

# **A Plea for Minimally Biased Empirical Philosophy**

Andrea Polonioli

a.polonioli@bham.ac.uk

## **Abstract**

Naturalistic philosophers rely on literature search and review in a number of ways and for different purposes. Yet this article shows how processes of literature search and review are likely to be affected by widespread and systematic biases. A solution to this problem is offered here. Whilst the tradition of systematic reviews of literature from scientific disciplines has been neglected in philosophy, systematic reviews are important tools that minimize bias in literature search and review and allow for greater reproducibility and transparency. If naturalistic philosophers wish to reduce bias in their research, they should then supplement their traditional tools for literature search and review by including systematic methodologies.

---

## **1. Introduction**

Naturalism has been a hugely successful research program in contemporary analytic philosophy. Naturalistic philosophers have been following Quine (1952) in challenging the separation of philosophical from scientific research and arguing that philosophy and science are best conceived of as engaged in essentially the same enterprise, pursuing similar goals and deploying similar methods. Yet, naturalism comes in different forms, and a distinction is often made between experimental and empirical philosophy (Prinz 2008; Rose and Danks 2013).

Experimental philosophers have recently adopted methodologies from the behavioural, social and cognitive sciences and engaged in a number of empirical projects to answer philosophically interesting questions (Knobe and Nichols 2008; Alexander 2012; Sytsma and Buckwalter 2016). Empirical philosophers use instead empirical results acquired by professional scientists in their philosophical theorizing. Notably, empirical philosophy has become extremely popular in several corners of philosophical research (Knobe 2015).

Whilst some argue that empirical philosophers should turn themselves into experimental philosophers (Machery and O'Neil 2014), others have expressed a number of qualms about the project of experimental philosophy (Polonioli 2016; Seyedsayamdost 2015;

Carmel 2011). Here, it will be argued that there would indeed be benefits for naturalistic philosophers if they expanded their methodological toolkit. However, the tools discussed here are not those typically advocated by experimental philosophers, but rather the systematic methodologies for literature search and review that are widely employed in the natural, life and health sciences.

More precisely, the paper presents and defends the following claims. First, empirical philosophers do not philosophize in a vacuum and, in fact, rely on literature search and review in a number of ways and for several purposes. Second, biases and cognitive limitations are likely to affect literature search and review in many critical ways. Over the past decades, psychologists have described numerous ways in which judgment formation and information search can be biased, and there are no reasons to doubt that also literature search and review should be biased in important ways, and even in the field of philosophy. Third, scientists have come to widely adopt systematic reviews to minimize bias in the activity of literature search and review, and these tools should also be of wide interest to empirical philosophers. More precisely, systematic research review is a highly structured approach to cumulating knowledge. Progress in knowledge acquisition is the result of the integration of efforts, and literature reviews are vehicles for summarizing research. For systematic reviews, a clear set of rules exists for searching studies and for determining which should be included in or excluded from the analysis. The reproducibility of an experimental result is a fundamental assumption in science and in a similar fashion systematic reviews aim to allow for high reproducibility of conclusions by minimizing bias and maximizing transparency.

In the remainder of this paper these claims will be carefully discussed, and then combined to offer a plea for a type of minimally biased philosophy: if naturalistic philosophers wish to reduce bias in philosophy, as it is here assumed that they should, then they should consider ways to supplement their traditional tools for literature search and review by including systematic reviews.

## **2. Empirical philosophy and literature review**

A hot topic in metaphilosophy concerns how best to describe the methods used by philosophers and their practices. Many of the recent discussions on this topic have focused on whether, to what extent, and how analytic philosophy rests on the use of intuitions (e.g., Cappelen 2012; Andow 2016). Still, it is important not to underestimate the

importance of literature search and review for the philosophical profession, at least in many areas of philosophical investigation.

