# Improving postgraduate researchers' inferences with a philosophical workshop

Robbie Clark[1,2], Margarida Hermida[3], Nicole Russell Pascual[4, 5]
Marcus Munafò[1,2], James Ladyman[6]

[1]School of Psychological Science, University of Bristol
[2]MRC Integrative Epidemiology Unit at the University of Bristol
[3]Department of Philosophy, King's College London
[4]Department of Psychology, University of Exeter
[5]Department of Psychology, University of Michigan
[6]Department of Philosophy, University of Bristol

ORCID Identifiers
Robbie Clark: https://orcid.org/0000-0002-2160-313X
Margarida Hermida: https://orcid.org/0000-0002-0259-109X
Nicole Russell Pascual: https://orcid.org/0000-0001-7873-4184
Marcus Munafò: https://orcid.org/0000-0002-4049-993X
Correspondence to: robbie.clark.2018@bristol.ac.uk

## Abstract

Postgraduate research training in the United Kingdom often narrowly focuses on domain-specific methods, neglecting wider philosophical topics such as epistemology and scientific method. Consequently, we designed a workshop on (inductive, deductive, and abductive) inference for postgraduate researchers. We ran the workshop three times with ($N$ = 29) attendees from across four universities, testing the potential benefits of the workshop in a mixed-method, repeated measures design. Our core aims were to investigate what attendees learned from the workshop, and whether they felt it had impacted on their research practices six months later. Overall, learning inferential logic benefitted postgraduate researchers in various ways and to varying degrees. Six months on, roughly half of attendees reported being more critical of key aspects of research such as inferences and study design. Additionally, some attendees reported more subtle effects, such as prompting new lines of thought and inquiry. Given that self-criticism and scepticism are fundamental intellectual virtues, these results evidence the importance of embedding epistemological training into doctoral programmes across the UK.

**Keywords:** *philosophy of science; inference; workshop; postgraduate curriculum*

# Introduction

Becoming a scientist in the in the UK typically involves highly specialised training in domain-specific methods. Scientists are often taught what to do, but not necessarily the underlying justification for their practices (Grüne-Yanoff, 2014; Johansson, 2015). This narrow focus can result in scientists becoming highly specialised within their area, without learning the philosophy of *their science*, nor *of science* more broadly. Discussions on the importance of teaching philosophy of science to researchers have recurred for decades across a range of disciplines (e.g., Andreoletti & Maugeri, 2019; Boniolo & Campaner, 2020; Kampourakis & Uller, 2020; Laplane et al., 2019; Spike, 1991). For example, Prather and colleagues (2009) found that graduate students in ecology, generally lacking a background in philosophy of science, sometimes found it difficult to explicate how their research questions and designs interacted with existing theories.

Understanding scientific methodology and its philosophical underpinnings can improve researchers' awareness of the deeper rationale for their research practices and discipline more broadly (Grüne-Yanoff, 2014; Meehl, 1993). In turn, this can help researchers to appraise the appropriateness of their practices (Boniolo & Campaner, 2020). This has been evidenced in the various 'crisis' debates within psychology. In response to concerns about the reliability of research, many researchers have returned to philosophical foundations in order to contextualise and address some of the issues (e.g., Albert et al., 2020; Derksen, 2019; Eronen & Romeijn, 2020; Flis, 2019; Morawski, 2019; Reiter, 2013). Nevertheless, formally educating researchers in philosophy of science remains the exception, rather than the rule. Moreover, even if philosophy is taught, there is no guarantee of its accuracy or extensiveness (Blachowicz, 2009).

Recent philosophical workshops to improve cross-disciplinary collaborations have provided promising evidence for their benefits. Research into a workshop on epistemology and ontology concluded that philosophical discussions on these topics can help to facilitate collaboration across disciplines (Eigenbrode et al., 2007; O'Rourke & Crowley, 2013). The authors state that philosophical discourse provides a common vocabulary for the discussion of methodological issues, which enables researchers from different disciplines to work together more effectively. Moreover, it encourages

attendees to consider their own ontological and epistemological assumptions, which otherwise remain implicit, unacknowledged, and therefore unchallenged.

Inspired by this work and philosophical theory, we developed, piloted, refined, and tested an interactive workshop on inference (induction, deduction, and abduction – *henceforth IDA*) for researchers at postgraduate level (and possibly beyond). The workshop begins with a section on scientific methodology. This highlights the context-dependence of research processes and that there is no single scientific method (Haig, 2019; Woodcock, 2014). The remainder of the workshop extensively covers IDA inference. Combined, these three types of reasoning underpin all inferences and are utilised in various ways depending on the type of scientific method one applies (Haig, 2019). Nevertheless, the logical form of these inference types is seldom taught to those outside of philosophy.

The choice to teach IDA inference was also motivated by the fact that Popperian hypothetico-deductivism is often taught in introductory courses as central to the scientific method (Blachowicz, 2009; Popper, 1959). Despite this, such courses rarely cover the logic underpinning Hume's problem of induction, the logical basis of falsification, nor the issues that arise in practice due to underdetermination problems (Harding, 1975; Hume, 1779). Ultimately, we hoped to elucidate the interconnectedness and fallibility of inferences. We wanted attendees to identify how they can arrive at more trustworthy conclusions, recognising where errors are most likely to occur, and the need for epistemic humility within science (Fjelland, 2022).

This study reports repeated testing of the workshop. For information on the piloting and refinement of the workshop, see Clark et al. (2022).

## Research aims

This study had three aims: two primary and one secondary. The primary aims were to investigate whether and how the workshop influenced attendees' (1) understanding and (2) research behaviour. The secondary aim was (3) to gather feedback on the content and format. More specifically, primary aim 1.a was to ascertain whether the workshop changed attendees' understanding of IDA inference. Primary aim 1.b was to understand the main lessons they took from the workshop. Primary aim 2.a was to explore attendees' predictions about whether their research behaviours would change

in light of the workshop. Primary aim 2.b was to investigate any reported influence on their research behaviours, six months after the workshop. We used this to assess whether predictions aligned with final reports.

# Method

## Study design

We conducted the workshop three times, collecting the same data for each as part of a mixed-methods design. Attendees completed surveys before the workshop (T1), immediately after (T2), and six months after (T3). Each survey included a 15-question test on inference, which we used as a quantitative, repeated measure, to operationalise changes in their understanding of IDA inference between timepoints T1, T2, and T3. The remaining data were all qualitative and analysed using conventional content analysis.

Additionally, at the end of the workshops, attendees completed a 20-minute writing exercise. We used this to gather self-report data on the main lessons they had taken from the workshop, and any predicted influence on their future research behaviours. The follow-up survey at T3 also asked attendees whether any of their predictions had come to fruition. The study protocol was pre-registered on the Open Science Framework (https://doi.org/10.17605/OSF.IO/QEJ9N).

