

The impact of the replication crisis on philosophy: two case studies

A symposium proposal for the BSPS conference, Durham 2019.

Dr. Suilin Lavelle, School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

Dr. Richard Morey, School of Psychology, Cardiff University

Dr. Hugh Rabagliati, School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

Short summary

Contemporary philosophy of mind is closely integrated with the empirical sciences. Many of the most respected theories of vision, thought, consciousness and free will draw on findings from neuroscience and psychology. But what happens when the psychological findings upon which this work is based do not replicate? How are the philosophical theories affected?

This symposium brings together a philosopher and two psychologists who are involved in large-scale replication studies that question data central to key areas in philosophy of mind. All three presenters examine the role of replication in advancing science, through the lens of work that is of direct relevance to philosophers.

Dr. Richard Morey presents a large scale replication attempt of the 'Action Sentence Compatibility Effect', which is often cited in favour of embodied cognition; and Dr. Hugh Rabagliati will discuss on-going complications in infant cognition replications, which impact key philosophical theories about nativism and concept acquisition. Dr. Suilin Lavelle is a philosopher whose work is directly affected by the on-going replication crisis in infant cognition, and will discuss (a) how Dr. Morey's and Dr. Rabagliati's work impacts the related philosophical theories, and (b) how the replication debates are a living example of scientists arguing about what constitutes 'data', bringing forward aspects of theory-choice so famously developed by Kuhn (1962).

Description of the symposium

How we understand scientific replication and its role in advancing knowledge is a lively, cutting-edge issue in contemporary philosophy of science (Bird, 2018; Fidler & Wilcox, 2018; Leonelli, 2018). This is, in large part, due to the perceived crisis of failed replications across the sciences. While bad science bears some responsibility for this phenomenon, other factors are also at work. It may be the case that some sciences, e.g. social psychology, are more likely to be affected by the base-rate fallacy yielding false positives (Bird 2018), despite the rigour of the scientific method used to gain results. Some observations and findings do not lend themselves easily to repeatability (Leonelli 2018), while others are fragile, with apparent minor changes to method causing them to disappear, or become significantly weaker (Buttelmann, Baillargeon, & Southgate, 2018). That a phenomenon does not repeat does not entail that it does not exist; but it does warrant closer investigation to better understand why it has not been duplicated.

This symposium looks at the replication debates through the lens of case studies that are of central interest to philosophers. The aim is not to rehearse arguments about the value or otherwise of replication, but to examine how these issues play out in two applied cases. These cases directly impact contemporary philosophical theories and are accessible to philosophers and those in related disciplines who may not have encountered them before. The symposium will consist in three talks: one for each case study, highlighting philosophical questions specific to that field (Drs. Morey and Rabagliati); and a talk developing the themes of these specific replication cases to their impact on philosophy of science more generally (Dr. Lavelle).

Case study one: The Action-sentence Comprehension effect and Embodied Cognition

Broadly speaking, embodied accounts of cognition see the body as being partially constitutive of cognitive processes. Such accounts oppose traditional claims that cognition consists in the processing of mental representations, and that such processing is done entirely in the brain. Examples of embodied theories of cognition include Andy Clark's extended mind thesis (2011), Daniel Hutto's 'Radical enactivism' (Hutto & Myin, 2012), and Vittorio Gallese's 'Embodied simulation' (Gallese, 2009)

The *action-sentence compatibility effect* (ACE) (Glenberg and Kaschak 2002) is a phenomenon that has been cited in support of embodied cognition (e.g. Niedenthal et al 2005; Gallese and Lakoff 2005). Glenberg and Kaschak maintained that participants took longer to understand a sentence when their response to that sentence required an action incongruent with the sentence. For example, upon reading the sentence 'close the drawer', participants who had to move their arm forwards to push the response button (congruent movement) were faster to respond than those who had to move their arm backwards (incongruent movement) to push the button. Advocates of embodied cognition take this effect to evidence the necessary role of the motor system in language comprehension. While one may question this interpretation of the effect, the work presented by Dr. Morey suggests that the effect itself is not as stable as Glenberg and Kaschak imply.

