



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Less theory, more observation**

A response to psychology's 'theory crisis'

**Citation for published version:**

Lavelle, JS 2024, 'Less theory, more observation: A response to psychology's 'theory crisis'', *Philosophy*.

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Philosophy

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## Abstract

There is a worry within psychology that its researchers experience too many degrees of freedom in formulating their hypotheses, resulting in experiments being designed to test implausible hypotheses which then do not successfully replicate. A popular diagnosis of this problem is that psychological theories are too vaguely specified, and that formalising them will add the constraints necessary to solve the problem. This paper argues for a different strategy, namely, for more theory-lite observational research to be conducted. This appears antithetical to the restraint urged by others, but I argue that it is a necessary pre-cursor to forming well-established foundational theories. I discuss two case-studies to support my arguments.

**Keywords:** *Replication, Theory-crisis, Developmental Psychology, Exploratory Research, Infant Observation, Methodology, Experiment, Underdetermination*

# Less Theory, More Observation: A Response to Psychology's 'Theory Crisis'

## 1 Introduction

It is no secret that many experiments in psychology are difficult to replicate, leading to what in recent years has become known as the 'Replication crisis' (Lilienfeld and Waldman 2017; McNutt 2014; Open Science Collaboration 2015). The replication crisis has caused a great deal of soul-searching among psychologists, but largely appears to have passed philosophers by. This lack of interest from philosophy is odd, especially in the light of how many philosophers of mind rely on empirical work to support their theories. Perhaps one reason for this comparative neglect is the thought that the crisis can be fixed with practical measures, such as pre-registration, open data, and pledges from journals to publish replications and null results. Designing and implementing such measures is not within the philosopher's skill-set.

However, relegating the remedial work entirely to practical measures would be a mistake, for there are also deep philosophical issues at the heart of the replication crisis concerning, for example, methodology, experiment, the relationships between data, phenomena and theories, and many others.<sup>1</sup> This paper examines one of these, namely, the recent diagnosis that the ‘replication crisis’ is actually a symptom of a deeper problem known as the ‘theory crisis’. The theory crisis claims that many theories in psychology are so vaguely specified that they cannot yield workable hypotheses. A solution which has gained recent popularity is that the tools of formalisation and computational modelling will add the required specificity, and thus solve the crisis. This paper argues that this solution is inadequate, and that a more useful response is to encourage research that aims to describe aspects of human cognition in naturalistic settings. I call this ‘observational’ research. I argue that observational research is necessary to ground better-specified foundational theories, and that such research is currently lacking in most areas of psychology. Section 2 describes the theory crisis as it is currently perceived, and the reasons why some researchers believe formalisation will resolve it. Section 3 presents my arguments for why the solution will not work, and introduces observational

---

<sup>1</sup>Philosophers who have fruitfully explored some of these themes include Marcus Eronen (2020, Eronen and Bringman 2021), Uljana Feest (2016, 2019, 2022), Samuel Fletcher (2021a,b), Sabina Leonelli (2018), and Edouard Machery (2020).

research. In section 4 I describe two case-studies which I think are paradigmatic of observational research, with section 5 using these to reinforce the arguments presented in section 3. By the paper's conclusion I hope to have persuaded philosophers who use empirical work from psychology to be more reflective about the nature of the data they draw upon, and to have shown that there are rich pickings within the replication debates for philosophers of science to engage with. If the problems discussed here inspire philosophers to re-examine the methodologies central to psychological experiment, then a good crisis will not have been wasted.

## **2 The theory-crisis**

### **2.1 What is the theory crisis?**

One analysis of the replication issues currently experienced by the psychological sciences is that they stem from a 'theory crisis', which refers to a specific problem of underdetermination. The problem of underdetermination most familiar from philosophy of science is the underdetermination of a theory by the data. This occurs when more than one theory is able to explain a particular data set, as when, for example, both behaviourist and mentalistic approaches claim to be able to account for primates' performance on social cognitive tasks (Fitzpatrick 2009).

Those arguing for a theory crisis focus on a different type of underdetermination, namely, when hypotheses can be accommodated by

more than one over-arching theory (Muthukrishna and Henrich 2019; Oberauer and Lewandowsky 2019). For example, the hypothesis that our ability to accurately attribute psychological states to other people is heightened when we feel disempowered (Rizzo and Killen 2018) can be a consequence of either the theory-theory or simulationist approach to mindreading. The following argument shows how the theory crisis is thought to cascade through the scientific method:

1. The psychological sciences lack detailed, widely accepted foundational theories.
2. 1. results in researchers having an unacceptably large degree of freedom in the process of forming hypotheses.
3. Without the constraints provided by well-established, detailed theories, researchers are free to formulate and test hypotheses that
  - (a) have a low prior likelihood of being true, and
  - (b) are compatible with more than one over-arching theoretical framework (Theoretical underdetermination).
4. Due to the significance level of psychological experiments ( $p = 0.05$ ) and publication biases, many of these hypotheses end up as published findings supported by data, when the results are in fact false positives (Asendorpf et al. 2013; Ioannidis 2005; Scheel et al. 2021) .

5. Experiments used to support these hypotheses often fail to replicate, constituting psychology’s ‘replication crisis’ (Bird 2021).

It is not the aim of this paper to defend this argument, and each of the premises is open to contention. Instead, I want to evaluate a specific response to it that targets premise 2, namely, a proposal on how to constrain researchers and thus break the move to 3.

## **2.2 A proposed solution to the ‘Theory crisis’**

As presented in the preceding argument, those arguing that psychology faces a theory crisis see the problem as stemming from researchers having too much freedom in developing their hypotheses; a lack of constraint that is consequent upon the field having relatively few well-established foundational theories. A solution to the problem popular among cognitive scientists is that formal theory and computational models will provide the desired constraints (Borsboom et al. 2021, Fried 2020, Robinaugh et al. 2021, Scheel et al. 2021). Michael Muthukrishna and Joseph Henrich write

‘A general theory of human behaviour would be evolutionarily plausible (via natural selection under phylogenetic constraints), often utilize formal models, and provide us with an ultimate framework that delivers proximate predictions. In addition to being a crucial part of the abductive scientific process, by forcing, often formally, statements of assumptions and logic,

constrained by the broader web of interconnected work, such a framework also contributes to one of the goals of the replication renaissance: constraining researchers' (Muthukrishna and Henrich 2019, p. 223).

Formal theories are said to serve the function of constraining researchers in two, related, ways. The first is by avoiding the vagaries that come with natural language:

‘Psychologists predominantly use verbal statements to express their theories, hypotheses, predictions and inferences. Because natural language is imprecise, this practice keeps causing confusion. Expressing theoretical assumptions and hypotheses in formal mathematical, computational or causal models can help reveal ambiguous definitions, hidden assumptions and internal inconsistencies’ (Scheel 2022, p. 2; see also van Rooij and Baggio 2021, pp. 685 - 6).