## Participants

We ran the workshop three times with different groups of attendees. Workshops 1 and 3 were conducted at the University of Bristol in November 2022 and February 2023 respectively. Workshop 2 was run at the University of Exeter in November 2022. Across all three workshops, 29 attendees completed all of the requisite workshop sessions and surveys to be included in the analysis (26 PhD students, 2 MRes students, and 1 post-doctoral researcher – see Table 1). Six more researchers could only attend one of the sessions. As they only experienced half of the workshop, we excluded them from the final sample and analysis.

**Table 1.** Attendee characteristics for all 3 workshops. Their anonymised identifiers (IDs), career stages, university affiliations, schools, and research areas. Workshop 1: B1-B8; Workshop 2: E1-E9; Workshop 3: B9-B22.

| ID | Career Stage | Affiliation | School / Faculty | Research Area |
|---|---|---|---|---|
| B1 | PhD Year 2 | University of Bristol | Department of Computer Science | Cybersecurity |
| B2 | PhD Year 4 | University of Bristol | Faculty of Arts, Dept of Archaeology and Anthropology | Social anthropology |
| B3 | PhD Year 1 | University of Bristol | School of Psychological Science | Sociology of science |
| B4 | Post-doc Researcher (<1 year post PhD) | University of Bristol | School of Psychological Science | Health behaviours |
| B5 | PhD Year 2 | University of Bristol | School for Policy Studies | Social policy |
| B6 | PhD Year 4 | University of Bristol | School of Physics | Biophotonics |
| B7 | PhD Year 5 | University of Bristol | Bristol Medical School | Molecular biology |
| B8 | PhD Year 2 | University of Bristol | Department of Aerospace Engineering | Aeroelastic tailoring |
| E1 | PhD Year 1 | University of Exeter | Faculty of Environment. Science and Economy | Physical geography |
| E2 | PhD Year 3 | University of Exeter | School of Education | Ethics education |
| E3 | PhD Year 1 | University of Exeter | Faculty of Environment, Science, Economics | Mathematical modelling |
| E4 | PhD Year 4 | University of Exeter | Sport and Health Sciences | Nutritional physiology |
| E5 | PhD Year 1 | University of Exeter | Faculty of Health and Life Sciences | Medical mycology and |

| | | | | |
|---|---|---|---|---|
| E6 | PhD Year 1 | University of Exeter | Faculty of Health and Life Sciences | Population genetics |
| E7 | PhD Year 1 | University of Exeter | Faculty of Health and Life Sciences | Medical mycology |
| B9 | PhD Year 3 | University of Bristol | Medical Research Council Integrative Epidemiology Unit | Epidemiology |
| B10 | PhD Year 1 | University of Bristol | School of Psychological Science | Cognitive psychology |
| B11 | Master's by Research | University of Bristol | School of Biology | Forest ecology |
| B12 | Master's by Research | University of Bristol | Faculty of Life sciences, School of Cellular and Molecular Medicine | Antibiotic resistance |
| B13 | PhD Year 1 | University of Bristol | School of Geographical Sciences | Digital and green transition |
| B14 | PhD Year 2 | University of Bristol | Centre for Ethics in Medicine, Department of Population Health Sciences | Medical ethics and law |
| B15 | PhD Year 1 | University of Bristol | Population Health Sciences | Mental health |
| B16 | PhD Year 3 | University of the West of England | School of Psychology | Counselling psychology |
| B17 | PhD Year 1 | University of Bath | Department of Psychology | Clinical psychology |
| B18 | PhD Year 1 | University of Bath | Department of Psychology | Social psychology |
| B19 | PhD Year 2 | University of Bath | Department of Psychology | Cognitive neuroscience |
| B20 | PhD Year 1 | University of Bath | School of Education | Leadership and management in higher education |
| B21 | PhD Year 2 | University of Bath | Department for Health | Nutrition and metabolic physiology |
| B22 | PhD Year 3 | University of Bath | Department of Psychology | Biological psychiatry |

## Recruitment

To be eligible, attendees had to be a postgraduate researcher (or higher) in an empirical science. We advertised through the Bristol Doctoral College (BDC) events programme and targeted mailing list across the four universities. In total, 55 people signed the consent form. Of those, 35 were able to attend at least one workshop session. The other 20 did not participate.

## Materials

### *Research materials*

We collected all of the study data using four online surveys:

1) Pre-workshop survey (sent before the workshop)
2) Reflective writing exercise (completed in the last 20 minutes of session 2)
3) Post-workshop survey 1 (sent immediately after the workshop)
4) Post-workshop survey 2 (sent 6 months after the workshop)

The reflective writing exercise asked attendees to write about two topics at length: (1) the main things they took from the workshop; and (2) whether they might change any of their future research practices/behaviours in light of the workshop. Following research on the intention-behaviour gap, we instructed attendees to set simple, discrete, and achievable goals (Gollwitzer & Sheeran, 2006; Sheeran & Webb, 2016). We also asked them to state the factors that would help them achieve each goal, and to consider potential obstacles to achieving the goal, and how they might be overcome.

The pre-workshop survey collected demographic data on attendees': career stages; funding sources; university and departmental affiliations; and specific research areas. Post-workshop survey 1 included an anonymous feedback form, on which attendees could write about anything they liked, disliked, or would change about the workshop's content and format. Post-workshop survey 2 asked attendees to consider whether the workshop had influenced any of their research practices/behaviours over the last 6 months. It also included a closed question on whether they had looked back at what they wrote during the reflective writing exercise. For the full set of out outcome measures, see the *Online Supplement: https://osf.io/9pxqr*.

### *Inference Test creation*

All three surveys included an Inference Test, which we created to measure primary aim 1.a. Questions 1-9 collectively tested respondents' knowledge of the *meaning* of the terms induction, deduction, and abduction. Questions 10-15 were more challenging, testing a nuanced understanding of the *problems* inherent in these types of inference. We created and refined the test as follows. RC created a first draft, which was discussed and revised with JL. MM then improved some phrasing. MH then took the test, scoring 100% and offering a 15th question. Satisfied with this iteration, we trialled it with 6 philosophy PhD students, who averaged 13/15. We assumed they should score highly, and that their mistakes would provide valuable insights. With their feedback, we tweaked the wording used in 5 of the questions. We also adjusted the scoring of 3 questions to account for conceptual overlap between induction and abduction. We then piloted the test with 5 psychology PhD students, who found this final version easy to understand. For more comprehensive information, see our OSF project page.

### *Workshop materials*

The workshops followed the same set of slides (https://osf.io/p5hru), as were used in Workshop 2 of our pilot study (Clark et al., 2022). The addition of the reflective writing exercise was the only substantive change. Everything else was aesthetic.