Empirical philosophy is certainly a case in point here: if we asked what empirical philosophers actually do when they carry out philosophical research, a plausible answer could not help but mention their engagement with literature search and review as an important aspect of it. As Prinz put it, “empirical philosophy works by citation” (2008, 200). Empirical philosophers search for and cite relevant empirical research. But there are different types of activity that require a thorough literature review to be conducted, and different uses of literature review exist.

Empirical philosophers use empirical findings to support their claims and premises. They also often contribute to scientific theorizing by providing novel hypotheses, synthesizing swathes of empirical and theoretical works, and suggesting empirical research. Most typically, empirical philosophers cite neuroscientists and psychologists, but they also call on linguists, evolutionary biologists, roboticists, and anthropologists, among others.

There are plenty of well-known works from empirical philosophers. Yet mentioning a few paradigmatic examples might be helpful here. For instance, philosopher Andy Clark (2013) extensively surveyed scholarly work to argue that brains are fundamentally prediction-error minimizing devices trying to self-generate the sensory streams that are currently arriving from the world. Further, philosophers have also appealed to empirical research to diagnose problems affecting their research community and ameliorate the prospects for philosophical research. In particular, Jennifer Saul (2013) linked the topic of underrepresentation of women in philosophy to empirical research on implicit bias. She stresses that ‘over the last few decades, psychologists have established very clearly that human beings, even those who hold strongly egalitarian ideals, are prone to a range of unconscious biases against members of groups that are stigmatized in certain areas’ (243). Moreover, Jesse Prinz (2004) has offered an up-to-date version of William James’s theory of emotion that he takes to be well supported by the wealth of evidence from empirical psychology and neuroscience that he discusses.

Yet literature review can also constitute an original and valuable piece of research in itself. Review articles come in many trades and different taxonomies are available as well: they can be critical, literature, mapping, generic, qualitative, rapid, scoping, state-of-the-art, systematic, systematized, and umbrella reviews (Grant and Booth 2009). Some

classifications follow instead a pragmatic approach. For instance, the ISI Web of Knowledge Science Citation Index categorizes a paper as a review if it either (i) contains more than 100 references; or (ii) appears in a review journal or the review section of a journal; or (iii) states in the abstract that it is a review.<sup>1</sup>

Review articles are published in philosophy journals too. For example, the journal *Philosophy Compass* publishes original, peer-reviewed survey articles of the most important research from across the entire discipline. In its section on Naturalistic Philosophy it is possible to find entries on topics such as the “Experimental Philosophy of Aesthetics” (Cova et al. 2015) or “Causation: Empirical Trends and Future Directions” (Rose and Danks 2012). Rather than providing a basis for the researchers’ own endeavors, this type of literature review creates a solid starting point for all other members of the community that are interested in a particular topic, and who can refer to these reviews to support some of their claims. Further, if philosophers wish to have impact outside their field too, it becomes advisable to have review articles that present the state of the art on a particular philosophical topic and the main results, so that scientists, policy makers, or any member of a particular profession can more easily become acquainted with the status of a philosophical debate and avoid taking onboard questionable assumptions in their practice. Such review articles can also prove to be useful tools for teaching purposes, providing students with a balanced treatment of a particular topic.

### 3. Heuristics and biases in literature search and review

Since empirical philosophers rely on literature search and review in their philosophical work, they seem to face a number of possible hurdles. The crisis of findings’ reproducibility in psychology (Open Science Collaboration 2015; Pashler and Harris 2012) and other fields (Baker 2016) clearly highlights a number of relevant issues. Notably, if empirical philosophers base their philosophical work on findings that fail to replicate, then their claims might rest on shaky grounds too. Still, there are other, less frequently appreciated, problems. In particular, what empirical philosophers also need to carefully consider is the fact that literature search and review are likely to be constrained by cognitive limitations and vulnerable to biases. Scientists sometimes warn of these risks. For instance, Roy Baumeister and Mark Leary wrote that:

Although literature reviews are less subject than empirical investigations to capitalizing on chance, they are probably more susceptible to the danger of confirmation bias. Many good literature reviews involve seeing a theoretical