Here the idea is that describing our hypotheses in natural language leads to terms that are too under-specified to be tested, with the consequence that too many data could then fit them. For example, György Gergely and Gergely Csibra proposed the hypothesis that infants expect others to act in the most efficient way to achieve their goals (Csibra et al. 1999; Gergely and Csibra 2003). Terms such as ‘goal’, ‘agency’ and ‘efficiency’ and measures of efficiency could be interpreted in multiple different ways, and one way to



reduce the scope of such interpretations would be to present them formally; in the words of Iris van Rooij and Giosuè Baggio: ‘utilities are numbers, beliefs are propositions, and meanings of linguistic terms can be formalised in terms of functions and arguments: numbers, propositions, functions and arguments are all well-defined mathematical concepts’ (2021, p. 686).

The second type of constraint comes from our prior knowledge about the organisms we are working with. One such constraint is evolutionary plausibility, as previously mentioned by Muthukrishnan and Henrich. Those authors argue that computational models can help researchers understand whether one’s theory about a particular cognitive capacity, e.g. the ability to recognise goal-directed movements, is evolutionarily plausible, by modelling various gene-culture interactions over thousands of generations and seeing if such a capacity emerges. If it does not, then one goes back to the formalisation stage and adjusts the key concepts until something which *is* evolutionarily plausible emerges. Thus advocates of such computer modelling techniques also support the formalisation of theories, as formalisation is required for the models to work. Another constraint, also discussed by van Rooij and Baggio (2021), is ‘tractability’. The phenomenon of joint attention illustrates this perfectly. Joint attention occurs when two or more people are attending to something, and each is aware both of the other’s attention to the object, *and* of the other’s awareness that the first person is aware that they are aware of the object. Attempts to characterise this phenomenon in terms of propositional

knowledge run into an intractable problem equivalent to the ‘problem of co-ordinated attack’ (Wilby 2010). Thus our ability to jointly attend to features of the world with others cannot be represented with propositions because that would result in something impossible to implement.

The view that psychology is facing a theory crisis stems from the worry that researchers have too few constraints on their theories of cognitive capacities. Formalisation is proposed as a solution because it imposes constraints: the concepts used in theories must be specifiable in formal ways so as to avoid ambiguities and, once formalised, such theories are then candidates for further stricture through computational models of evolutionary plausibility, tractability and others (e.g. learnability, physical realisation). These are definitely benefits that can be reaped from the formalisation approach. But in the rest of this paper I want to explore an alternative solution to the theory crisis, which at first glance looks to pull in the opposite direction by advocating for *less* theory, not more.<sup>2</sup> Despite this, I believe the two approaches are complementary, and will explain why in section 5.

---

<sup>2</sup>Other solutions that reject formalisation have also been proposed, focussing on different aspects of underdetermination; see Oude Maatman (in prep).

## **3 Theory and observation**

### **3.1 Why are there so few established theories in psychology?**

The formalisation solution to the theory crisis presented in 2.2 certainly has its merits. My worry is that it does not reach down to the root of the problem. As already explained, it targets premise two by offering ways to constrain researchers in the absence of well-established theories. But what of premise one? Is it the case that there are few well-established theories in psychology and, if so, then why?

In contrast to many scientific disciplines, such as physics, chemistry, astronomy and biology, psychology as we know it today is a relatively young science. Modern psychology's emphasis on experiments, operationalisation of concepts and hypothesis testing really came to the fore in the wake of Popper's critiques of psycho-analysis, in a move to distinguish itself as a falsifiable (and thus respectable) science (although its experimental heritage more broadly is of course much older, going at least back to Wilhelm Wundt). The youth of the discipline naturally brings some distinctive features. First, unlike its older cousins, psychologists have fewer catalogues of systematic observations of human behaviours across the globe. Astronomy serves as a helpful contrast: we have documentation going back thousands of years describing the night skies, detailing observations that are systematic in their descriptions of what is seen, where and at what time

of the year, and comparisons with previous observations. Psychology does not have this heritage. Observations can be gleaned from historical records, but these were not created with the primary aim of documenting human behaviour in terms familiar to modern psychologists. In the last few decades linguistic corpora of recordings of natural interactions between people, or children learning their native language have been created, but the number of these pale in comparison with the centuries of data that other sciences have to work with. Second, and as a direct consequence of this, there are fewer impregnable theories in psychology than in other disciplines. To borrow an example from Lavelle (2022): if a child drops a mento sweet into a bottle of fizzy pop and the sweet simply sinks, the immediate reaction is not that they have discovered some hitherto unknown principle of chemical reactions. Rather, one is more likely to believe that something has gone wrong with the experiment, perhaps that the drink was not sufficiently carbonated. The connection between these two points is that psychology has few well-established phenomena that serve as core explananda for its theories, whereas the older sciences have lots. Fewer does not mean zero: examples of well-established phenomena include the limits of short-term memory (Baddeley 2007), babies' preference of infant-directed-speech (Many Babies 2020), and the visual processing of spatial relations (Dror and Schreiner 1998). Rather, the point is one of scale: in contrast to other sciences the number of well-established phenomena is small. And because the older sciences have experienced a higher degree of agreement for many

years about which phenomena exist and demand explanation, there has also been more time to develop well-established theories that explain them (Oberauer and Lewandowsky 2019, Rozin 2001).

### **3.2 More observation**

My preferred solution to the theory crisis is that psychology needs a greater number of well-established phenomena, and that this step is prerequisite to having well-established theories, formalised or otherwise. These phenomena would come in the form of theory-lite observations of people in their everyday environments. The emphasis is on observation with minimal intervention, in direct contrast to effects that are uncovered through experiments conducted in laboratories. The importance of such work has been discussed in the literature (Borsboom et al. 2021; Haig 2013), although not in direct connection with the current theory-crisis. One exception is an excellent paper by Marcus Eronen and Laura Bringman who write that ‘by discovering new phenomena and gathering more robust evidence for those already discovered, the possible space of theories will be constrained’ (Eronen and Bringmann 2021, p. 784). I agree whole-heartedly, and in the following sections develop my own analysis of what such data consists in and how it serves this constraining function.

My notion of a ‘theory-lite’ observation is intended to respect the view that observations can never be theory-free, for reasons well-rehearsed in the philosophy of science (Hanson 1958; Van Fraassen 1980). ‘Theory-lite’ is

meant to capture the idea that the observations are not collected with the aim of supporting a particular theory. Instead, it is a first pass at describing what people are doing in different situations. From this, one can begin to see if particular patterns emerge which warrant further investigation. If established, it is these patterns which form the basis of a set of explananda that a wide range of researchers, regardless of their theoretical commitments, can agree to be core to their discipline.<sup>3</sup> Agreement may be hard to find, but this feature should be openly acknowledged and embraced because it respects the idea that different researchers' backgrounds (culturally and theoretically) affects what they perceive to be a pattern in human behaviour. This is exactly what occurs in qualitative research, where differences in how researchers code data is brought into the open as a topic of discussion rather than being hidden away. This point is developed further in section 4 where the case studies illustrate different levels of researcher agreement about the patterns revealed in the data.