## Procedure

Figure 1 details the structure and rationale for the workshop activities. Each workshop involved 2 sessions, 3-hours each, run one week apart. The workshop was almost identical to our pilot Workshop 2 (Clark et al., 2022). The only procedural addition was the reflective writing exercise during the last 20 minutes of session 2. The reflective writing exercise was intended to give attendees a chance to reflect on what they had taken from the workshop, while simultaneously allowing us to collect research data.

Immediately after the second workshop session, we sent attendees the post-workshop survey, which included a link to an anonymous feedback form. All attendees completed this survey within two weeks of receipt. Six months after they had attended the workshop, we sent attendees the second post-workshop survey. All attendees completed this survey within three weeks. Once they had responded, we sent their £25 voucher and the study debriefing information.

| Workshop Activity | Output |
|---|---|
| **Session 1:**<br>1. Overview of the workshop's structure and rationale.<br>2. Intro to scientific methodology and presentation of an example methodological schematic.<br>3. Pictorially represent individual research processes, creating methodological schematics.<br>4. Discuss and critique each other's schematics in pairs, editing as necessary.<br>5. Interactive seminar on inductive inference.<br>6. Free discussion.<br>**Session 2:**<br>1. Quick recap of induction.<br>2. Interactive seminar on deductive inference.<br>3. Interactive seminar on abductive inference.<br>4. Summary slides, comparing all 3 types of inference.<br>5. Free discussion.<br>6. Reflective writing exercises (1. Main lessons from the workshop and 2. Predicted changes to research practices). | **Session 1:**<br>1. Signpost procedure and rationale.<br>2. Highlight that research processes differ, and exemplify the next activity.<br>3. Visualise one's research process, and later identify where inferences are made therein.<br>4. Give a chance to improve schematics and form relationships with other attendees.<br>5. See *Interactive Seminar Notes*.<br>6. Consolidate learning and allow attendees to voice any remaining thoughts or questions.<br>**Session 2:**<br>1. Check and refresh memory.<br>2. See *Interactive Seminar Notes*.<br>3. See *Interactive Seminar Notes*.<br>4. Consolidate learning and contrast types of inference.<br>5. Invite critical discussion (e.g., on whether personal practices might now change).<br>6. Reflect on key lessons (hopefully consolidating learning and retention) and possibly set intentions for future research practices. |
| ***Interactive Seminar Notes*** | ***Interactive Seminar Output*** |
| *Sequence of the 3 'Interactive seminars' on inference:*<br>1. Attendees discuss in pairs what they understand by that inference type.<br>2. The facilitator introduces the concept, providing examples, asking questions, and eliciting lots of further examples from attendees.<br>3. In pairs, attendees identify instances of that inference type in the *H. pylori* example. Then discuss as group.<br>4. Attendees independently consider, and then discuss in pairs, how they use that inference type in their own research, annotating schematics.<br>5. Attendees discuss how one can be more confident in such inferences, both in an abstract sense, and in terms of their own research.<br>❖ *Deduction contains additional slides on deductive falsification. Attendees are asked to consider whether they use, or could use, falsification in their research.*<br>❖ *10-minute breaks occur every hour.* | *Rationale for the Interactive seminars' sequence:*<br>1. Set the tone of the session (discursive/ cooperative), and check knowledge.<br>2. Ensure each attendee has a strong baseline understanding of the inference type.<br>3. Consolidate understanding of that inference type and practise identifying inferences across a research process.<br>4. Recognise the multitude of inferences one makes, their interdependence, possible assumptions, and risks.<br>5. Identify common problems and how to increase trustworthiness of inferences, both in general and in terms of personal practices.<br>❖ *Introduce those unfamiliar to the concept and the possibility that they could use falsification in their own research.*<br>❖ *Ensure chances to relax, reflect, socialise, and get refreshments.* |

**Fig. 1** Implementation plan, detailing the sequence of activities for each workshop session, and the intended output from each activity. The bottom panels detail the sequence and rationale of the sections on induction, deduction, and abduction.

## Analysis

### *Inference Test scores*

RC scored attendees' Inference Test results for questions 1-3 and 7-15 using our pre-determined marking scheme. Questions 4-6 asked attendees to provide an example of an inductive, deductive, and abductive inference, respectively. These were qualitatively double coded, independently by RC and MH. Our marks were the same on 236/261 responses. We resolved all of the 25 discrepancies by discussion. Inter-coder reliability was Cohen's $\kappa = 0.78$ (see *Online Supplement* for the calculation).

### *Qualitative analyses*

RC analysed all of the qualitative data using conventional content analysis, allowing the data to determine the codes (Hsieh & Shannon, 2005). He began by creating high-level categories for each research question, as well as a 'miscellaneous' category for all data which did not pertain to these questions. He then freely coded data, grouping codes under the higher-level categories (i.e., research questions) they pertained to. Having done this for all of the data, he then re-visited each code, checking whether each quote reflected the concepts. This process helped to refine the code labels, aggregate certain codes, and identify redundant ones. He then began thematically grouping codes based on their conceptual similarity, paying particular attention to codes which reflected more attendees' experiences. In doing so, themes were created out of the most common ideas, giving voice to dissenting voices, but also dropping some codes that were unique, uninformative, or did not fit the higher categories.

# Results

## Understanding of inference (induction, deduction, abduction)

Primary aim 1.a was to learn whether and how the workshop influenced attendees' understanding of inference. We operationalised and measured their understanding of inference before, after, and 6-months after the workshop using a 15-question Inference Test. Figure 2 descriptively presents these test results.

On average, attendees scored much higher on the test after attending the workshop ($M_{before} = 5.6$, $M_{after} = 12.1$, $M_{6months} = 10.3$, $SD_{before} = 3.9$, $SD_{after} = 2.6$, $SD_{6months} = 4.0$). Six months on, mean scores were a bit lower than immediately post-workshop and had a larger range. Still, they were considerably higher than before the workshop, showing a

high level of retention. Nine attendees scored 8/15 or above before the workshop, creating a slight ceiling effect which is visible in the post-workshop plots. Consequently, it is worth noting that the *median* scores were 4, 13, and 11 respectively.

