Simon WEIXLER, Daniel SOMMERHOFF & Stefan UFER, München

Eine Intuition für das empirische Gesetz der großen Zahlen? Systematische Analyse des Einflusses multipler Aufgaben- & Personenmerkmale beim „hospital problem“

In der Forschung zum empirischen Gesetz der großen Zahlen (eGdgZ) wird seit Jahrzehnten kontrovers diskutiert (vgl. Lem, 2015), ob eine Intuition im Sinne eines natürlichen Gespürs für dieses Gesetz existiert. Um dies zu untersuchen werden oftmals Aufgaben eingesetzt, die auf Kahneman und Tverskys (1972, S. 443) „hospital problem“ basieren. Da sich die eingesetzten Aufgaben jedoch zumeist in ein oder mehreren Merkmalen unterscheiden und zudem die untersuchten Personengruppen verschieden sind, verwundert es kaum, wenn die erhaltenen Lösungsraten unterschiedlich sind und zu widersprüchlichen Interpretationen führen (Lem, van Dooren, Gillard, & Verschaffel, 2011): Hohe Lösungsraten werden als Argument für die Existenz einer Intuition für das eGdgZ herangezogen (z. B. Sedlmeier & Gigerenzer, 1997), niedrige Lösungsraten als Argument zur Widerlegung, wobei letztere mit der Identifikation gleicher relativer Häufigkeiten für beide Krankenhäuser (z. B. 60%) als (vermeintlich) entscheidendes Merkmal für die Problemlösung erklärt werden (z. B. Fischbein & Schnarch, 1997).

Ziel der hier berichteten Studie war, die widersprüchlichen Ergebnisse zu erklären und Aufgabenmerkmale zu identifizieren, die Schülerinnen und Schülern die Problemlösung erleichtern. Dazu wurde in einer Querschnittstudie mit $N = 242$ Studierenden des Lehramts Mathematik an Realschulen bzw. Gymnasien der Einfluss verschiedener Aufgaben- und Personenmerkmale auf die Lösungsrate systematisch untersucht (ausführlich in Weixler, Sommerhoff & Ufer, 2019). Weiterhin sollte exploriert werden, ob Aufgaben mit bestimmten Merkmalen das Bearbeiten nachfolgender Aufgaben erleichtern.

1. Aufgaben- und Personenmerkmale bei Varianten des „hospital problem“

Das eGdgZ besagt, dass eine große Stichprobe besser geeignet ist als eine kleine, um einen Parameter der Grundgesamtheit zu schätzen (Sedlmeier & Gigerenzer, 1997). Zugrunde liegt hierbei eine „statisch-komparative“ Sichtweise kleiner und großer Stichproben (vgl. Schnell & Prediger, 2012), welche in Kahneman und Tverskys (1972) „hospital problem“ aufgegriffen wird. Dabei variierten in der Vergangenheit jedoch oftmals verschiedene Merkmale der Aufgabenstellung innerhalb und zwischen Studien (s. Lem et al., 2011), u. a.:

- *Problemkontext*. Anstelle des Kontexts „Geburten“ wurde u. a. der Kontext „Münzwürfe“ verwendet (z. B. Fischbein & Schnarch, 1997; Watson & Callingham, 2013). Einen systematischen Vergleich dieser beiden Kontexte konnten wir in bisherigen empirischen Studien nicht finden.
- *Verhältnis der Stichprobengrößen*. Ein größeres Verhältnis zwischen der großen und der kleinen Stichprobe könnte den Unterschied in den Stichprobengrößen salienter machen. Murray, Iding, Farris und Revlin (1987) berichten diesbezüglich von positiven Tendenzen in den Lösungsraten von Problemvarianten mit größerem Verhältnis.
- Im Fokus stehende *relative Häufigkeit*. In Kahneman und Tverskys (1972) „hospital problem“ wird eine Geburtenhäufigkeit von mehr als 60% Jungen betrachtet. Die zu erwartende relative Häufigkeit liegt bei ca. 50%. Je stärker die im Fokus stehende relative Häufigkeit von der zu erwartenden relativen Häufigkeit abweicht, desto salienter könnte das Außergewöhnliche an der Beobachtung werden (Lem, 2015).
- *Darstellung der relativen Häufigkeit 100%*. Im Fall der maximalen Abweichung von der zu erwartenden relativen Häufigkeit ist anstelle der numerischen Angabe („100% Jungen“) eine verbale Beschreibung („nur Jungen“) möglich (vgl. z. B. Engel & Sedlmeier, 2005). Die Stichprobengrößen verbleiben hierbei als einzige numerische Angaben im Aufgabentext.
- *Reihenfolge, in der die Stichprobengrößen aufgeführt werden*. In Kahneman und Tverskys (1972) „hospital problem“ wird die größere Stichprobe zuerst aufgeführt, in anderen Problemvarianten (z. B. Fischbein & Schnarch, 1997; Rubel, 2009) hingegen die kleinere. Dass dieses Merkmal kontrolliert wurde, konnten wir in bisherigen empirischen Studien nicht finden.

Unsere zentrale Hypothese war, dass sich Unterschiede in den Lösungsraten auf die Salienz für die Lösung relevanter Aufgabenmerkmale zurückführen lassen. Neben Aufgabenmerkmalen lässt sich auch ein Einfluss von Personenmerkmalen auf die Lösungsrate vermuten, u. a. von allgemeinen kognitiven Fähigkeiten und Bildungsstand, gemessen in Form des *Abiturschnitts* (Roth et al., 2015), und vom *Geschlecht* (Watson, 2000).