---

<sup>1</sup> (<http://wokinfo.com/essays/impact-factor/>).

pattern or principle in multiple spheres of behavior and evidence, and putting together such a paper undoubtedly involves an aggressive search for evidence that fits the hypothesized pattern. (1997, 319)

These considerations are not at all untethered. As it turns out, over the past five decades psychologists have documented many ways in which our reasoning and decision-making can be systematically biased by cognitive, motivational and affective factors (Hastie and Dawes 2010). People have been shown to deploy heuristics that in several contexts lead to a number of biases and, in turn, poor or ill-grounded decision-making.

Philosophers have been hugely interested in this psychological literature (Samuels et al. 2002; Lee et al. 2013; Saul 2013; Polonioli 2014). Still, a point that has not been properly acknowledged in the philosophical literature is that there are also good reasons to think that biases and cognitive limitations might generally have a huge and negative impact on literature search and reviews. After all, literature search and review are just a specific case of information search and related judgment and decision-making, where the latter have been described as riddled with biases. In brief, scholars and scientists typically search for relevant information, process it, and form a number of judgments about the information reviews. Arguably, the same processes that underlie many other instances of information search and assessment are likely to be at work on these occasions as well.

Search for evidence can be biased in critical ways. An obvious and well-known bias is the abovementioned *confirmation bias*, which consists in the “seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand” (Nickerson 1998, 175). When people seek new information, their information search processes are often biased in favor of the seekers’ previously held beliefs, expectations, or desired conclusions. For example, people have been shown to favor information that supports their social stereotypes (Johnston 1996), attitudes (Lundgren and Prislin 1998), expectations in negotiations (Pinkley, Griffith and Northcraft 1995), and self-serving conclusions. Arguably, if a researcher is carrying out a literature review on the positive psychological effects of holding a particular sort of belief, such as believing in free will, confirmation biases might result in her ignoring undesired information and her focusing on what seems to support her previously held belief.

Moreover, consider research on *implicit biases*, where the latter are typically understood as ‘largely unconscious tendencies to automatically associate concepts with one another’ (Saul 2013, 244). Whilst there are still a number of open questions regarding the nature

and frequency of implicit biases, what seems to be supported by copious evidence is that decision makers are often biased by people's gender, ethnic background, and sexual preference, for example when they select job candidates based on CVs, and also by people's appearance (especially weight and height in relation to gender) when they interview job candidates (Steinpreis and Ritzke 1999). If these biases can result in judging members of stigmatized groups negatively, it is not difficult to see how they could impact on the search and assessment of scholarly information. Research published by researchers from particular groups might be just ignored or its value might be slighted.<sup>2</sup>

Further, consider how people have been shown to heavily rely on what is most salient or available information to them in their judgment and decision-making (Tversky and Kahneman 1973). In particular, recent occurrences, because they are cognitively more salient, often lead people to misrepresent the probability of certain types of events. For instance, "the subjective probability of traffic accidents rises temporarily when one sees a car overturned" (11). People are apt to accept one of two competing views merely because they heard evidence supporting the one view more recently (Tversky and Kahneman 1982). In light of this, it is not unlikely that researchers trying to organize findings in support of a particular hypothesis will rely on partial, although quite salient or easily accessible, information.

Finally, researchers attempting to deliver an evenhanded treatment of the literature are supposed to discuss which views are more and less widely shared. But evidence from several studies suggests that social observers tend to perceive a form of *egocentric bias* with respect to the relative commonness of their responses. The best-known example comes from a 1977 study in which Ross, Greene and House asked students to walk around a campus wearing a sandwich board with the word 'repent' on it. Students could agree to wear the board, doing the experimenters a favor, or disagree and participate in a later study. Those who offered to wear the sign (50%) estimated that more than half of their peers would also agree to do so (average estimation 63.5%). Those who declined thought that about a quarter of those asked would accept (average estimation 23.3%). Apparently, students overestimated how similar others' preferences were to their own. These findings might be taken to suggest, for example, that researchers trying to assess the popularity of a particular view or account might produce inaccurate assessments of its popularity (Gilovich 1990).