Other researchers have made similar calls for more observational work in the psychological sciences, albeit in contexts different to the theory crisis discussed here. Alan Kingstone discusses it with respect to the problem of

---

<sup>3</sup>I agree with Thomas Kuhn (1962) and more recently Hasok Chang (2012), that different groups of researchers will prioritise different phenomena as *the* explanandum most central to their project. But while, say, Priestly and Lavoisier disagreed about the importance of explaining the circulation of nutrients in the environment, both agreed that the phenomenon existed.

generalising psychological effects uncovered in the lab to the real world, with a call to ‘first directly study how people behave in their natural real-world environments before moving into the lab’ (Kingstone et al. 2008, p. 320). His view is that psychologists have focussed too narrowly on control and invariance in their lab-based experiments, which can cause effects which are actually artefacts of very specific laboratory situations to be inappropriately characterised in too general a way,<sup>4</sup> and have not paid enough attention to how situation and context moderate cognition. The only way we can better understand these influences, he claims, is to ‘spend a good deal of time observing and describing what other people are doing’ (ibid).

Another way in which observational work has been emphasised is through the distinction between hypothesis-generating and hypothesis-testing work (which is discussed in more detail in section 5). Hypothesis-testing work occurs when researchers design an experiment to test a specific hypothesis that they have formulated. Hypothesis-generation research occurs when one looks for new patterns of interest in a data set, and use these to shape their thoughts about possible new hypotheses to

---

<sup>4</sup>See Feest (2016) for an analysis of the difference between the general claim that ‘listening to Mozart improves your IQ’, and the more specific one that ‘listening to up-beat classical music can improve performance on a very specific measure of intelligence’.)

test.<sup>5</sup> Observational work is a species of hypothesis-generating research, as it consists in collecting data that serves as the touchstone for new theories. This characterisation of observational work by reference to the type of data involved means that it can include lab-based work, as it may require equipment or a specific set-up. For example, in one of the studies described below, toddlers needed to interact with brightly coloured objects in a lab that had a white background, so that the analysis of how much of their visual field was filled by a particular object could be tracked by a computer. Despite this, the data was still collected with a descriptive aim: researchers wanted to know how toddlers interacted with objects, so that they could use these data to ground theories about the cognitive mechanisms guiding these interactions.

Research that has been gathered within a hypothesis-generation framework cannot be used for hypothesis-testing. When researchers analyse data gathered with an exploratory aim, they do not know what patterns they are going to find. When they find a pattern of potential interest, they

---

<sup>5</sup>Other terms used to capture the spirit of hypothesis-generating research include ‘discovery-oriented research’ (Oberauer and Lewandowsky 2019), ‘exploratory research’ (Burian 1997; Feest and Steinle 2016; Steinle 1997), and ‘non-confirmatory research’ (Scheel 2022). These terms are by no means synonymous as they highlight different aspects of non-theory-testing research. A more careful analysis of the differences is beyond the scope of this paper, but can be found in Feest and Steinle (2016).



cannot point to the data they are analysing as support for the existence of that pattern. The pattern's existence needs to be corroborated independently of the exploratory data, which is when hypothesis-testing experimental work comes in. To use exploratory research in a confirmatory way would be a problematic form of HARK-ing (hypothesising after the results are known (Kerr 1998)) that is, implying that the theory was posited prior to data collection (Wagenmakers et al 2018; Wicherts et al 2016).<sup>6</sup>

Addressing the root of the theory crisis means understanding why there are so few constraints on researchers, and finding ways to resolve this. Formalisation offers one kind of constraint. But this section has argued that focusing on formalisation is premature. Before we are even in a position to create theories that can be formalised, we need a clearer account of those theories' explananda. Observational work is one way of gathering the necessary explananda. The next section explores what this work might look like for a particularly tricky area of psychology, namely, developmental psychology.

## 4 Two case studies

In the previous section I argued that what lies at the root of psychology's theory-crisis is not a dearth of rigorous theorising, but rather a paucity of systematic observations to ground phenomena that researchers widely agree

---

<sup>6</sup>HARK-ing can be useful in some research contexts, see Baruch (2023).

to be worth explaining. In this section I present two types of work which I think are promising in helping rectify this lacuna: head-mounted video cameras on infants (see also Scheel 2022), and the field of ‘Infant Observation’. I have chosen to focus on work with babies for several reasons. First, infant cognition, as a field of study, is particularly susceptible to the problems outlined in section 3.1, being an especially young discipline within psychology. Many of the experimental techniques designed to get data from these non-verbal, un-instructable participants are barely thirty years old at best. Second, infancy research has been affected by the replication crisis more generally, with researchers struggling to replicate landmark findings.<sup>7</sup> Third, while there is a lack of observational work in psychology in general, I think the problem is especially acute in infancy work. Many of the techniques associated with observational work, such as qualitative interviews, are just not appropriate for babies. Even within those fields which specialise in detailed observations of people — I am thinking here of social and cultural anthropology — children, and particularly babies, are usually studied only in virtue of their relationship to adults (Allerton 2016). Finally, developmental research has had a significant impact on contemporary philosophy of mind, especially in debates about nativism, core cognition, social cognition, to name but a few. Gaining a better understanding of how the problems with psychological data outlined in previous sections could be addressed should be of interest

---

<sup>7</sup>For examples see the ‘Many Babies’ projects: [www.manybabies.org](http://www.manybabies.org)

to philosophers who rely on them for their theories.

## 4.1 Head-mounted cameras

Since the early 2010's technology has allowed for small, light video cameras to be embedded into headbands that can be worn by young infants and toddlers. The recorded footage gives unique insights into what an infant sees during her daily life, from a perspective alien to most adults.<sup>8</sup> For instance, Fausey and colleagues (2016) discovered that for very young infants, other people's faces dominate their visual field, but that as the infant becomes more independent and mobile, this fades and hands become the dominant feature that they look at in social interactions. With hindsight this makes perfect sense: an infant's visual field will be constrained by her physical posture and limitations. But these data reveal just how striking this feature of infant visual experiences is, and serves as a starting point for new theories about so much of infant cognition, for example what sorts of social information is available to them at different stages of their development.<sup>9</sup>

Another example of hypotheses emerging from this observational data

---

<sup>8</sup>For an explanation of why researchers believe the video footage is a good proxy for the infant's visual field see Smith and colleagues (2015).

<sup>9</sup>The transition to looking at hands correlates with new types of play around shared objects, in contrast to face-to-face play that dominates earlier interactions (Fausey et al. 2016, p. 105)

comes from the same research team, this time looking at the difference between an adult and toddler's visual field as they play with shared toys. In this lab-based experiment, both parent and child wore the cameras and sat opposite each other at a toddler-height table, which had three toys on it. The parent and child interacted with the toys. Analysis of the toddler footage revealed that when the toddler was playing with the toys, she tended to only have one toy dominant in her visual field (lifting the toy, or orienting herself to facilitate this):

‘The toddler view is one in which, at any one moment, one toy is much larger than the other toys in the image and the largest object in the image changes often. In contrast, the parent view is broad, stably containing all three objects, with each taking up a fairly constant and small portion of the head camera field. [...] In sum, the adult view includes and is equal distance from all of the objects on the table top; but in marked contrast, the child's view often contains one dominating object that is closer to the head and eye and thus often blocks the view of the other objects’ (Smith, Yu, and Pereira 2011, p. 12).

Once again it becomes clear how different a toddler's view of the world is from an adults', with the data revealing patterns which had not occurred to adult researchers but which make sense once understood from the perspective of the different bodies and cognitive processing of toddlers.