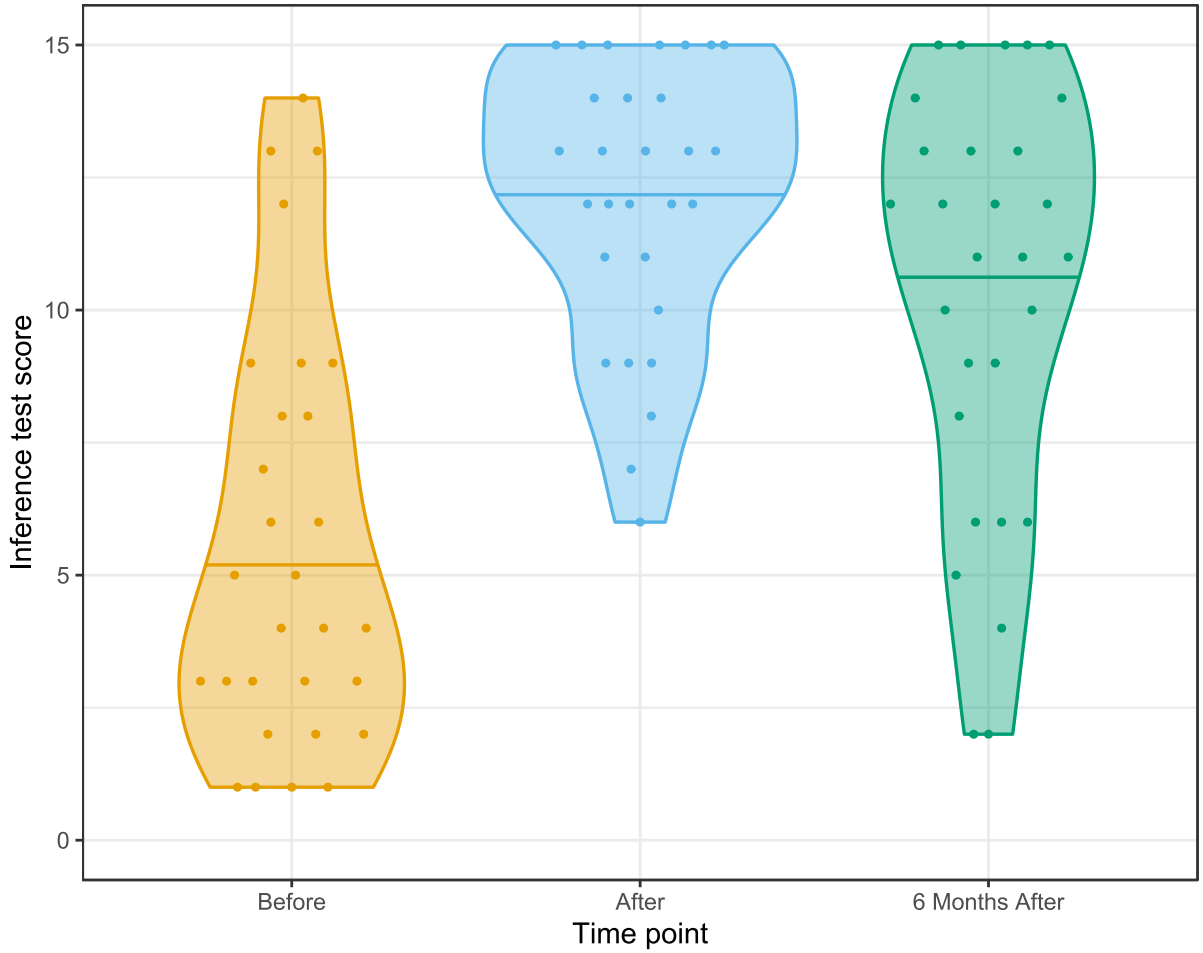


**Fig. 2** Attendees' scores on the Inference Test at each timepoint (before, after, and 6-months after the workshop). Attendees' total scores (out of 15) are presented as dots. The three shapes represent the distribution of scores at each timepoint. Hence, width indicates the frequency of scores near that value. The horizontal bars are means.

The remaining results all involved qualitative analysis and are thematically structured based on the research aim they related to (see Table 2). Attendee feedback is reported in the *Online Supplement S.1.*

**Table 2.** The Major themes related to each Research aim.

| Research aim | Major theme |
|---|---|
| 1.b. Influence on understanding | Main lesson 1: Understanding IDA |
| | Main lesson 2: Critical thought |
| 2.a. Predicted influence on behaviour | Predicted influence: Addressing uncertainty |
| 2.b. Reported influence on behaviour | Reported influence 1: Critical of design |
| | Reported influence 2: Critical of inferences |
| | Reported influence 3: Indirect influences |
| | Reported influence 4: No influence |

## Influence on understanding

Primary aim 1.b was to qualitatively explore other ways in which the workshop influenced attendees' understanding. At the end of the workshop, attendees were asked to write about the main lessons they took from the workshop. RC analysed these responses and collated the most prominent ideas into two, interrelated themes. The following sections, ***understanding IDA*** (induction, deduction, and abduction) and ***critical thought***, explore attendees' reports on the benefits of learning about IDA, many of which corresponded with our intended learning outcomes.

### *Main Lesson 1: Understanding IDA*

> "The session significantly improved my understanding of the different types of inference, their applications, usefulness, and limitations." (B2)

Most attendees mentioned that their understanding of induction, deduction, and abduction had improved. For many attendees, this was their first formal introduction to these concepts; their underlying logic, strengths, and weaknesses. Several mentioned how the terminology and associated understanding provided them with a framework with which to think about arguments, and to "critically articulate [their] sentiments" (B1).

> "Now I know what each is, I can recognise when I am using them, and think critically about the assumptions and premises I am using to underpin my inferences." (B10)

One of the primary goals of the workshop was to move attendees from a basic understanding of IDA to a nuanced understanding of inferential fallibility. Multiple

attendees' responses confirmed that this was both achieved and also beneficial to them. For example, B21 "hadn't appreciated before that inductive inferences can never be 100% certain, and these often provide the premises for abductive and deductive inferences". Hence, learning about each type of inference and their interconnectedness served to elucidate the foundations of epistemic fallibilism.

> "The workshop has highlighted how much of our research is based on inductive reasoning, and the fragility of our hypotheses and conclusions as a result of this. This isn't necessarily new information… but the consequences of this hadn't been spelled out before." (B19)

Another goal of the workshop was to apply the concepts directly to attendees' work, encouraging them to draw conclusions which were pertinent to them. Multiple responses supported the success of this approach. Several attendees whose research involved hypothesis testing took away a "greater appreciation for generating more specific hypotheses in which the variables [one is] testing are clearly defined in terms of the measurements [one is] performing" (B17). This idea of conceptual or design specificity was frequently mentioned. For our almost exclusively postgraduate attendees, the workshop served as a reminder to set an achievable project scope, rather than trying "to address all possible variables" (B11).

> "I know I will make my research easier and more reliable by increasing the specificity of my research questions. These workshops have made me think about the scope of what I can actually achieve within the confines of PhD research and how I can narrow my research down into something achievable." (B10)

### *Main Lesson 2: Critical thought*

> "The main ideas I have taken away from this workshop is that often in science and when conducting research we are not always fully transparent about the prior assumptions we have made. This can include the theories, literature, or common knowledge that we assume to be true that then leads us to a particular research question, study, sample, or dataset. This seems important given that our conclusions are only as valid as the assumptions we use to draw our conclusions." (B4)

As intended, establishing a good understanding of the logic of and nuance underpinning IDA resulted in many attendees reporting a more critical attitude towards inferences. Our hope was that breaking arguments down into a logical structure would help attendees to visualise how the strength of inferences is contingent on the individual premises therein. In line with this aim, almost half of attendees reflected on the different kinds of assumptions made in research, the importance of "questioning" them (E6), and making them as "specific and explicit as possible" (E7). Considering inferences in this logical way suggested to B18 that they "should think more deeply about the logic chain used in [their] research."