---

<sup>2</sup> The importance of the topic of discrimination against authors has recently gained importance and attention in philosophy: <http://www.diversityreadinglist.org>.

Overall, confirmation biases, salience and availability biases, as well as false consensus biases are among the many effects and distortions that might impact on literature search and review. There are no compelling reasons to assume that empirical philosophers would be immune from such biases. After all, it seems that everybody is affected by biases to some degree, irrespective of factors like general intelligence or open mindedness (Stanovich and West 2008). Though expertise in specific domains might have positive effects on judgment and decision-making, some findings indicate that experienced professionals often display either roughly the same biases as college students or the same biases at somewhat reduced levels. More precisely, psychological research has demonstrated that a wide variety of biases often affect assessments that many professionals, including physicians, investors, accountants, option traders, real estate agents, engineers, and psychologists, are trained to make (e.g., McNeil et al. 1982; Choi and Pritchard 2003; Bazerman et al. 2002; Fox et al. 1996). More recently, it has been shown that also philosophers tend to commit the very same biases (Schwitzgebel and Cushman 2012; 2015; but see also Livengood et al. 2010). In light of this, it seems highly plausible that when empirical philosophers search for and assess scholarly content, they can be affected by these biases.

But there is also more direct evidence on the impact of biases in scholarly contexts. For instance, Roumbanis (2016) explored the impact of anchoring effects, whereby the first numerical value that an individual encounters tends to influence his or her judgment of what is to be assessed, in the process of peer reviewing research proposals. Moreover, literature on biases in literature selection has suggested that factors such as the reputation and prominence of an author as problematic could influence citation decisions (the so-called “Matthew effect”; Merton 1968). More recently, focus has been placed on testing whether female scholars are systematically denied credit for their work (the so-called “Matilda effect,” Rossiter 1993). In the case of Matilda effects, unconscious automatic processes are assumed to trigger gender bias. In light of this, an author’s gender indirectly influences citation behavior, as it functions as a simple cognitive heuristic to assess scientific quality, challenging the assumption that citations represent payments of intellectual debt and, to that end, would strictly follow the criteria of relevance and quality (cf., e.g., the discussion in Baldi 1998). There is growing evidence on the impact of such biases in scholarly contexts (cf. Lincoln et al. 2012; Potthoff and Zimmermann 2017).

As it turns out, in the context of literature search and review, reliance on heuristics seems quite likely to occur, also because of the increasing and huge amount of research materials published. Herbert Simon claimed that ‘human rational behavior is shaped by a scissors whose blades are the structure of task environments and the computational capabilities of the actor’ (1990, 7). Whilst an important generalization that comes out of efforts to study human information processing is that an individual is a limited information processing system (Newell and Simon, 1972), it is also the case that researchers have to face a rather complex environment in the case of literature search and review, as the relevant literature is huge and the database of papers and noteworthy content is not only massive but also growing fast (De Solla Price 1986; Genova et al. 2016). The volume of research available in most fields is expanding rapidly and there has been an increase in the rate of outputs publication, although it is less clear to what extent this might reflect an increase in knowledge accumulation or rather a tendency to “slice” one research project into too many papers (i.e., “salami publishing”). What is clearer, instead, is that in light of people’s memory and cognitive limitations, recall of huge amounts of relevant information and literature might be far from optimal. Moreover, in light of people’s computational limitations, analyses of huge databases of scholarly content might also be far from optimal.

Clearly, not all of the processes interfering with the reliability of outcomes of literature search and review need to be unconscious. In addition to the abovementioned unconscious biases, empirical philosophers could also be deliberately adopting questionable strategies in processing the literature, resulting for instance in more favorable treatment of close colleagues. Yet unconscious biases are particularly worrisome, as they are hardly detected by the agent. Correction of distorting factors seems harder to occur than one might think, and it has recently been suggested that rationalization of biased choices or judgments might be common in the work of philosophers and scientists as well (Schwitzgebel 2016).