Smith and colleagues used this pattern of behaviour (of moving the toy so it dominates the visual field) to ground a hypothesis about how children learn about objects, suggesting that the point at which the object was steady and dominant in the toddler's visual field was the optimal time to learn its name. They tested this hypothesis in later work, arguing that the data did indeed support it (Yu and Smith 2012). This is a clear example of initial hypothesis-generating research preceding hypothesis-testing.

I consider these studies to be paradigmatic of the observational work discussed in section 3.2. Data was gathered with the aim of finding out about infants' visual experiences, and the analysis uncovered unexpected and interesting phenomena worthy of further investigation. Although the data were gathered in a largely unstructured way (especially in the cases where infants wore the cameras during their daily life), the analyses were quantitative in nature. Fausey's data was coded by four people, who noted when hands or faces (or both) were present in the footage. There was very high inter-coder reliability (over 90% in all categories), indicating a strong level of agreement about what the data represented. In Smith's study, algorithms determined how much the toys dominated the participant's visual fields, and two independent human coders checking the algorithms outputs each agreed 100% with its coding.

Smith's work demonstrates 'theory-lite' research, as the researchers began with an open-ended question they believed to be of theoretical interest (e.g. 'how toddlers' own actions may play a role in selecting visual

information'), which implies that they already thought there would likely be some phenomenon of interest to explore, even though the parameters are left deliberately vague. The research questions and algorithms systematise the data in ways that result in certain patterns becoming especially salient. Hence they constitute systematic observational data of the sort that can ground further theoretical development.

## **4.2 Infant observation**

The second source of observational data I'd like to discuss comes from the practice of 'Infant Observation', pioneered by Esther Bick (1964). Bick was an analyst at London's Tavistock clinic and incorporated Infant Observation into the training for first-year psychoanalysis students. Each student would spend an hour a week with a family of a new-born infant for at least a year. During this time they would observe the baby's life: her interactions with others, her movements, eye-gaze, vocalisations, and note these down during and after their visit. This method of closely documenting a child, or group of children's interactions, has since expanded beyond infancy and the clinical setting, and has been used by researchers in anthropology, educational psychology, sociology and other disciplines to address questions such as 'What are the boundaries of the curriculum that young children experience and enact within the early childhood education setting?' (Stephenson 2011, p. 136); how to 'facilitate the voices of babies and young children' in the creation of government policy (McFadyen &

Anderson 2022, p.104) ; or how to understand children’s aspirations (again with a view to informing government policy) (Sim 2016). Here is an excerpt from Alison Stephenson’s piece ‘Taking a “Generous” Approach in Research with Young Children’:

‘It was only over time that I recognised how attuned Anakin (one year) was to the older boys, observing them from a distance, and increasingly experimenting with the roles they took. Outside in the playground he heard one of the boys chanting “Da, di de da da”, and a second later he said “da da”.

I wrote:

*He is clearly turned in to them - ignoring the two girls . . . in his vicinity. Anakin comes up and sits beside me on the seat. Even here he shouts “da, da”.*

Over time, I saw how he and Emjay (3 months younger) increasingly gained confidence in displays of shared resistance to teachers that older boys delighted in. When James (four years) was sent from mat-time to sit on a chair, he put a tissue box on his foot like a shoe:

*Towards the end of mat-time, Anakin and Emjay got up and wandered over there. A teacher called their names—she was holding [a baby]—but they did not show any sign of responding. They wandered on past James . . . Anakin picked out a*

*cardboard box, similar in size to James's.*

While the long period of observing enabled me to recognise such changing patterns, I was always aware of how many more I must be missing' (2011, p. 151).

The difference between this kind of descriptive report and reports of experimental effects based in laboratories is striking. The observer is trying to capture what she perceives as neutrally as she can. Naturally there are features that shape her attention: she focuses on the children's vocalisations, socialisation attempts and interactions with peers and teachers. Early students of Infant Observation, trained by Bick, were encouraged by her to attend to those interactions central to the psychoanalytic tradition:

'As I describe the initial visit, Mrs. Bick asks questions which on subsequent visits act like a zoom lens of a camera to move the baby into very close, clear focus. Her questions are: "How is mother holding the baby? Where is his head? How close to mother's body is he? Where is he looking? And what are his hands and legs doing when she changes position? What kind of movement or stillness do you see in the baby's body? Show us, we want to know." Through her questions Mrs. Bick elicits more detailed descriptions of the quality of mother's holding of the baby as well as additional comments on the various ways



baby “holds himself together”’ (Magagna 2018, p. 33).

A researcher’s theories and prior concepts will affect her data collection, even in a process intended to be less theory-heavy than experiments. Bick’s students are training to be psychoanalysts, and as such she guides their focus towards movements of the baby relevant to psychoanalytic concepts such as attachment and containment. But Infant Observation need not be restricted, as a method, to those with a commitment to psychoanalysis.

Firmly based in the qualitative tradition of research, there are important differences between how these data are gathered and analysed in contrast to the laboratory setting. First, the time-scale is very different. Whereas a baby might only be in a lab for half an hour, observations span several months, or even years.<sup>10</sup> Observers have time to get to know their participants, their individual preferences and character traits, family structure, cultural expectations, etc. These features are often deliberately omitted from laboratory experiments unless the experimenter thinks they are a relevant variable, and even then they would be reported as a response to a pre-test questionnaire. By contrast, Infant Observation details how these features are lived out in the participant’s daily experiences, how and when they change, etc. Similarly, while a lab-environment is totally novel to most children, observations take place in familiar settings such as the infant’s home, nursery, or routine outings to the shops or park. Again, the

---

<sup>10</sup>To give a sense of this scale, ten weeks is considered a ‘an extremely short observation’ (Klauber 2022, p. 84).

aim is to capture the child's lived experience. Instead of examining her social competence in an interaction with a puppet in a lab, the observer sees how the toddler interacts with siblings, parents, shop assistants and strangers.

Second, researchers may start out with a research question, but not one as fully formed as a hypothesis tested in a lab. Similar to Smith's work with infant head-cameras, the initial question is framed descriptively, e.g. as 'exploring the dynamics of relationship formation of three infants as they transitioned into a child-care nursery' (Degotardi 2011, p. 18), with the goal of collecting observations that may help illuminate this aim. But due to the open-ended nature of the research, what the observer chooses to focus on will change through the observation period. For example, in her work with under-three-year olds, investigating their engagement with the pre-school curriculum, Stephenson introduced an exercise where she walked through the nursery with each child, asking them to photograph their favourite spaces. It quickly emerged that what children liked about their chosen spaces wasn't necessarily the activity it was associated with but the people, in particular the other children, that they interacted with there. Furthermore, her relaxed approach meant she could get responses from children who initially did not seem to engage with the activity:

'This interaction underlined the importance of prolonged contact. On one level, spending 20 uninterrupted minutes with Cassidy allowed him to dip in and out of his thinking about

monsters, which was perhaps a response to my asking who he did not usually play with. If I had abandoned the process because Cassidy was not drawing his responses, or had moved away when he apparently lost interest, I would have missed his repeated re-introduction of the topic. And if I had not had contact with him over several weeks, I would not have known of his earlier interaction with a “monster” (Stephenson 2011, p. 151).

