.12563

Majdandžić, M., & van den Boom, D. C. (2007). Multimethod longitudinal assessment of temperament in early childhood. *Journal of Personality*, *75*(1), 121–168. https://doi.org/10.1111/j.1467-6494.2006.00435.x

Martin, R. P., Wisenbaker, J., & Huttunen, M. (1994). Review of factor analytic studies of temperament measures based on the Thomas-Chess structural model: Implications for the Big Five. In C. F. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 157–172). Lawrence Erlbaum Associates, Inc.

McCall, R. B., Parke, R. D., Kavanaugh, R. D., Engstrom, R., Russell, J., & Wycoff, E. (1977). Imitation of live and televised models by children one to three years of age. *Monographs of the Society for Research in Child Development*, *42*(5), 1–94. https://doi.org/10.2307/1165913

McDougall, W. (1908). *An introduction to social psychology.* Methuen.

Mehrpour, S., & Vojdani, M. (2012). Globalization and EFL learning motivation: A new perspective on integrative vs. instrumental motivation among Iranian learners of English. *Open Journal of Modern Linguistics*(2), 43–50. https://doi.org/10.4236/ojml.2012.22006

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, *24*(4), 470–476. https://doi.org/10.1037/0012-1649.24.4.470

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850. https://doi.org/10.1037/0012-1649.31.5.838

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*(4312), 75–78. https://doi.org/10.1126/science.198.4312.75

Melzer, A., Prinz, W., & Daum, M. M. (2012). Production and perception of contralateral reaching: A close link by 12 months of age. *Infant Behavior and Development*, *35*(3), 570–579. https://doi.org/10.1016/j.infbeh.2012.05.003

Mink, D., Henning, A., & Aschersleben, G. (2013). Visual habituation tasks: The role of infant temperament. *Infant Behavior and Development*, *36*(3), 377–390. https://doi.org/10.1016/j.infbeh.2013.03.005

Monroy, C., Chen, C.-H., Houston, D., & Yu, C. (2021). Action prediction during real-time parent-infant interactions. *Developmental Science*, *24*(3), Article e13042. https://doi.org/10.1111/desc.13042

Montirosso, R., Cozzi, P., Putnam, S. P., Gartstein, M. A., & Borgatti, R. (2011). Studying cross-cultural differences in temperament in the first year of life: United States and Italy. *International Journal of Behavioral Development*, *35*(1), 27–37. https://doi.org/10.1177/0165025410368944

Nakagawa, A., Matsuki, T., Tomida, M., Miyachi, T., Ebara, T., & Kamijima, M. (2024). Development of temperamental regulation of infants at 6 and 24 months: Associations with maternal soothing and distress. *Health Science Reports*, *7*(9), Article e70036. https://doi.org/10.1002/hsr2.70036

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, *42*(3), 555–565. https://doi.org/10.1037/0012-1649.42.3.555

Nielsen, M. (2009). The imitative behaviour of children and chimpanzees: A window on the transmission of cultural traditions. *Revue De Primatologie.* Advance online publication. https://doi.org/10.4000/primatologie.254

Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental Psychology*, *47*(4), 1050–1053. https://doi.org/10.1037/a0023866

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science*, *11*(5), 722–731. https://doi.org/10.1111/j.1467-7687.2008.00722.x

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Clark, S., & Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology*, *26*(10), 1334–1338. https://doi.org/10.1016/j.cub.2016.03.047

Óturai, G., Kolling, T., & Knopf, M. (2013). Relations between 18-month-olds' gaze pattern and target action performance: A deferred imitation study with eye tracking. *Infant Behavior and Development*, *36*(4), 736–748. https://doi.org/10.1016/j.infbeh.2013.08.005

Óturai, G., Kolling, T. & Knopf, M. (2018). Developmental trend towards exact imitation in the second year of life: Evidence from a longitudinal study. *International Journal of Behavioral Development*, *42*(4), 388–395. https://doi.org/10.1177/0165025417713727

Over, H. (2020). The social function of imitation in development. *Annual Review of Developmental Psychology*, *2*(1), 93–109. https://doi.org/10.1146/annurev-devpsych-033020-024051

Over, H., & Carpenter, M. (2009). Priming third-party ostracism increases affiliative imitation in children. *Developmental Science*, *12*(3), F1–F8. https://doi.org/10.1111/j.1467-7687.2008.00820.x