> "Inferences in general rely on the strength of our observations or evidence. Strengthening an inference can be as simple as gathering more data or curtailing the breadth of the inference to specific situations." (B7)

One of the most commonly discussed assumptions was the foundational literature on which one's research stands. Hence, a commonly reported lesson was the need to be more critical of literature one cites; scrutinising their methodology and inferences more carefully. Several attendees mentioned that their improved understanding of the logic of inference would help them in being more critical.

> "The workshop made me realise there are proper and structured ways to explain findings, discuss results and make conclusions about the theories from my research. Following this course… I will try to critically analyse every paper that will be included in my extensive literature review." (B14)

Attendees reported it was important to be critical not only of others' inferences, but also of one's own. On the importance of reflexivity, B7 highlighted that "the most fallible part of the equation is the brain doing the thinking. This means assessing our own assumptions and biases is a key part of making our work sound" (B7). Additionally, several attendees acknowledged the importance of precision, harking back to the ideas of specificity and achievable project scope.

> "The workshop highlighted how important it is to be precise in our inferences and conclusions… to be specific about the parameters about which our findings might be true, to recognise the limitations of our findings and to be aware of when our theories may no longer hold." (B19)

## Predicted influence on behaviour

Primary aim 2.a was to investigate whether and how the workshop might influence attendees' research behaviour. At the end of the workshop, attendees wrote about predicted changes to their research behaviours. The analysis of these responses resulted in a single main theme: ***addressing uncertainty***. Although the content of this theme contains the most popular ideas, each mentioned by multiple attendees, some ideas should be mentioned that were not included. Firstly, three attendees did not foresee their practices changing in light of the workshop. One was "too close to the end of [their] research to make changes to research practices now" (B6). The other two felt that acquiring the terminology of IDA had been useful, but that their practices would not change.

Also noteworthy was that two attendees found the act of diagrammatically breaking down their research into methodological schematics particularly useful. P8 said that "drawing a schematic of the research process is very useful for reflection", and that they would "try to keep updating" it throughout their PhD. The other felt that it was "an incredibly useful way of guiding the reader of a paper/thesis chapter through the analytical decisions made", and that they would use this as a figure in their thesis (B9).

### *Predicted Influence on behaviour: Addressing uncertainty*

Many of the predicted influences were the practical implications of the lessons discussed in the theme ***critical thought.*** The most commonly mentioned ideas to achieving this involved creating a solid derivation chain within experiments. This means maximising confidence in each premise in a study plan and the logic behind it. Hence, roughly a quarter of attendees mentioned that they were likely to be more critical of literature on which their research was founded. Similarly, many discussed wanting to take "a step back" and examine what they assume to be true or false when designing their studies (B4). In addition to interrogating one's assumptions, B21 also mentioned detailing how "confident [they feel] in these assumptions".

The other most prominent idea was *a priori* study specificity (i.e., being very detailed at the design phase). These goals all mapped onto the main lessons around having a clear and achievable project scope. For example, one of B11's goals was to "make more clear-cut statements/hypotheses". Five attendees linked *a priori* specificity to the aim of facilitating the possibility of deductive falsification. They mentioned how the workshop

had made them think more about falsification, and the importance of considering "what it would mean if [one] got negative results" (B21). Hence, they recognised the value in using precise language in predictions (B11) and considering the assumptions underpinning a theory or hypothesis (B20), ideally "before the research is conducted" (B4). These ideas all linked into *a priori* considerations of what would constitute evidence for or against a hypothesis.

> "[My goal is to] create specific testable research aims, hypotheses (or predicted outcomes) that can be tested statistically but from which a clear conclusion can be made." (B17)

Attendees also made predictions about how they might report their research differently. These predictions mapped onto the importance of recognising and minimising assumptions. E6 set the goal of "acknowledging uncertainty and being careful with language used. Being very specific about results and confidence in results". Two attendees linked these ideas specifically to inductive inferences; the importance of being "up front about the generalisability of [their] study to a wider population" (B21), and considering "the limits of the conclusions that can be made based on the data that has been collected" (B17).

## Reported influence on behaviour

Primary aim 2.b was to learn whether attendees felt that the workshop had influenced their research behaviours. In the 6-month follow-up survey, attendees reflected on this. Overall, the main themes related back to many of the attendees' key lessons and predictions. The first two themes discuss how attendees felt that they had been more ***critical of design*** and ***critical of inferences.*** Theme 3 covers the varied ***indirect influences*** that attendees felt the workshop had. The final theme collates the five responses from those who felt that the workshop had ***no influence*** on their research.

### *Reported Influence 1: Critical of design*

Over a third of attendees felt more critical of various aspects of their research design since the workshop. Several mentioned thinking more about their research questions, how to develop them, and how their methodology will allow them to address them. Relatedly, B17 felt that the workshop "definitely helped" them "to be more forward thinking" about their research design, and encouraged them to "[manage their expectations] of what can be answered in a single research study". Three attendees

mentioned being generally more critical of the rationale behind their research practices. For example, E6 mentioned "rethinking methods and being more aware of inferences we make, and questioning why and how this could be improved". Likewise, B7 reported "being more critical of the why" behind their research. When justifying analysis decisions, B9 had utilised the mind-mapping schematic exercise in their thesis.

> "It is a useful way to justify analyses. Where there are lots of primary, secondary, and sensitivity analyses, displaying them like a mind map as we did in the workshop has been helpful to formalise the decision-making process." (B9)

Two others mentioned paying more attention to the underlying assumptions in their research. B22 felt encouraged "to re-examine (or even become more aware of) the premises and assumptions" underpinning their research questions, hypotheses, and interpretations of the data. Similarly, B19 found themselves "thinking more about tying projects together" so that the underlying premises can be tested first, before then testing hypotheses. Lastly, B13 seemed to benefit from learning about the different inference types, particularly deduction. Now, they concentrate on deducing their "own hypotheses according to existing theories."

"Thanks for this workshop, I learned how to conduct my PhD research in an appropriate way based on deductive reasoning. Now, I am doing data cleaning and analysis after having hypotheses. If not attending this workshop, I may not have a full picture about the process of reasoning and research." (B13)

### *Reported Influence 2: Critical of inferences*

> "Overall, I think the workshop influenced the way I draw conclusions from my own data, and take a more critical approach to the literature that I use." (B2)

The other most frequently reported perception was of having been more critical of inferences, both one's own and others', for example "when reading journals" (B11). This most often manifested in critiquing whether one's inferences are substantiated by the data and design. Hence, multiple attendees mentioned "being less strident with the interpretations" they are able to draw (B19), "not making conclusions further than the results indicate" (B17), and considering "if there are other relevant explanations" (B2). Multiple attendees mentioned this specifically in terms of data analysis. For example, B4 felt "more aware of the limitations associated with statistical inference." Likewise, E1

reported paying closer attention to their statistical methods, "ensuring they are suitable" for their data and research questions. A few concluded that they would use more precise language, "ensuring that the phrasing and conclusion is supported by evidence and reasoning" (E4).