Stephenson’s research focussed on young children’s engagement with the pre-school curriculum. After spending 5 months with the children, she came to realise that ‘relationships with peers might be at the very heart of curriculum for children’ (ibid), a hypothesis which emerged from her sustained observations. It is a paradigmatic example of hypothesis generation from a series of systematic observations.

Another striking feature of Infant Observation is the inherently subjective nature of the descriptions. Observers describe how the scene appears to them, with no attempt to reconcile their descriptions with another person’s. This contrasts sharply with the quantitative analyses discussed in 4.1, where a high level of agreement between coders is taken as evidence that the patterns of behaviour are really there. With an observation, though, another viewer may disagree with how a researcher has described the scene, or with their interpretation of an interaction (e.g. one might disagree with Stephenson that Cassidy’s thoughts about

monsters shape his conceptualisations of different peer relations). Thus, one might argue that Infant Observation cannot serve as a foundation for hypothesis-generating research, because it does not clearly lead to sets of widely agreed phenomena to serve as explananda for theories. But here I want to develop a point from 3.2, which is that different people picking out different patterns should be celebrated rather than ignored. The observer's perspective is as much a part of the explananda as the pattern they perceive. Openly acknowledging this means that the potential limitations of a particular set of observations are made more apparent, a feature that needs to be accommodated by subsequent hypothesis-testing work. On the other hand, patterns might emerge from multiple different observation accounts that can be agreed upon by a group of researchers, leading to hypothesis-generating work in the more familiar and straightforward way.

While there is a general lack of hypothesis-generating research in psychology, the problem is particularly acute in developmental psychology. Several factors no doubt play a role: the (relative) newness of the field, difficulties in gaining approval to conduct the work, and the limitations of working with participants who can't yet talk or reliably follow instructions. It is only in the past decade that technology has allowed light, unobtrusive cameras to be worn by babies.<sup>11</sup>

---

<sup>11</sup>This technology has also led to new observational research with non-human animals, including mice (Sattler 2021), dogs (Williams 2011), and fur seals (Volpov 2021).

But in some respects observational work with babies is easier than hypothesis-testing experiments. There are few, if any, requirements that participants do a particular activity, and such requirements when they do exist are very minimal (e.g. ‘play with these toys’). Furthermore, because babies are such difficult and resource-intensive experimental participants, it would be beneficial if as many constraints as possible were in place before hypothesis-testing begins. Observational work is an over-looked source of such constraint, especially in this field.

## 5 Theory and observation re-visited

The previous section described research that I think constitutes the type of observational work that is better placed to help with the theory-crisis than the formalisation approach described in section 2.2. This part of the paper re-examines some of the points made in section 2 in the light of these case-studies.

Friedrich Steinle said of exploratory research that it ‘typically takes place in those periods of scientific development in which — for whatever reasons — no well-formed theory or even no conceptual framework is available or regarded as reliable’ (Steinle 1997, p. 70).<sup>12</sup> In the light of my arguments pressing for more observational work in developmental

---

<sup>12</sup>Steinle’s ‘exploratory’ work overlaps with my conception of observational work (see footnote 3).

psychology, practitioners in that field may take umbrage with the implication that their research lacks reliable conceptual frameworks. Yet in the wake of the replication ‘crisis’ that has swept through that field and many others in the psychological sciences, where findings that have grounded previously well-established theories are proving hard to replicate (e.g. Schuwerk et. al. 2022), such a conclusion may not be as far-fetched as it originally sounds. The problem, I believe, is one of presentation.<sup>13</sup> Why *should* psychologists be expected to have reliable conceptual frameworks given the relative youth of the discipline (as discussed in section 2.2)? It is a valuable aspiration for the field, but it is far from clear that it can be realised now. This connects with a point made in section 2.1, that psychology as a discipline seems driven by an urge to be seen as a respectable science (in part due to Popper’s dismissal of Freudian psychoanalysis), which has forced it into methods that are inappropriate for its current developmental stage. This is captured in the following passage from Paul Rozin:

‘It is characteristic of an advanced science to have many (but not all) of its studies located at the later stages in this process (i.e. natural or laboratory experiments, more and more refined and formal theorising). However, these activities only make sense if the earlier stages have provided an appropriate direction

---

<sup>13</sup>Lavelle (2022) argues a similar point in relation to the view that many failed replications constitute a ‘crisis’ for psychology.

for the later research. I claim that in modern social psychology, an understandable urge to become a more advanced science has led to a slighting of the critical early stage work. [...]

Prematurely advanced science stifles creativity, closes the eyes of the field to new phenomena, is prone to generate long lines of research that ultimately have little to do with the basic target of the field (i.e., the social world), and generally pulls people prematurely away from the real world where it all starts' (Rozin 2001, p. 5).

My case-studies were intended to showcase research that constitutes the 'early stage work' advocated by Rozin, that is, theory-lite observations of human behaviour.

One way to challenge my position would be to contest the claim that there is a paucity of data in psychology. For example, Eiko Fried (2020, p.217) describes psychology as 'data rich and theory poor' , while the philosopher Robert Cummins (2001/2011, p. 93) comments that 'In psychology we are overwhelmed with things to explain, and somewhat underwhelmed by things to explain them with', with both writers drawing inspiration from Paul Meehl's (1978) paper which argues along similar lines. My response is that they are right with regard to data collected with the aim of testing a particular hypothesis. Most effects are found because a group of researchers want to test the scope of their theories. For example, Simone Schnall and colleagues hypothesised that moral judgments are

based on emotional responses rather than rational reasoning processes, which led them to the more specific hypothesis that people experiencing disgust would evaluate actions more harshly than people in clean environs (Schnall, Benton & Harvey 2008, Schnall et al 2008). They designed experiments with the aim of testing this hypothesis, and the data supported it, leading them to claim that they had uncovered a new effect, namely, the correlation of moral judgment and disgust. This example is paradigmatic of how effects are discovered. The problem that has been well-established by now, is that uncovering these effects depends on relying on all kinds of auxiliary hypotheses concerning measurement, the operationalisation of the main factors, individual differences in how disgust experienced, etc. In addition to this there is the criticism that lies at the heart of the theory crisis, namely, is the hypothesis (that cleanliness correlates with lenient moral judgement) underdetermined by the theory that moral judgments are based on emotional responses? The hypothesis that people experiencing disgust are less bothered by moral transgressions because they are distracted by their disgust, seems to fit just as well. Even if it doesn't, the point still remains that the tie between theory and hypothesis is loose, at best. These objections do not affect my argument because they concern hypothesis-testing work. My call is for more research conducted at the stage which comes prior to this, namely, observational work. There is a glut of data and effects that have been generated in the name of hypothesis-testing; this does not affect the argument that there is a



lack of data appropriate for hypothesis-generation.