Over, H., & Carpenter, M. (2013). The social side of imitation. *Child Development Perspectives*, *7*(1), 6–11. https://doi.org/10.1111/cdep.12006

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, *9*(6), 137–157. https://doi.org/10.1177/014272378900900603

Paulus, M. (2011). Imitation in infancy: Conceptual considerations. *Theory & Psychology*, *21*(6), 849–856. https://doi.org/10.1177/0959354310395990

Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011). Imitation in infancy: Rational or motor resonance? *Child Development*, *82*(4), 1047–1057. https://doi.org/10.1111/j.1467-8624.2011.01610.x

Paulus, M., Schuwerk, T., Sodian, B., & Ganglmayer, K. (2017). Children's and adults' use of verbal information to visually anticipate others' actions: A study on explicit and implicit social-cognitive processing. *Cognition*, *160*, 145–152. https://doi.org/10.1016/j.cognition.2016.12.013

Piaget, J. (1962). *Play, dreams and imitation in childhood*. Routledge.

Planalp, E. M., van Hulle, C., Gagne, J. R., & Goldsmith, H. H. (2017). The infant version of the Laboratory Temperament Assessment Battery (Lab-TAB): Measurement properties and implications for concepts of temperament. *Frontiers in Psychology*, *8,* Article 846. https://doi.org/10.3389/fpsyg.2017.00846

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science, 264*(5166), 1733–1739. https://doi.org/10.1126/science.8209254

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development*, *29*(3), 386–401. https://doi.org/10.1016/j.infbeh.2006.01.004

Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and Child Development*, *17*(4), 387–405. https://doi.org/10.1002/icd.582

Putnam, S. P., Sanson, A. V., & Rothbart, M. K. (2002). Child temperament and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting: Children and parenting* (2nd ed., Vol.1, pp. 255–277). Lawrence Erlbaum Associates Publishers.

Putnam, S. P., Sehic, E., French, B. F., Gartstein, M. A., Lira Luttges, B., & 486 Additional Partners in the Global Temperament Project. (2024). The Global Temperament Project: Parent-reported temperament in infants, toddlers, and children from 59 nations. *Developmental Psychology, 60*(5), 916–941. https://doi.org/10.1037/dev0001732

Rogers, S. J., Hepburn, S. L., Stackhouse, T., & Wehner, E. (2003). Imitation performance in toddlers with autism and those with other developmental disorders. *Journal of Child Psychology and Psychiatry*, *44*(5), 763–781. https://doi.org/10.1111/1469-7610.00162

Rothbart, M. K. (2011). *Becoming who we are: Temperament and personality in development.* Guilford Press.

Rothbart, M. K., & Bates, J. (2006). Temperament. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (6th ed., Vol. 3, pp. 99-166). John Wiley & Sons. https://doi.org/10.1002/9780470147658.chpsy0303

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology*, *103*(1), 55–66. https://doi.org/10.1037/0021-843X.103.1.55

Rothbart, M. K., & Goldsmith, H. H. (1985). Three approaches to the study of infant temperament. *Developmental Review*, *5*(3), 237–260. https://doi.org/10.1016/0273-2297(85)90012-7

Rothbart, M. K., & Mauro, J. A. (1990). Questionnaire approaches to the study of infant temperament. In J. Colombo & J. W. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction* (pp. 411–429). Lawrence Erlbaum Associates, Inc.

Rothbart, M. K., & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (pp. 37–86). Erlbaum. https://doi.org/10.4324/9780203728185-2

Sanson, A. V., Smart, D. F., Prior, M., Oberklaid, F., et al. (1994). The structure of temperament from age 3 to 7 years: Age, sex, and sociodemographic influences. *Merrill-Palmer Quarterly, 40*(2), 233–252.

Saudino, K. J. (2009). The development of temperament from a behavioral genetics perspective. *Advances in Child Developement and Behavior*, *37*, 201–231. https://doi.org/10.1016/S0065-2407(09)03705-7

Schmidt, J. A., Aschersleben, G., & Henning, A. (2024). Early-life temperament: A longitudinal investigation of stability and structure. *International Journal of*

Behavioral Development, Article 01650254241290999. https://doi.org/10.1177/01650254241290999

Schroeder, K. B., Asherson, P., Blake, P. R., Fenstermacher, S. K., & Saudino, K. J. (2016). Variant at serotonin transporter gene predicts increased imitation in toddlers: relevance to the human capacity for cumulative culture. *Biology Letters*, *12*(4), Article 20160106. https://doi.org/10.1098/rsbl.2016.0106