> "In combination with my overall research experience, I now find myself in a better position to thoroughly evaluate the strengths and limitations of the evidence generated by my own research as well as that of others." (B22)

### *Reported Influence 3: Indirect influences*

When designing this study, we felt that it would be difficult to discretely assess long-term influences of the workshop. This is partly because effects of abstract knowledge are likely to be both subtle and diffuse, being interconnected with myriad other factors. Moreover, such a diverse range of individuals will inevitably absorb and act upon the ideas differently, mediated by personal factors. Although we cannot directly quantify the downstream effects of acquiring abstract knowledge, a number of attendees' responses provided some glimpses as to how the workshop might have indirectly influenced their actions.

One trend was that the workshop had prompted attendees to complete further reading, and had "impacted conversations" (B19) since the workshop. For example, B3 said that the workshop influenced them to think more about epistemology. Combined with their research interests, this had become a topic in their research. Likewise, B4 felt encouraged to read more about the subtleties of statistical inference. They elaborated on this at length, exemplifying how they had extended their understanding far beyond the main lessons on the workshop, linking key ideas from the workshop with the nuance of their discipline. Finally, E1 reported that "the workshop cultivated a deeper appreciation for cross-disciplinary research and collaboration". This resulted in "an effort to engage with scholars from other disciplines" (E1).

Other evidence was again linked to a retention of the main lessons, mediated by personal factors. For example, E5 found that it gave them "a lot of confidence in having a philosophical framework" and feels sure that these ideas have given them "clarity in how [they] write and think". Others echoed this idea that it influenced their writing, and/or generally elevated their "proficiency in drawing insightful conclusions" (B14).

It is worth noting that some of these reports were caveated with feeling that the workshop had not changed attendees' methodological practices, but rather, that it had influenced the way they thought. For example, B4 began their response with: "although the workshop did not change my research methods ..."

> "I'm not sure if it has transferred into any changes to my study, but it has definitely made me reflect on how we know what we know." (B21)

> "I believe it has subtly shaped my approach to research. While writing, I wasn't explicitly conscious of employing either an inductive or deductive method. However… filling this questionnaire, I remember that the key lesson I learned from the workshop is that: an awareness of the varying strengths of different arguments. Some arguments hold more weight than others. The robustness of our evidence greatly influences the credibility of our inferences. This notion ties closely to the process of hypothesis generation. Relying on literature closely aligned with our research and drawing from reliable studies enhances the likelihood of formulating dependable hypotheses." (B17)

***Reported Influence 4: No influence***

The final theme consists of the five reports which outright stated that the workshop had not influenced them. These often came with no further context or elaboration, and in three cases these same individuals had written less than average in their answers to the other questions. Two of the five were those who had predicted no influence. This included B6, who had previously been too close to the end of their studies to implement any changes. The only response which provided informative context came from B8, who enjoyed the workshop, but felt too hurried in their studies:

> "The workshop was one of the top two I've done during my time in Bristol. However, it has not influenced my research practices because I've not dedicated time to the implementation of what I've learned into my daily work. Unfortunately I feel I'm constantly in a hurry to do complete my research work and I struggle to find time to reflect on and modify my practices." (B8)

# Discussion

Overall, the test scores and qualitative reports converged on the conclusion that most attendees attained a selection of our intended outcomes (see Figure 3, 'Outcome' panel). The most commonly reported of these was of being more critical, both of research design and inferences as a whole. This was mentioned by around half of attendees, six months on from the workshop. We also used the inference tests to quantitatively approximate changes in attendees' understanding of inference (IDA) across time. The differences in scores before and after the workshop were stark, with the median score increasing from 4 to 13 out of 15. Better still, the median dropped by only 2/15 after 6 months, showing a high level of retention across most attendees.

| Input | Activity | Output | Outcome |
|---|---|---|---|
| *The necessary components to run the workshop are:* | *The core workshop activities are:* | *The immediate aims of the activities are to:* | *Long term we hope that these activities will promote:* |
| * **a facilitator** who understands the content and delivery style.<br><br>* 4+ **researchers** with research projects that they can pictorially represent.<br><br>* necessary **materials** (presentation slides, A3 paper, coloured pens).<br><br>* necessary **facilities** (a room with tables for paired work and a large monitor for the slides). | * a brief presentation on scientific methodology.<br><br>* pictorial representation of individual research processes (attendees create methodological schematics).<br><br>* three discursive seminars on inference (induction, deduction, and abduction – respectively).<br><br>* free discussions. | * convey ideas about scientific methodology and three types of inference.<br><br>* help attendees to visualise their research processes and the inferences made therein.<br><br>* teach the rationale and specific problems of each inference methodology.<br><br>* promote critical thought around research, inference, and epistemology. | * understanding of the logic behind inferential fallibility and the need for epistemic humility.<br><br>* appreciation of manifold assumptions in research; the need to recognise and minimise them.<br><br>* appreciation of such things as: triangulation; methodological transparency; replication; publishing null results.<br><br>* corresponding changes in research behaviours (likely contingent on the individual). |

**Fig. 3** The logic model from our previous study, showing the pathway from the workshop's core components to the key outcomes we hoped to achieve.

Attendees' reflections after the workshop provided additional information about the large variety of ideas they took from the workshop, and potential influences on their research. As in our previous workshops (Clark et al., 2022), many attendees noted that learning about IDA had helped them to understand the logic behind inferential fallibility. While keeping things philosophically simple, the workshop included the problem of induction, and how the underdetermination of theory by data can inhibit falsification (Harding, 1975; Hume, 1779). This framework for thinking about inferences led attendees to various conclusions of their own. Many of these were associated with research design and critical thinking.

Immediately after the workshop, the most common conclusion was the need to scrutinise the assumptions and logic underpinning experiments, creating a solid chain from evidence to conclusion (Scheel et al., 2021). Accordingly, many attendees predicted that they would go on to assess literature more critically, as well as their own conclusions. This included drawing appropriate inferences, substantiated by the data, interrogating one's own assumptions, and being precise with one's language. Some also noted the intention to be precise in the operationalisation of hypotheses, which is crucial for hypothesis tests to have clear outcomes (Scheel, 2022). Several mentioned the need to have a prior understanding of the criteria under which one would accept or reject a hypothesis, thereby facilitating deductive falsification through the specification of prior assumptions (Uygun Tunç & Tunç, 2020). Others noted the intention to set an achievable project scope; trying to do less with a single study, but with greater confidence in the conclusion.