A more pressing objection would be that there is more observational work in psychology than is acknowledged in this paper. Caspar Van Lissa (2022) comes close to this, by arguing that when one looks more carefully at research purporting to be hypothesis-testing, it is in fact much closer to hypothesis-generation, and that this is the case for the majority of papers published in developmental psychology journals in the past decade. There are two responses. The first is that I am not sure that the work Van Lissa references is really gathered in the theory-lite manner which I believe is critical for observational work to serve the functions outlined in this paper. Secondly, Van Lissa may be correct without affecting the claims in this paper. For what is the good of observational work if it is not reported and acknowledged as such? The aim of observational work is to provide data that can ground phenomena that are widely agreed upon, which means that multiple people need to analyse and discuss it. If the work is not marked as a candidate for these processes, then given the vast amount of research published daily it is not going to be found by other interested parties. By contrast, a newly available open access corpus of linguistic data (e.g. transcripts or recordings) would be clearly marked and publicised in ways that made it clear that others can use it to find new patterns worthy of further research. Labelling is important.

There are three further points I want to make before closing. The first relates to the replication crisis. Many of the measures put in place to

remedy the replication crisis are targeted at hypothesis-testing work. Thus, pre-registration of methods and statistical analyses is intended to reduce HARK-ing and changing statistical methods to make the data better support the hypothesis, and making data openly available is meant to improve transparency by showing how the raw data relates to the published analyses. But if what is needed is more hypothesis-generating work, then these measures will have a limited impact on the root causes of the theory-crisis. A limited impact does not mean no impact: while finding a pattern worthy of further investigation is the aim of hypothesis-generating research, actually ascertaining if that pattern exists is the role of hypothesis-testing, which *has* been improved by measures introduced during the replication crisis. The two modes of research are complementary, and improvements in one will benefit the other (Wagenmakers et al. 2018).

The second point is that, while the title of this piece is ‘Less theory, more observation’, it is not a zero-sum game. Theorising and observational research can and should happen in tandem: one does not preclude the other. But there are some areas where formal theories have been introduced prematurely, and these are the places where the need for observational work is most pressing. The point is one of emphasis: while most researchers in developmental psychology are familiar with different techniques for formalising theories of learning (Tenenbaum et al. 2006) or social knowledge (Jin et al. 2024), I have met only one who had heard (vaguely) of Infant Observation. (For my part, I have been following developmental psychology

for several decades and only discovered Infant Observation very recently.) While it is unreasonable to ask people trained in quantitative analysis to also become experts in qualitative methods, there is a need for more collaboration and dialogue between these traditions (Yarkoni 2022). Facilitating this would be more beneficial in the short-term than the push to formalise. While I reject the view that more formal theory is required to resolve the current crisis, I nevertheless accept the claim that psychology lacks foundational theories and that this must be remedied (premise 1 of the theory crisis argument, see section 2.1). ‘*More observation*’ is offered as one path for achieving this.

Finally, as much of this paper focuses on developmental psychology there is the question of how well the presented arguments generalise to other fields. The question splits in two: is there a theory-crisis of similar proportions in other sub-disciplines of psychology and, if there is, does ‘more observation’ serve the same ameliorative functions as argued for here? Given the scale of the replication crisis in the discipline, I think it fair to claim that each sub-discipline is subject to a theory-crisis, the depth of which will vary between fields. Consequently, each will benefit from open-ended research characteristic of the observational work described here. However, the potential to be ‘theory-lite’ will depend on the technical specifications of the field. Observing and documenting the behaviours of a group of people is not-theory neutral (see 4.2), but it is less theory-laden than observations that require bespoke instruments and extensive training

to use them, as would be the case in cognitive neuroscience. Moreover the growing literature in cognitive ontology demonstrates the urgent need for wider samples of human behaviour to be observed and described in order to ground scientific cognitive concepts, which in turn trickles down to those fields such as neuroscience which utilise those concepts in a more technical domain (Chiao and Ambady 2007; Dewhurst 2018; Khalidi 2023; Feest 2025).

## **6 Conclusions**

Psychology's theory crisis reveals something more fundamentally problematic with the field: its lack of systematised observations grounding widely agreed upon phenomena to be explained. The phenomena we have plenty of, namely effects (section 5), are often uncovered in situations quite removed from the real world, rely on many auxiliary hypotheses, and are products of hypothesis-testing experiments where the hypotheses under test are underdetermined by wider, over-arching theories. The diagnosis presented in this paper does not call for more formalised theory and computer modelling, contra the voices presented in section 2.2. Instead, what is required is more observational data, gathered in a theory-lite way, which is openly available to multiple coders with different theoretical backgrounds and aims. When patterns emerge from these data, and are confirmed through hypothesis-testing experiments, then we have more

robust phenomena to serve as explananda for foundational theories. These then serve as the constraints for more specific hypotheses.

The main aim of this paper is to persuade readers that more observational work is a better remedy for the current theory crisis than formalisation. Section 4 offered examples that I think are paradigmatic of this; more abound. However, integrating this work will not be straightforward, as it often depends on very different skill-sets and frameworks to those familiar to many psychologists. Primatology could serve as a useful template, where field and lab-based researchers manage fruitful collaborations despite their differences in training and approach.<sup>14</sup> Philosophers too, particularly those who rely on psychological data in their research, need to become more aware of the different types of observational work relevant to their field. I want to highlight Van Lissa's remark (section 5) that much of developmental psychology is hypothesis-generating rather than hypothesis-testing, it's just not explicitly labelled as such. There is a rich heritage in philosophy of science regarding exploratory research (see footnote 3), and philosophers are well-placed to use this to better understand the nature of the research that they integrate into their theories of cognition.

It is not the case that all of psychology is affected to the same extent by the theory-crisis, and some sub-disciplines have more well-established theories than others. But in cases where the theory-crisis is particularly

---

<sup>14</sup>Thanks to Hannes Rakoczy for this suggestion.

acute, more observations rather than more theories offer the most promising way forward.

## **Acknowledgments**

This research was supported by a Humboldt Experienced Researcher fellowship at the Ruhr Universität Bochum, and a British Academy grant (SRG2000688). Versions of this paper were presented at the BA Leverhulme funded conference ‘Philosophical Issues in Replication’, the visiting speaker seminar at the University of Stirling, and the Philosophy, Psychology and Informatics group at the University of Edinburgh; I am grateful to participants at these events for their thoughtful comments. Thanks to Carrie Figdor for speedy and helpful feedback on a draft, and to Uljana Feest for discussion.

## **Biographical note**

Jane Suilin Lavelle is a philosopher in the School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

## References

- Catherine Allerton. 'Encountering children'. *Children: Ethnographic Encounters*. Ed. by Catherine Allerton. Encounters : Experience and Anthropological Knowledge. Routledge, 2016, pp. 1–15.
- Jens B. Asendorpf, Mark Conner, Filip De Fruyt, Jan De Houwer, Jaap J.A. Denissen, Klaus Fiedler, Susann Fiedler, David C. Funder, Reinhold Kliegl, Brian A. Nosek, Marco Perugini, Brent W. Roberts, Manfred Schmitt, Marcel A.G. Van Aken, Hannelore Weber, and Jelte M. Wicherts. 'Recommendations for Increasing Replicability in Psychology'. *European Journal of Personality* 27 (2 2013), pp. 108–119.
- Alan D. Baddeley. *Working Memory, Thought, and Action*. Oxford: Oxford University Press, 2007.
- Yehuda Baruch. 'HARKing can be Good for Science: Why, When, and How C/Should We Hypothesizing After Results are Known or Proposing Research Questions After Results are Known'. *Human Resource Management Journal* (2023).
- Esther Bick. 'Notes on Infant Observation in Psycho-Analytic Training'. *International Journal of Psychoanalysis* 45 (1964), pp. 558–566.
- Alexander Bird. 'Understanding the Replication Crisis as a Base Rate Fallacy'. *British Journal for the Philosophy of Science* 72.4 (2021), pp. 965–993.