Schwier, C., van Maanen, C., Carpenter, M., & Tomasello, M. (2006). Rational imitation in 12-month-old infants. *Infancy*, *10*(3), 303–311. https://doi.org/10.1207/s15327078in1003_6

Seehagen, S., Konrad, C., Herbert, J. S., & Schneider, S. (2015). Timely sleep facilitates declarative memory consolidation in infants. *Proceedings of the National Academy of Sciences*, *112*(5), 1625–1629. https://doi.org/10.1073/pnas.1414000112

Shiner, R. L. (2015). The development of temperament and personality traits in childhood and adolescence. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, Vol. 4. Personality processes and individual differences* (pp. 85–105). American Psychological Association. https://doi.org/10.1037/14343-004

Shiner, R. L., Buss, K. A., McClowry, S. G., Putnam, S. P., Saudino, K. J., & Zentner, M. (2012). What is temperament now? Assessing progress in temperament research on the Twenty-Fifth Anniversary of Goldsmith et al. (1987). *Child Development Perspectives*, *6*(4), 436–444. https://doi.org/10.1111/j.1750-8606.2012.00254.x

Shiner, R. L., & DeYoung, C. G. (2013). The structure of temperament and personality traits: A developmental perspective. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology, Vol. 2: Self and other* (pp. 113–141). Oxford University Press.

Sonne, T., Kingo, O. S., & Krøjgaard, P. (2016). Empty looks or paying attention? Exploring infants' visual behavior during encoding of an elicited imitation task. *Infancy*, *21*(6), 728–750. https://doi.org/10.1111/infa.12141

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, *44*(5), 430–444. https://doi.org/10.1037/h0062885

Stępień-Nycz, M., Rostek, I., Białecka-Pikul, M., & Białek, A. (2018). The Polish adaptation of the Early Childhood Behavior Questionnaire (ECBQ): Psychometric properties, age and gender differences and convergence between the questionnaire and the observational data. *European Journal of Developmental Psychology*, *15*(2), 192–213. https://doi.org/10.1080/17405629.2017.1292906

Sukigara, M., Nakagawa, A., & Mizuno, R. (2015). Development of a Japanese version of the Early Childhood Behavior Questionnaire (ECBQ) using cross-sectional and longitudinal data. *Sage Open*, *5*(2), Article 2158244015590443. https://doi.org/10.1177/2158244015590443

Sung, J., Beijers, R., Gartstein, M. A., de Weerth, C., & Putnam, S. P. (2015). Exploring temperamental differences in infants from the United States of America (US) and the Netherlands. *European Journal of Developmental Psychology*, *12*(1), 15–28. https://doi.org/10.1080/17405629.2014.937700

Taylor, G., & Herbert, J. S. (2013). Eye tracking infants: Investigating the role of attention during learning on recognition memory. *Scandinavian Journal of Psychology*, *54*(1), 14–19. https://doi.org/10.1111/sjop.12002

Taylor, G., & Herbert, J. S. (2014). Infant and adult visual attention during an imitation demonstration. *Developmental Psychobiology, 56*(4), 770–782. https://doi.org/10.1002/dev.21147

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2021). Infants’ preference for social interactions increases from 7 to 13 months of age. *Child Development*, *92*(6), 2577–2594. https://doi.org/10.1111/cdev.13636

Thomas, A., & Chess, S. (1977). *Temperament and development*. Brunner/Mazel.

Thomas, A., Chess, S., Birch, H. G., Hertzig, M. E., & Korn, S. (1963). *Behavioral individuality in early childhood*. *Behavioral individuality in early childhood*. New York University Press. https://doi.org/10.1037/14328-000

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*(4), i-109. https://doi.org/10.1037/h0092987

Tomasello, M. (1999). *The cultural origins of human cognition.* Havard University Press.