Of course the embodied cognition movement does not rest on the success of one experiment. But Morey and colleagues' null replication through a large-scale replication paradigm serves as a critical case-study through which philosophers can explore and consider the type of data they need to support their theories. If embodied theories of cognition predict variability in the ACE then one might not expect the ACE to appear in all populations, and so a large-scale non-replication of the effect could be expected. Perhaps there are boundary conditions on the ACE that are revealed when the effect fails to replicate in another lab; subtle cues that prime participants to be more embodied in their cognition than usual. If this is so, then embodied cognition may not be a default mode of thought, but one that can be spontaneously activated if required. These possibilities each affect our understanding of embodied cognition, and philosophers who place weight on the ACE need to consider more closely how they expect it to be manifested, and then to look carefully at the results from different replication attempts to decide how to use the data.

Case study two: looking time studies and infant cognition

Nativist-leaning accounts of cognition attribute great significance to young infants' conceptual abilities. If a concept appears in early infancy, so the argument goes, the infant is unlikely to have had experiences of sufficient structure to allow her to learn about that concept, and so the concept must be innate (Carey, 2009;

Carruthers, 2013; Laurence & Margolis, 2001). Therefore, when psychologists find evidence of infants understanding concepts, this quickly becomes incorporated into nativist theories as further evidence for their views. One widely discussed example in the literature is Onishi and Baillargeon's (2005) claim that 15 month old infants have the concept BELIEF, and expect others to act in accordance to their beliefs, regardless of whether those beliefs are true or false. This contrasts with the long accepted view that children only begin to attribute false beliefs to others around their fourth birthday (Wimmer & Perner, 1983).

However, as Dr. Rabagliati will argue, infants are fickle experimental participants, and the methods used to ascertain their conceptual abilities are extremely sensitive and yield noisy data. As revealed in a special issue of *Cognitive Development* (2018), attempts at replicating infant data have been mixed, at best. Dr. Rabagliati's presentation highlights some of the methodological and statistical challenges of infant work, and the questions this raises for those philosophers whose work relies heavily on the early emergence of complex and abstract concepts. These themes are further developed in Dr. Lavelle's presentation, where she argues that competing theorists are, true to Kuhn's observations (1962), prioritising different aspects of the infant data in their theories. She also applies Leonelli's work on the epistemic value of replication to the specific case of developmental psychology.

References

- Bird, A. (2018). Understanding the Replication Crisis as a Base Rate Fallacy. *The British Journal for the Philosophy of Science*.
- Buttelmann, D., Baillargeon, R., & Southgate, V. (2018). Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development, 46*, 112-124.
- Carey, S. (2009). *The Origin of Concepts*. OUP.
- Carruthers, P. (2013). Mindreading in Infancy. *Mind and Language, 28*, 141 - 172.
- Clark, A. (2011). *Supersizing the Mind*. O.U.P.
- Fidler, F., & Wilcox, J. (2018). *Reproducibility of Scientific Results*. (E. Zalta, Editor) Retrieved from Stanford Encyclopedia of Philosophy (Winter 2018 Edition): <https://plato.stanford.edu/archives/win2018/entries/scientific-reproducibility>
- Gallese, V. (2009) Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanalytic Dialogues, 19*, 519-536.
- Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 5*, 455-479.
- Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic bulletin & review, 9*, 558-565.
- Hutto, D., & Myin, E. (2012). *Radicalizing Enactivism*. MIT Press.
- Laurence, S., & Margolis, E. (2001). The Poverty of the Stimulus Argument. *British Journal for the Philosophy of Science, 52*, 217 - 276.
- Leonelli, S. (2018). Rethinking Reproducibility as a Criterion for Research Quality. In L. Fiorito, S. Scheall, & C. Suprinyak (Eds.), *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise*. (pp. 129-146). Emerald Publishing Ltd.
- Niedenthal, P., Barsalou, L., Winkelman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and social psychology review, 9*, 184-211.
- Onishi, K., & Baillargeon, R. (2005). Do 15 month-old infants understand false beliefs? *Science, 308*, 255-58.

Abstracts

A large-scale, multi-lab test of the Action-Sentence Compatibility Effect: Results and implications

Dr. Richard Morey

The Action-Sentence Compatibility Effect (ACE; Glenberg and Kaschak, 2002) is a speeding of response times to evaluate sentences when the actions described in those sentences are congruent with the actions necessary for a response. For instance, if a participant must evaluate the meaningfulness of the sentence “You handed Meghan the book,” and the correct response is away from the participant’s body, the described movement (handing to) is compatible with the response. The ACE effect is presumed to be due to strong links between cognitive systems for understanding language about motor actions and motor systems themselves. Although the ACE has been described in dozens papers, Papesh (2015) recently reported a number of failures to replicate the effect concluding that the effect may not be as robust as previously believed. A large-scale, multilab attempt to replicate the effect designed by the proponents of ACE did not yield evidence for the effect. I will discuss what the null replication teaches us about the value of large-scale replications for scientific discovery, for how we should view “solutions” for the replication crisis.