- Denny Borsboom, Han L. J. van der Maas, Jonas Dalege, Rogier A. Kievit, and Brian D. Haig. ‘Theory Construction Methodology: A Practical Framework for Building Theories in Psychology’. *Perspectives on Psychological Science* 16.4 (2021), pp. 756–766.
- Richard M Burian. ‘Exploratory Experimentation and the Role of Histochemical Techniques in the work of Jean Brachet, 1938-1952’. *History and Philosophy of the Life Sciences* 19.1 (1997), pp. 27–45.
- Hasok Chang. *Is Water H<sub>2</sub>O? Evidence, Realism and Pluralism*. Dordrecht: Springer, 2012.
- Joan Y Chiao and Nalini Ambady. ‘Cultural Neuroscience: Parsing Universality and Diversity Across Levels of Analysis’. *Handbook of Cultural Psychology*. 2007, pp. 237–254.
- Gergely Csibra, György Gergely, Szilvia Bíró, Orsolya Koós, and Margaret Brockbank. ‘Goal attribution without agency cues: the perception of ‘pure reason’ in infancy’. *Cognition* 72.3 (1999), pp. 237–267.
- Sheila Degotardi. ‘Two Steps Back: Exploring Identity and Presence While Observing Infants in the Nursery’. *Educational Research with our Youngest: Voices of Infants and Toddlers*. Ed. by Eva Johansson and E. Jayne White. Springer, 2011, pp. 15–38.
- Joe Dewhurst. ‘Context-Sensitive Ontologies for a Non-reductionist Cognitive Neuroscience’. *Australasian Philosophical Review* 2 (2 2018), pp. 224–228.



- Itiel E. Dror and Christopher S. Schreiner. 'Neural Networks and Perception'. *Systems Theories and a Priori Aspects of Perception*. Ed. by J. Scott Jordan. Vol. 126. Advances in Psychology. North-Holland, 1998, pp. 77–85.
- Markus I. Eronen. 'Causal discovery and the problem of psychological interventions'. *New ideas in psychology* 59 (2020), pp. 100785–.
- Markus I. Eronen and Laura F. Bringmann. 'The Theory Crisis in Psychology: How to Move Forward'. *Perspectives on Psychological Science* 16.4 (2021), pp. 779–788.
- Caitlin M. Fausey, Swapnaa Jayaraman, and Linda B. Smith. 'From Faces to Hands: Changing Visual Input in the First Two Years'. *Cognition* 152 (2016), pp. 101–107.
- Uljana Feest. 'Data quality, experimental artifacts, and the reactivity of the psychological subject matter'. *European Journal for Philosophy of Science* 12.1 (2022), pp. 1–25.
- Uljana Feest. *Operationism and the Epistemology of Exploration in Experimental Psychology*. Chicago: University of Chicago Press, 2025.
- Uljana Feest. 'The experimenters' regress reconsidered: Replication, tacit knowledge, and the dynamics of knowledge generation'. *Studies in History and Philosophy of Science Part A* 58 (2016), pp. 34–45.
- Uljana Feest. 'Why replication is overrated'. *Philosophy of Science* 86 (5 2019), pp. 895–905.

- Uljana Feest and Friedrich Steinle. 'Experiment'. *Oxford Handbook of Philosophy of Science*. Ed. by P. Humphreys. Oxford: Oxford University Press, 2016, pp. 274–295.
- Simon Fitzpatrick. 'The primate mindreading controversy: a case study in simplicity and methodology in animal psychology'. *The Philosophy of Animal Minds*. Ed. by Robert W. Editor Lurz. Cambridge: Cambridge University Press, 2009, pp. 258–277.
- Samuel C Fletcher. 'How ( not ) to measure replication'. *European Journal for Philosophy of Science* 11 (2021), pp. 1–27.
- Samuel C. Fletcher. 'The role of replication in psychological science'. *European Journal for Philosophy of Science* 11 (1 2021).
- Eiko I Fried. 'Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature'. *Psychological Inquiry* 31 (4 2020). doi: 10.1080/1047840X.2020.1853461, pp. 271–288.
- György Gergely and Gergely Csibra. 'Teleological reasoning in infancy: The naïve theory of rational action'. *Trends in cognitive sciences* 7.7 (2003), pp. 287–292.
- Brian D Haig. 'Detecting Psychological Phenomena: Taking Bottom-Up Research Seriously'. *American Journal of Psychology Summer* 126 (2 2013), pp. 135–153.
- Norwood Russell Hanson. *Patterns of Discovery*. Cambridge: Cambridge University Press, 1958.

- John P. A. Ioannidis. ‘Why Most Published Research Findings Are False’.  
*Chance* 18 (4 2005), pp. 40–47.
- Chuanyang Jin, Yutong Wu, Jing Cao, Jiannan Xiang, Yen-Ling Kuo,  
Zhiting Hu, Tomer Ullman, Antonio Torralba, Joshua B Tenenbaum,  
and Tianmin Shu. ‘MMToM-QA: Multimodal Theory of Mind Question  
Answering’ (2024).
- Norbert L. Kerr. ‘HARKing: Hypothesizing After the Results are Known’.  
*Personality and Social Psychology Review* 2.3 (1998), pp. 196–217.
- Muhammad Ali Khalidi. *Cognitive Ontology : Taxonomic Practices in the  
Mind-Brain Sciences*. Cambridge University Press, 2023.
- Alan Kingstone, Daniel Smilek, and John D. Eastwood. ‘Cognitive  
Ethology: A new approach for studying human cognition’. *British  
Journal of Psychology* 99.3 (2008), pp. 317–340.
- Trudy Klauber. ‘Learning from infant and young child observation:  
Enriching development of future psychoanalytic psychotherapists and  
other professionals’. *Infant Observation* 25.2 (2022), pp. 83–88.
- Thomas Kuhn. *The Structure of Scientific Revolutions*. 3rd ed. Chicago:  
University of Chicago Press, 1962.
- Jane Suilin Lavelle. ‘When a crisis becomes an opportunity: the role of  
replications in making better theories’. *British Journal for the  
Philosophy of Science* 73 (4 2022), pp. 965–986.
- Sabina Leonelli. ‘Rethinking Reproducibility as a Criterion for Research  
Quality’. *Including a Symposium on Mary Morgan: Curiosity,*

- Imagination, and Surprise*. Bingley, UK: Emerald Publishing Limited, 2018, pp. 3–10.
- Scott O Lilienfeld and Irwin D Waldman, eds. *Psychological science under scrutiny: Recent challenges and proposed solutions*. John Wiley & Sons, 2017.
- Edouard Machery. ‘What is a replication?’ *Philosophy of Science* 87.4 (2020), pp. 545–567.
- Jeanne Magagna. ‘Three years of infant observation with Esther Bick’. *Infant observation : creating transformative relationships*. Ed. by Frances Thomson-Salo. Boca Raton, FL: Routledge, 2018.
- ManyBabies Consortium. ‘Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference’. *Advances in Methods and Practices in Psychological Science* 3.1 (2020), pp. 24–52.
- Anne McFadyen, Kathryn Gray Armstrong, and Beatrice Anderson. ‘The Voice of the Infant’. *Infant Observation* 25.2 (2022), pp. 104–122.
- Marcia McNutt. ‘Reproducibility’. *Science* 343 (2014), p. 229.
- Paul E Meehl. ‘Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology’. *Journal of consulting and clinical psychology* 46.4 (1978), pp. 806–834.
- Michael Muthukrishna and Joseph Henrich. ‘A problem in theory’. *Nature Human Behaviour* 3.3 (2019), pp. 221–229.
- Klaus Oberauer and Stephan Lewandowsky. ‘Addressing the theory crisis in psychology’. *Psychonomic bulletin & review* 26.5 (2019), pp. 1596–1618.