Six months after the workshop, over half of the attendees reported ways in which some of these predictions had manifested. The overarching themes were being more critical**,** both of research design and of inferences in general. To this end, attendees mentioned adopting a range of specific practices, such as being more critical of the literature, assumptions, and/or rationale underpinning a study. They also reported thoroughly justifying the research design, analytical decisions, research questions, and hypotheses *a priori*. Encouragingly, these reports corresponded with what some participants across all of our workshops had speculated might occur, and strongly suggest that the workshop is meeting one of the core outcomes (Popper, 1970; Siegel, 1989).

Moreover, a sub-set of attendees drew discrete, practical implications from the workshop. Some reported more subtle influences, such as inspiring further reading and discussion. A few attendees reported that the workshop was not very beneficial, at least insofar as it did not influence their research. Reasons for this included not feeling that it was relevant to their discipline, being too busy, or being too far along the research process to implement changes. This last point corresponds with an idea which has recurred in all of our studies: that the workshop is more helpful while designing studies, and earlier in one's doctoral training. The other negative responses provided little explanation. Looking at their inference test scores, three of these individuals never

scored over 10. Hence, it could be that they required more resources and other types of scaffolding, or perhaps they did not find the content interesting.

## Methodological considerations

By incorporating distal and direct measures, this study addressed some limitations of our previous research. Nevertheless, we still greatly relied on self-report data, which creates some uncertainty in interpretation. Firstly, we cannot rule out the possibility that some attendees overstated the positives out of kindness. Partly mitigating this concern, some responses demonstrated their understanding through the substance of their responses. Moreover, the convergence of so many attendees on similar conclusions (i.e., the Major Themes) implies that the general trend is real. Even still, our sample was very small compared to the scale of our recruitment. This suggests that sampling bias may have resulted in us having an especially philosophically interested and motivated sample. It is therefore doubtful that the proportion of positive outcomes would be so high, were participation to be mandated.

Another minor issue with self-report was variability in response lengths. A few responses were too short and ambiguous to be used (e.g., “The way I explain possible explanations have been changed” (E2)). Similarly, some of the negative responses provided no further information than “no”. Conversely, some of the positive responses were lengthy, providing multiple ideas for numerous themes. Given how the data are thematically presented, this could create a biased perception of the proportion of attendees who found the workshop beneficial. We aimed to partially mitigate this by our direct quotes and occasional quantification of attendees associated with each idea.

Due to the constraints imposed by measurement error on the Inference test, we have used it only as a descriptive approximation of general trends. Given that multiple choice answers allow guessing, the Inference test cannot be taken as a precise measure. Indeed, multiple respondents reported in the additional comments that they had guessed or were not confident in their responses, and yet had not selected “I have no idea”. Hence, some correct responses could reflect chance accuracy, rather than actual knowledge. While this does not invalidate our conclusions, it leads us to treat these tests as a rough indicator rather than definitive evidence.

## Future directions

Future research should interrogate discrete hypotheses based on the main themes this study identified. Hypotheses could be tested by asking attendees specific questions, before and after the workshop, requiring sufficiently detailed answers as to confirm understanding (e.g., 'Are inductive inferences ever certain? Clearly explain your rationale'). Another step would be to understand the different demographics for which the workshop is more or less beneficial. Disciplinary differences would be interesting to explore. One attendee found the content very 'positivist' and less relevant to their relativist epistemic standpoint. It would therefore be good to understand whether the workshop is relevant to, or needs tailoring for, different epistemological views.

A large amount of feedback from across all of our workshops suggests that they are both sought after and lacking for postgraduate researchers, at least within the South-West of England. It may be worth considering whether these sorts of training workshops could be scaled up and included as a more visible (and potentially requisite) component of doctoral training in empirical research. Additional exploratory angles could also heed attendee feedback to develop similar workshops on other philosophy of science topics such as causation and explanation.

## Conclusion

Overall, learning about inferential logic seems to benefit postgraduate researchers in empirical fields. The extent and nature of outcomes from our workshop differed across individuals. A small subset of attendees reported little influence, while others acknowledged such subtle effects as prompting new lines of thought and inquiry. Most encouragingly, around half of attendees reported being more critical of inferences and research design six months later. Self-criticism, scepticism, and intellectual humility are all fundamental intellectual virtues. Teaching developing researchers about the underlying justification for these virtues is therefore highly important. Nevertheless, such training remains uncommon across the UK. As such, this novel research makes a compelling case for integrating more philosophical training into postgraduate education.

## Ethics statement

Ethics approval for the study was obtained from the School of Psychological Science Research Ethics Committee at the University of Bristol (approval code: 12255). Participants gave informed consent prior to the study using an online consent form.

## Open practices

The open study data are online and should be cited as:
Robbie Clark, James Ladyman, Marcus Munafo (2024): Testing a workshop on Inference and Scientific methodology. https://doi.org/10.5523/bris.1rqtuoxvdpj572uvcpqgufxl0y

The study protocol was preregistered: https://doi.org/10.17605/OSF.IO/QEJ9N. Other study materials and information can be found on our OSF Project page: https://doi.org/10.17605/OSF.IO/Y9ZFH

## Conflicting interests

As designers of the study and workshop, the authors have prior theoretical beliefs about the potential benefits of this workshop and therefore hopes for its success. Otherwise, the authors have no financial or non-financial interests to disclose.

## Funding

This research was supported by a £3000 Southwest Doctoral Training Partnership (SWDTP) Student Led Collaboration Fund Grant.

## Acknowledgements

These workshops could not have happened without a tremendous amount of help organising them. We owe huge thanks to Molly Conisbee, Paul Spencer, Maxine Sims, Sarah Watts, and Alex Maclellan for their help and support. Huge thanks to everyone who attended the workshop, particularly those who travelled from Bath, twice, to attend it. Finally, thank you to Jackie Thompson for your critical feedback on an earlier version of the manuscript.