2. Fragestellungen

Im Zentrum unserer Studie standen folgende Fragestellungen: (1) Inwiefern beeinflussen (a) Aufgaben- und (b) Personenmerkmale die Lösungsrate bei Varianten des „hospital problem“? (2) Lassen sich Interaktionen zwischen Aufgaben- und Personenmerkmalen feststellen? (3) Gibt es bestimmte

Problemvarianten, die zu höheren Lösungsraten bei nachfolgenden Aufgaben führen (in Relation zu vergleichbaren, vorangegangenen Aufgaben)?

3. Methode

Die Teilnehmer bearbeiteten jeweils eine randomisierte Sequenz aus 16 Varianten des „hospital problem“, in welcher der Problemkontext (Geburten vs. Münzwürfe), das Verhältnis der Stichprobengrößen (100 und 10 vs. 1000 und 10), die im Fokus stehende relative Häufigkeit (70% vs. 90% vs. 100%) sowie – im Fall der maximalen Abweichung von der zu erwartenden relativen Häufigkeit – die Darstellung der relativen Häufigkeit 100% (numerisch vs. verbal) systematisch variierte. Die Formulierung der Aufgabentexte war hierbei an das „Problem 5B. The Effect of Sample Size“ von Fischbein und Schnarch (1997, S. 99) angelehnt. Die Reihenfolge, in der die Stichprobengrößen in den Aufgabentexten aufgeführt waren, variierte innerhalb der Sequenz wie auch zwischen den Teilnehmern. Die Bearbeitungen der Aufgaben wurden dichotom codiert. Die Daten wurden mit Generalisierten Linearen Misch-Modellen in R unter Einsatz des „lme4“ Pakets analysiert.

4. Ergebnisse

(1a) Für den Problemkontext zeigte sich kein signifikanter Einfluss auf die Lösungsrate. Für ein größeres Verhältnis zwischen der großen und der kleinen Stichprobe ergab sich hingegen eine signifikant höhere Lösungsrate, ebenso für die relative Häufigkeit 100% im Vergleich zur relativen Häufigkeit 70%. Auch für die Reihenfolge „zuerst große, dann kleine Stichprobe“ ergab sich eine signifikant höhere Lösungsrate. Für die numerische Darstellung der relativen Häufigkeit 100% ergab sich eine signifikant niedrigere Lösungsrate als für die verbale Darstellung. (1b) Für den Abiturschnitt zeigte sich kein signifikanter Einfluss auf die Lösungsrate, wohingegen sich für weibliche Teilnehmer eine signifikant niedrigere Lösungsrate ergab als für männliche. (2) Für weibliche Teilnehmer ergab sich zudem im Problemkontext „Münzwürfe“ eine signifikant geringere Lösungsrate als im Problemkontext „Geburten“. Ein besserer Abiturschnitt hatte einen stärkeren Einfluss auf die Lösungsrate im Fall der relativen Häufigkeit 100% als im Fall der relativen Häufigkeit 70%. (3) Für Aufgaben mit der relativen Häufigkeit 70% und 90% zeigte sich eine signifikant höhere Lösungsrate, wenn diese nach der ersten Aufgabe in der Sequenz mit relativer Häufigkeit 100% auftraten.

5. Diskussion

Die Ergebnisse unterstützen unsere Hypothese, dass sich Unterschiede in den Lösungsraten auf die Salienz für die Lösung relevanter Aufgabenmerkmale zurückführen lassen. Insbesondere erscheinen Varianten, bei denen die Abweichung von der zu erwartenden relativen Häufigkeit maximal, das Verhältnis zwischen der großen und der kleinen Stichprobe groß sowie die Reihenfolge „zuerst große, dann kleine Stichprobe“ ist, geeignet, um Schülerinnen und Schülern die Problemlösung zu erleichtern. Intuitionen scheinen – wenn sie existieren – vor allem in diesem Fall wirksam zu werden. Aus instruktionaler Sicht bieten sich diese Varianten als Ausgangspunkt für eine mathematische Diskussion des Phänomens an, sowie als Übergang zu Aufgaben, bei denen für die Lösung relevante Merkmale weniger salient sind.

Literatur

Engel, J., & Sedlmeier, P. (2005). On middle-school students’ comprehension of randomness and chance variability in data. *ZDM*, *37*(3), 168–177.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*, 96–105.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*, *3*, 430–454.

Lem, S. (2015). The intuitiveness of the law of large numbers. *ZDM*, *47*(5), 783–792.

Lem, S., van Dooren, W., Gillard, E., & Verschaffel, L. (2011). Sample size neglect problems: A critical analysis. *Studia Psychologica*, *53*(2), 123–135.

Murray, J., Iding, M., Farris, H., & Revlin, R. (1987). Sample-size salience and statistical inference. *Bulletin of the Psychonomic Society*, *25*(5), 367–369.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137.

Rubel, L. H. (2009). Middle and high school students’ thinking about effects of sample size: An in and out of school perspective. In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings of the 31st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 636–643). Atlanta, GA: Georgia State University.

Schnell, S., & Prediger, S. (2012). From “everything changes” to “for high numbers, it changes just a bit” – Theoretical notions for a microanalysis of conceptual change processes in stochastic contexts. *ZDM*, *44*(7), 825–840.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making*, *10*, 33–51.

Watson, J. (2000). Intuition versus mathematics: The case of the Hospital Problem. In J. Bana & A. Chapman (Eds.), *Proceedings of the 23rd annual conference of the Mathematics Education Research Group of Australasia* (pp. 640–647). Sydney: MERGA.

Watson, J., & Callingham, R. (2013). Likelihood and sample size: The understandings of students and their teachers. *Journal of Mathematical Behavior*, *32*(3), 660–672.

Weixler, S., Sommerhoff, D., & Ufer, S. (2019). The empirical law of large numbers and the hospital problem: systematic investigation of the impact of multiple task and person characteristics. *Educational Studies in Mathematics*, *100*(1), 61–82.