- Open Science Collaboration. ‘Estimating the reproducibility of psychological science’. *Science* 349.6251 (2015), pp. 1–8.
- Freek JW Oude Maatman. ‘Why Formal Modelling Cannot Solve the Theory Crisis’.
- Michael T. Rizzo and Melanie Killen. ‘How social status influences our understanding of others’ mental states’. *Journal of Experimental Child Psychology* 169 (2018), pp. 30–41.
- Donald J. Robinaugh, Jonas M. B. Haslbeck, Oisín Ryan, Eiko I. Fried, and Lourens J. Waldorp. ‘Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction’. *Perspectives on psychological science* 16.4 (2021), pp. 725–743.
- Iris van Rooij and Giosuè Baggio. ‘Theory before the test: How to build high-verisimilitude explanatory theories in psychological science’. *Perspectives on Psychological Science* 16.4 (2021), pp. 682–697.
- Paul Rozin. ‘Social Psychology and Science: Some Lessons From Solomon Asch’. *Personality and Social Psychology Review* 5.1 (2001), pp. 2–14.
- Nicholas J. Sattler and Michael Wehr. ‘A Head-Mounted Multi-Camera System for Electrophysiology and Behavior in Freely-Moving Mice’. *Frontiers in Neuroscience* 14 (2021).
- Anne M Scheel, Mitchell R M J Schijen, and Daniël Lakens. ‘An Excess of Positive Results: Comparing the Standard Psychology Literature With Registered Reports’. *Advances in Methods and Practices in Psychological Science* 4 (2 2021), p. 25152459211007467.

- Anne M. Scheel. ‘Why most psychological research findings are not even wrong’. *Infant and Child Development* 31.1 (2022), pp. 1–9.
- Simone Schnall, Jennifer Benton, and Sophie Harvey. ‘With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments’. *Psychological science* 19.12 (2008), pp. 1219–1222.
- Simone Schnall, Jonathan Haidt, Gerald L. Clore, and Alexander H. Jordan. ‘Disgust as Embodied Moral Judgment’. *Personality and Social Psychology Bulletin* 34.8 (2008). PMID: 18505801, pp. 1096–1109.
- Tobias Schuwerk, Dora Kampis, Manuel Bohn, Cynthia Fisher, Charlotte Grosse Wiesmann, Daniel C Hyde, Louisa Kulke Friedrich-Alexander, Kyle Mahowald, Olivier Mascaró, Julia Prein, Gal Raz, Rebecca Saxe Dana Schneider Friedrich-Schiller, Victoria Southgate, Francis Yuen, Amanda Rose Yuile, Lucie Zimmer, and Michael C Frank. ‘In-principle acceptance of Registered Report: Action Anticipation Based on an Agent’s Epistemic State in Toddlers and Adults’. *Child Development* (2022).
- Anne-Marie Sim. ‘Learning to be a child in Greater London’. *Children: Ethnographic Encounters*. Ed. by Catherine Allerton. Encounters : experience and anthropological knowledge. Routledge, 2016, pp. 73–85.
- Linda B. Smith, Chen Yu, and Alfredo F. Pereira. ‘Not your mother’s view: The dynamics of toddler visual experience’. *Developmental Science* 14.1 (2011), pp. 9–17.

- Linda B. Smith, Chen Yu, Hanako Yoshida, and Caitlin M. Fausey.  
 ‘Contributions of Head-Mounted Cameras to Studying the Visual  
 Environments of Infants and Young Children’. *Journal of Cognition and  
 Development* 16 (3 2015), pp. 407–419.
- Friedrich Steinle. ‘Entering new fields: Exploratory uses of  
 experimentation’. *Philosophy of science* 64.S4 (1997), S65–S74.
- Alison Stephenson. ‘Taking a ‘Generous’ Approach in Research with Young  
 Children’. *Educational Research with our Youngest: Voices of Infants  
 and Toddlers*. Ed. by Eva Johansson and E. Jayne White. Springer  
 Science+Business Media B.V., 2011, pp. 135–156.
- Joshua B. Tenenbaum, Thomas L. Griffiths, and Charles Kemp.  
 ‘Theory-based Bayesian models of inductive learning and reasoning’.  
*Trends in cognitive sciences* 10.7 (2006), pp. 309–318.
- Bas C. Van Fraassen. *The scientific image*. Clarendon library of logic and  
 philosophy. Oxford: Clarendon, 1980.
- Caspar J. Van Lissa. ‘Developmental data science: How machine learning  
 can advance theory formation in Developmental Psychology’. *Infant and  
 Child Development* (2022).
- Beth L. Volpov, Andrew J. Hoskins, Brian C. Battaile, Morgane Viviant,  
 Kathryn E. Wheatley, Greg Marshall, Kyler Abernathy, and  
 John P. Y. Arnould. ‘Identification of Prey Captures in Australian Fur  
 Seals (*Arctocephalus pusillus doriferus*) Using Head-Mounted

- Accelerometers: Field Validation with Animal-Borne Video Cameras'. *PLOS ONE* 10.6 (2015), pp. 1–19.
- Eric-Jan Wagenmakers, Gilles Dutilh, and Alexandra Sarafoglou. 'The Creativity-Verification Cycle in Psychological Science: New Methods to Combat Old Idols'. *Perspectives on Psychological Science* 13.4 (2018), pp. 418–427.
- Jelte M Wicherts, Coosje LS Veldkamp, Hilde EM Augusteijn, Marjan Bakker, Robbie Van Aert, and Marcel ALM Van Assen. 'Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking'. *Frontiers in psychology* (2016), p. 1832.
- Michael Wilby. 'The simplicity of mutual knowledge'. *Philosophical explorations* 13.2 (2010), pp. 83–100.
- Fiona J. Williams, Daniel S. Mills, and Kun Guo. 'Development of a head-mounted, eye-tracking system for dogs'. *Journal of Neuroscience Methods* 194.2 (2011), pp. 259–265.
- Tal Yarkoni. 'The generalizability crisis'. *Behavioral and Brain Sciences* 45 (2022).
- Chen Yu and Linda B. Smith. 'Embodied attention and word learning by toddlers'. *Cognition* 125.2 (2012), pp. 244–262.