Tomasello, M., & Carpenter, M. (2005). Intention reading and imitative learning. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science: Imitation, human development, and culture* (Vol. 2, pp. 133–148). MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–735. https://doi.org/10.1017/S0140525X05000129

Užgiris, I. C. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development*, *4*(1), 1–12. https://doi.org/10.1177/016502548100400101

Vonderlin, E., Ropeter, A., & Pauen, S. (2012). Assessment of temperament with the Infant Behavior Questionnaire Revised (IBQ-R) - the psychometric properties of a German version [Erfassung des frühkindlichen Temperaments mit dem Infant Behavior Questionnaire Revised]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, *40*(5), 307–314. https://doi.org/10.1024/1422-4917/a000187

Vroman, L. N., Lo, S. L., & Durbin, C. E. (2014). Structure and convergent validity of children's temperament traits as assessed by experimenter ratings of child behavior. *Journal of Research in Personality*, *52*, 6–12. https://doi.org/10.1016/j.jrp.2014.06.002

Wagner, J. B., Luyster, R. J., Yim, J. Y., Tager-Flusberg, H., & Nelson, C. A. (2013). The role of early visual attention in social development. *International Journal of Behavioral Development*, *37*(2), 118–124. https://doi.org/10.1177/0165025412468064

Whiten, A., & Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. *Advances in the Study of Behavior, 21,* 239–283. https://doi.org/10.1016/S0065-3454(08)60146-1

Whiten, A. (2005). The second inheritance system of chimpanzees and humans. *Nature*, *437*(7055), 52–55. https://doi.org/10.1038/nature04023

Yu, B., & Downing, K. (2012). Determinants of international students' adaptation: examining effects of integrative motivation, instrumental motivation and second language proficiency. *Educational Studies*, *38*(4), 457–471. https://doi.org/10.1080/03055698.2011.643111

Yu, Y., & Kushnir, T. (2014). Social context effects in 2- and 4-year-olds' selective versus faithful imitation. *Developmental Psychology*, *50*(3), 922–933. https://doi.org/10.1037/a0034242

Yu, Y., & Kushnir, T. (2020). The ontogeny of cumulative culture: Individual toddlers vary in faithful imitation and goal emulation. *Developmental Science*, *23*(1), Article e12862. https://doi.org/10.1111/desc.12862

Zentner, M., & Bates, J. E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. *International Journal of Developmental Science*, *2*, 7–37. https://doi.org/10.3233/DEV-2008-21203

Zmyj, N., Daum, M. M., & Aschersleben, G. (2009). The development of rational imitation in 9-and 12-month-old infants. *Infancy*, *14*(1), 131–141. https://doi.org/10.1080/15250000802569884

Zmyj, N., Schölmerich, A., & Daum, M. M. (2017). The relationship between attention and deferred imitation in 12-month-old infants. *Infant Behavior and Development*, *48*, 175–183. https://doi.org/10.1016/j.infbeh.2017.02.0

List of Tables

Article 1 – Sieber, F., & Zmyj, N. (2022). Stability and Structure of Infant and Toddler Temperament in Two Longitudinal Studies in Germany. *Infant Behavior and Development, 67*, 101714.

Table 1 *Scale definitions of the IBQ-R (Gartstein & Rothbart, 2003)* 36

Table 2 *Scale defintions of the ECBQ (Putnam et al., 2006)* 37

Table 3 *IBQ-R exploratory factor loadings at 9 and 12 months*...45

Table 4 *Descriptive Statistics and Correlations of the IBQ-R scales at 9 and 12 months*.....46

Table 5 *IBQ-R exploratory factor loadings at 12 months*..50

Table 6 *ECBQ exploratory factor loadings at 18 and 24 months*.......................................52

Table 7 *Descriptive Statistics, Correlations and significant age effects of the IBQ-R and ECBQ scales at 12, 18, and 24 months* 54

Article 2 – Sieber, F., Czarnomski, J., Schölmerich, A., Daum, M. M., & Zmyj, N. (2024). The two functions of imitation in the second year of life: A longitudinal study. *Developmental Psychology*. Advance Online Publication.

Table 1 *Number of Children, Children's Sex, Children's Mean Age (Months; Days) with Standard Deviation in Parentheses (Days) and Age Range (Days) at 12, 18, and 24 Months* 89

Table 2 *Overview of Materials and Tasks of the Imitation Test*...91

Table 3 *Mean Scores and Standard Deviations of IBQ-R scales at 12 months* 101

Table 4 *Mean Scores and Standard Deviations of ECBQ scales at 18 and 24 months*.......102

Table 5 *Mean Scores and Standard Deviations of Task Orientation Test Variables* 103

Table 6 *Mean Scores and Standard Deviations of Puppet Game Test Variables* 103

Table 7 *Hierarchical Regression Analysis Predicting the Imitation Score at 12 months*.... 109