Interpreting research on infant cognition in light of the replication crisis

Dr. Hugh Rabagliati

Experiments on the cognitive and social abilities of human infants have provided some of the most vivid illustrations of how psychological experiments can contribute toward answering key questions in philosophy of mind. These experiments, for example, have suggested that infants possess innate knowledge (e.g., so-called “core knowledge” of physical objects, Spelke, 1990), have provided evidence for symbolic mental representations (Marcus et al., 1999), and suggested precocious skills at learning to represent the beliefs of others (Onishi & Baillargeon, 2004).

But now, against a background of tumult in the social and biological sciences, the collective evidence provided by studies of infant cognition is being re-evaluated. Recent work in disciplines such as medicine (Ioannidis, 2005), psychology (Open Science Foundation, 2015), and economics (Camerer et al., 2018), have suggested limits to the global scientific record: Statistical evaluations suggest that many (perhaps most) research findings may be overstated, direct replications of prior work frequently fail to duplicate the previously-found results, and it has become painfully clear that, in many fields, incentives such as publication bias had caused unintentional deviations from scientific best practices.

This paper will provide an overview of how these issues have impacted, and will continue to impact, research into infant cognition. Part 1 describes the challenges of measuring cognition in infancy, with reference to the dominant experimental paradigm, i.e., the analysis of infant looking times. It will discuss how looking times (e.g., to novel, unfamiliar, or unexpected stimuli) are typically interpreted, and the mechanics of collecting looking time data, focusing on aspects of data collection that may not be obvious to those looking in from outside the field.

Part 2 describes reasons for caution when evaluating infant cognition research. These include concerns about publication bias, about researcher degrees of freedom (e.g., how researchers make decisions when processing data, such as excluding participants), and about so-called HARKing

(hypothesising after the results are known) in which researchers, when writing up the results of an experiment, retrofit their supposed hypotheses to dress up post-hoc interpretations as predictions.

Part 3 describes solutions to these problems. Solutions include 1) Large-scale, multi-lab replication studies, in which groups of laboratories work together to standardize an experimental protocol and collect far larger datasets than ever before; 2) Meta-analytic aggregation of prior work, which allow scientists to draw firmer conclusions by combining the results of many different published studies, and adjusting for publication bias; 3) Improving the psychometric validity of experimental paradigms, which is to say, showing that the paradigm measures what it was intended to measure.

Thus, this paper hopes to provide a primer allowing non-experts to give a more informed, nuanced and critical evaluation of research into infant cognition.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Altmejd, A. (2018). *Nature Human Behaviour*, 2(9), 637.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). *Science*, 283(5398), 77-80.

Onishi, K. H., & Baillargeon, R. (2005) *Science*, 308(5719), 255-258.

Open Science Collaboration. (2015). *Science*, 349(6251), aac4716.

Spelke, E. S. (1990). *Cognitive science*, 14(1), 29-56.

Replication, variation and theory choice

Dr. Suilin Lavelle

Many philosophers draw on work from the sciences to support their theories. In the philosophy of mind alone, findings in neuroscience, developmental psychology, cognitive ethology and anthropology inform respected theories of cognition and consciousness. The apparent crisis facing the psychological sciences as well-cited findings to replicate should therefore be of paramount concern to those philosophers who rely on them in developing their positions. This paper expands on themes presented by my colleagues working in embodied cognition and infant cognition, to better understand how philosophers whose work is directly affected by replication failure should react to the crisis.

The complexities of working with infants, and the messy data emerging from large-scale replication attempts of infant work, provides rich material for philosophers of science. From the perspective of replication there are important questions about the epistemic value of failed replications: while this has been discussed in science more generally (Leonelli 2018), these insights have yet to be applied to developmental psychology. There are critical debates to be faced about what should be considered 'data': the infants whose looking time suggests they can attribute beliefs to others, or those infants who do not look in the 'right' way? This question has a direct impact on philosophical theories. Those of a nativist leaning tend to focus on infants' (and childrens') successes on false belief tasks, taking it as evidence of an innate BELIEF concept. Those who prefer empiricist, learning theories, find childrens' errors to be key data, taking it to evidence a child trying out new hypotheses, sometimes successfully and sometimes not. These questions, about what constitutes data and what the core explanandum of theories of cognition should be, contribute to on-going, live examples of theory choice, and reflect core questions in the philosophy of science.