## Contributions

All co-author roles are listed below using the CRediT: Contributor Roles Taxonomy

| Role | Contributor |
|---|---|
| Conceptualisation: | lead Robbie Clark; supporting Marcus Munafò, James Ladyman |
| Data curation: | Robbie Clark |
| Formal analysis: | lead Robbie Clark; supporting, Margarida Hermida |
| Funding acquisition: | lead Robbie Clark; supporting Marcus Munafò, James Ladyman |
| Investigation: | Robbie Clark, Margarida Hermida |
| Methodology: | lead Robbie Clark; supporting Nicole Russell Pascual, Marcus Munafò, James Ladyman, Margarida Hermida |
| Project administration: | Robbie Clark |
| Resources: | lead Robbie Clark; supporting Nicole Russell Pascual, Marcus Munafò, James Ladyman, Margarida Hermida |
| Software: | *Not applicable* |
| Supervision: | Marcus Munafò, James Ladyman |
| Validation: | James Ladyman, Margarida Hermida, Robbie Clark |
| Visualisation: | Robbie Clark |
| Writing – original draft | Robbie Clark |
| Writing – review and editing | Nicole Russell Pascual, Marcus Munafò, James Ladyman, Margarida Hermida, Robbie Clark |

# References

Albert, K., Brundage, J. S., Sweet, P., & Vandenberghe, F. (2020). Towards a critical realist epistemology? *Journal for the Theory of Social Behaviour*, *50*(3), 357–372. https://doi.org/10.1111/jtsb.12248

Andreoletti, M., & Maugeri, P. (2019). Does medicine need philosophy? *Oral Diseases*, *25*(6), 1419–1422. https://doi.org/10.1111/odi.13143

Blachowicz, J. (2009). How Science Textbooks Treat Scientific Method: A Philosopher's Perspective. *The British Journal for the Philosophy of Science*, *60*(2), 303–344. https://doi.org/10.1093/bjps/axp011

Boniolo, G., & Campaner, R. (2020). Life Sciences for Philosophers and Philosophy for Life Scientists: What Should We Teach? *Biological Theory*, *15*(1), 1–11. https://doi.org/10.1007/s13752-019-00333-7

Clark, R. W. A., Ladyman, J., & Munafo, M. (2022, May 18). *Refining a Workshop Intervention for Scientists on Inference and Scientific Methodology* [Preprint]. https://philsci-archive.pitt.edu/20630/

Derksen, M. (2019). Putting Popper to work. *Theory & Psychology*, *29*(4), 449–465. https://doi.org/10.1177/0959354319838343

Eigenbrode, S. D., O'rourke, M., Wulfhorst, J. D., Althoff, D. M., Goldberg, C. S., Merrill, K., Morse, W., Nielsen-Pincus, M., Stephens, J., Winowiecki, L., & Bosque-Pérez, N. A. (2007). Employing Philosophical Dialogue in Collaborative Science. *BioScience*, *57*(1), 55–64. https://doi.org/10.1641/B570109

Eronen, M. I., & Romeijn, J.-W. (2020). Philosophy of science and the formalization of psychological theory. *Theory & Psychology*, *30*(6), 786–799. https://doi.org/10.1177/0959354320969876

Fjelland, R. (2022). Teaching Philosophy of Science to Science Students: An Alternative Approach. *Studies in Philosophy and Education*, *41*(2), 243–258. https://doi.org/10.1007/s11217-021-09802-8

Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology*, *29*(2), 158–181. https://doi.org/10.1177/0959354319835322

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. In *Advances in Experimental Social Psychology* (Vol. 38, pp. 69–119). Academic Press. https://doi.org/10.1016/S0065-2601(06)38002-1

Grüne-Yanoff, T. (2014). Teaching philosophy of science to scientists: Why, what and how. *European Journal for Philosophy of Science*, *4*(1), 115–134. https://doi.org/10.1007/s13194-013-0078-x

Haig, B. D. (2019). The importance of scientific method for psychological science. *Psychology, Crime & Law*, *25*(6), 527–541. https://doi.org/10.1080/1068316X.2018.1557181

Harding, S. (1975). *Can Theories be Refuted?: Essays on the Duhem-Quine Thesis*. Springer Science & Business Media.

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Hume, D. (1779). An enquiry concerning human understanding. In *Essays and treatises on several subjects, Vol 2: Containing An enquiry concerning human understanding, A dissertation on the passions, An enquiry concerning the principles of morals, and*

*The natural history of religion* (pp. 3–212). Unknown Publisher. https://doi.org/10.1037/11713-001

Johansson, L.-G. (2015). *Philosophy of science for scientists*. Springer Berlin Heidelberg.

Kampourakis, K., & Uller, T. (Eds.). (2020). *Philosophy of Science for Biologists*. Cambridge University Press. https://doi.org/10.1017/9781108648981

Laplane, L., Mantovani, P., Adolphs, R., Chang, H., Mantovani, A., McFall-Ngai, M., Rovelli, C., Sober, E., & Pradeu, T. (2019). Why science needs philosophy. *Proceedings of the National Academy of Sciences*, *116*(10), 3948–3952. https://doi.org/10.1073/pnas.1900357116

Meehl, P. E. (1993). Philosophy of Science: Help or Hindrance? *Psychological Reports*, *72*(3), 707–733. https://doi.org/10.2466/pr0.1993.72.3.707

Morawski, J. (2019). The replication crisis: How might philosophy and theory of psychology be of use? *Journal of Theoretical and Philosophical Psychology*, *39*(4), 218–238. https://doi.org/10.1037/teo0000129

O'Rourke, M., & Crowley, S. J. (2013). Philosophical intervention and cross-disciplinary science: The story of the Toolbox Project. *Synthese*, *190*(11), 1937–1954. https://doi.org/10.1007/s11229-012-0175-y

Popper, K. (1959). *Karl Popper: The Logic of Scientific Discovery*. Routledge.

Popper, K. R. (1970). Normal Science and its Dangers. In A. Musgrave & I. Lakatos (Eds.), *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965* (Vol. 4, pp. 51–58). Cambridge University Press. https://doi.org/10.1017/CBO9781139171434.007

Prather, C. M., Choate, D. M., Michel, M. J., & Crowl, T. A. (2009). Putting the "Ph" back into "PhD": Framing graduate research in a theoretical context. *Frontiers in Ecology*

*and the Environment*, *7*(7), 389–390. https://doi.org/10.1890/1540-9295-7.7.389

Reiter, B. (2013). The Epistemology and Methodology of Exploratory Social Science Research: Crossing Popper with Marcuse. *Government and International Affairs Faculty Publications.*, *99*.

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development*, *31*(1), e2295. https://doi.org/10.1002/icd.2295

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science*, *16*(4), 744–755. https://doi.org/10.1177/1745691620966795

Sheeran, P., & Webb, T. L. (2016). The Intention–Behavior Gap. *Social and Personality Psychology Compass*, *10*(9), 503–518. https://doi.org/10.1111/spc3.12265

Siegel, H. (1989). The Rationality of Science, Critical Thinking, and Science Education. *Synthese*, *80*(1), 9–41.

Spike, J. (1991). The need for teaching philosophy in medical education. *Theoretical Medicine*, *12*(4), 359–365. https://doi.org/10.1007/BF00489894

Uygun Tunç, D., & Tunç, M. N. (2020). *A Falsificationist Treatment of Auxiliary Hypotheses in Social and Behavioral Sciences: Systematic Replications Framework* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/pdm7y

Woodcock, B. A. (2014). "The Scientific Method" as Myth and Ideal. *Science & Education*, *23*(10), 2069–2093. https://doi.org/10.1007/s11191-014-9704-z