Table 8 *Hierarchical Regression Analysis Predicting the Imitation Score at 18 months*.... 109

Table 9 *Hierarchical Regression Analysis Predicting the Imitation Score at 24 months*.... 110

Table 10 *Hierarchical Regression Analysis Predicting the Imitation Score Longitudinally from 12 to 24 months*........ 111

Table 11 *Hierarchical Regression Analysis Predicting the Imitation Score Longitudinally from 18to 24 months*........ 111

Table S1 *Number of Participants Excluded at Each Time Point and Reasons for Exclusion*........ 129

Table S2 *Overview of Items Included in the Rationally Derived Sociability Scale Including their Original Scale Assignment*........ 131

Table S3 *Results of ICC Calculations of the Task Orientation Test Scores at 12, 18, and 24 Months*........ 137

Table S4 *Results of ICC Calculations of the Puppet Game Test Scores at 12, 18, and 24 Months*........ 138

Table S5 *Correlations of FIT 12 Test Score and Attentional and Social Aspects of Temperament and Intercorrelations of Variables from the Same Test Procedure*.139

Table S6 *Correlations of FIT 18 Test Score and Attentional and Social Aspects of Temperament and Intercorrelations of Variables from the Same Test Procedure*.140

Table S7 *Correlations of FIT 24 Test Score and Attentional and Social Aspects of Temperament and Intercorrelations of Variables from the Same Test Procedure*.141

Article 3 – Sieber, F., Czarnomski, J., Daum, M.M., & Zmyj, N. (2024). The relation between goal-predictive gaze behavior and imitation – A live eye-tracking study in 12-months-olds. *PsyArXiv Preprints.*

Table 1 *Hierarchical Regression Analysis Predicting the Imitation Score* 155

List of Figures

Article 3 – Sieber, F., Czarnomski, J., Daum, M.M., & Zmyj, N. (2024). The relation between goal-predictive gaze behavior and imitation – A live eye-tracking study in 12-months-olds. *PsyArXiv Preprints.*

Figure 1 *Experimental setup during calibration* 151

Figure 2 *Experimental setup during demonstration* 151

Figure 3 *Exemplary definition of the AoIs for one trial* 153

Figure 4 *Box plots representing mean gaze-arrival times of imitators and non-imitators* .. 155

List of Articles and Own Contribution to the Publications

Article 1

Sieber, F., & Zmyj, N. (2022). Stability and structure of infant and toddler temperament in two longitudinal studies in Germany. *Infant Behavior and Development*, *67*, 101714. https://doi.org/10.1016/j.infbeh.2022.101714

Conception and design of study

Analysis and interpretation of data

Drafting the manuscript

Revising the manuscript

Article 2

Sieber, F., Czarnomski, J., Schölmerich, A., Daum, M. M., & Zmyj, N. (2024). The two functions of imitation in the second year of life: A longitudinal study. *Developmental Psychology.* Advance online publication. https://doi.org/10.1037/dev0001856

Data curation

Formal analysis

Project administration

Validation

Visualization

Writing – original draft

Writing – review and editing

Article 3

Sieber, F., Czarnomski, J., Daum, M. M., & Zmyj, N. (2024). The relation between goal predictive gaze behavior and imitation – A live eye-tracking study in 12-months-olds. *PsyArXiv.* https://doi.org/10.31234/osf.io/4mfys

Data curation

Formal analysis

Project administration

Validation

Visualization

Writing – Original Draft Preparation

Writing – Review & Editing

Eidesstattliche Versicherung

Hiermit versichere ich schriftlich und eidesstattlich gemäß § 11 Abs. 2 PromO v. 08.05.2013:

1. Die von mir vorgelegte Dissertation ist selbstständig verfasst und alle in Anspruch genommenen Quellen und Hilfen sind in der Dissertation vermerkt worden.
2. Die von mir eingereichte Dissertation ist weder in der gegenwärtigen noch in einer anderen Fassung an der Technischen Universität Dortmund oder an einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegt worden.
3. Weiterhin erkläre ich schriftlich und eidesstattlich, dass mir der „Ratgeber zur Verhinderung von Plagiaten“ und die „Regeln guter wissenschaftlicher Praxis der Technischen Universität Dortmund“ bekannt und von mir in der vorgelegten Dissertation befolgt worden sind.

______________________________ ______________________________

Ort, Datum Franziska Sieber