The second theme raised by these talks is that of variation. Advocates of embodied cognition cite the ACE as evidence for their view, but as Dr. Morey argues, the ACE may not be a stable effect. While one failed replication does not cast doubt on the entire embodied cognition project, there are important questions to be considered about the reasons for the failed replication. The most pressing of these concerns variation in embodied cognition. Perhaps the failed replication can be understood in terms of variation across individuals' propensity to engage in embodied thought processes. Or perhaps the variation is within individuals, with subtle cues priming one towards or against embodied thought processes. If variation is the key, then advocates of embodied cognition need to accommodate this within their theories. Alternatively, if the ACE simply does not exist at all, then this also needs to be accounted for.

Participant biographies

Dr. Suilin Lavelle

Dr. Suilin Lavelle is a lecturer in philosophy of mind and cognitive science in the school of Philosophy, Psychology and Language Sciences at the University of Edinburgh. Her main area of expertise is Social Cognition, with a special interest in how we grasp of mental state concepts. Her recent book 'The Social Mind', published by Routledge, has been highly praised for its integration of psychology and philosophy, with one reviewer describing it as 'written by a philosopher who has engaged deeply with empirical research, this book shows why the social mind is such an exciting interdisciplinary topic'. In the book, she argues that the explanandum of Social Cognition has been poorly characterised by anglo-american philosophy as 'the attribution of psychological states to other people'. In fact, attributing psychological states to others is not central to many of our social interactions, and this characterisation in turn may be an artefact of a particular way of thinking culturally specific to North American/Northern European cultures. However, the ability to attribute mental states to others (and ourselves) is universal, albeit not as commonly used as philosophy textbooks maintain, and she supports a broadly theory-based account of this ability. She supports these positions with reference to work in developmental psychology, anthropology and cognitive ethology.

Due to her reliance on empirical work for her philosophy, Dr. Lavelle has developed a keen interest in debates about replication in psychology. She is working on a paper discussing the mixed results in replicating infant responses in spontaneous response false belief tasks, and the impact of this work on contemporary philosophical accounts of infant cognition. She also argues this debate is a living example of scientists arguing over what constitutes 'data', thus reflecting aspects of theory choice so famously developed by Kuhn (1962).

Dr. Lavelle teaches a wide range of courses in philosophy of mind and philosophy of science at both under-graduate and post-graduate level. She has three PhD students working on projects within Social Cognition, including a student co-supervised with the linguistics department.

Selected publications

The Social Mind (2018) Routledge.

'Theory of mind across cultures.' (2016) *Routledge Handbook of Social Mind*. (Ed. J. Kiverstein).
Routledge

'Is a modular cognitive architecture compatible with the direct perception of mental states?' (2015)
Consciousness and Cognition

'Contrastive Explanation and the Many Absences Problem' (first author; co-authored with George Botterill and Suzanne Lock). (2013) *Synthese*, 190, 3495-3510.

'Theory-theory and the direct perception of mental states'. (2012) *Review of Philosophy and Psychology*, 3, 213 – 230.

'Two challenges to Hutto's Enactive account of pre-linguistic social cognition.' (2011) *Philosophia*, 40, 459 - 472

Dr. Richard Morey

Dr. Richard Morey is a Reader in the School of Psychology at Cardiff University. His work is at the intersection of cognitive psychology, statistics, and philosophy of statistics, with his recent work addressing problems in statistical inference and philosophy of science. In 2017, he was awarded the Psychonomic Society Early Career Impact award by the Federation of Associations in Behavioral and Brain Sciences (FABBS) for his contributions to statistical reform and education. He is best known for his and his colleagues development of Bayesian methods statistical analysis, including the BayesFactor software which implements a number common analysis techniques. The methods underlying the BayesFactor software have been implemented in a number of other statistical software packages, including IBM SPSS.

His work in the philosophy of science and statistics argues for the importance of Bayesian statistical approaches, hypothesis testing in general, and of a proper understanding of classical approaches. He has also added to debates on meta-statistical methods for detection of publication bias and questionable research practices.

A major part of Dr. Morey's work is statistical education and outreach, including workshops around the world on Bayesian inference. In the past several years he has served as a statistical advisor for the journal Psychological Science, an associate editor at the Journal of Experimental Psychology: General, and a digital associate editor for the Psychonomic Society.