To be sure, heuristics do not just lead to critical biases. In a number of contexts, people have also been shown to deploy heuristics that enable them to gather and assess information effectively (Gigerenzer et al. 2000)<sup>3</sup>. Fast-and-frugal heuristics that take into account only few cues and little of the available information might lead to accurate predictions and estimates in a number of contexts, and such adaptive heuristics might also be at work in the context of literature search and review. This is an important point. Yet,

---

<sup>3</sup> Also, Mercier and Sperber (2011) argue that biases such as confirmation can be adaptive as truth might not be the relevant goal in many conversational contexts. Still, in case of scholarly research, relevant goals seem to be epistemic in nature.

whilst it is important not to overlook this more positive aspect of our decision-making performance (Robins and Craik 1993; Christensen-Szalanski and Beach 1984; Lopes 1991), such acknowledgment should not be read as being at odds with the recognition of the impact of the abovementioned biases and their problematic nature. In addition, it is also unclear to what extent the most plausible heuristics at work in this context could be *successful*. The most prominent cues to be used as proxies for the value of the relevant content would likely be the number of articles' citations, the Journal Impact Factor, or similar metrics. But it can take long time for an article to accumulate citations, especially in the humanities, meaning that it is hard to use such number as a cue to identify relevant published material. Further, whilst it is sometimes argued that social media activity provides useful indications for the future citations, recent research suggests that altmetrics are best conceived of as measuring a different kind of research impact (Erdt et al. 2016). Moreover, the view that Journal Impact Factor is a reliable indicator of the quality of journals has been heavily criticized (e.g., Moustafa 2014) and a number of authors have pointed out that it does not necessarily correlate with several aspects of the journal's quality (Brembs et al. 2013).

Hence, there are good reasons to carefully consider the heuristics and biases that might be operating in the context of literature search and review. Because of these factors, and especially in an environment characterized by growing scholarly production, researchers might find it difficult to process information in a way that serves well and furthers goals of accuracy and truth. But if common research practices do not serve these goals well, then this should look like a worrying situation for researchers. The question arises as to what options are available to remedy the situation described above. Arkes et al. (2006; 2010) referred to work on the inaccuracy of judgment and decision-making (e.g., Dawes et al. 1989) to urge that we should "examine the benefits of a more routinized, mechanical method for evaluating scientific materials such as research presentations at professional conventions or proposals submitted to federal funding agencies" (2006, 430). The next section examines ways to improve the reliability of the survey of research materials in the context of literature reviews by appealing to systematic methodologies.

#### **4. Narrative and systematic methods for literature review**

Biases threaten the reliability of literature search and review. But different types of search and review differ in terms of vulnerability. In philosophy, as well as in many other fields in the humanities, literature search and review is typically "narrative" in character (but see Feltz and Cova 2014), whereby a content expert writes about a particular topic offering a

comprehensive narrative synthesis of previously published materials, usually not describing the methods used and the criteria for inclusion and review of the literature. In brief, the authors of narrative reviews are free to include and exclude research as they like, and they are free in their evaluation of research too. As an example of this approach, consider that in her synthesis of the literature on framing effects, Joanna Demaree-Cotton writes the following: “I have included all relevant studies of which I am aware” (2014, 9). No reference to clear objective criteria of inclusion and methods used in the selection is provided here.

Outside philosophy, however, it has been frequently pointed out that traditional narrative reviews, in spite of some clear benefits, are also prone to error and bias. More precisely, whilst some researchers have suggested ways to reduce bias in literature search and review by improving traditional narrative approaches to review (Baumeister 2013), many others have argued that more rigorous and unbiased types of analysis should be offered to replace traditional narrative reviews (Kitsiou et al. 2013; Templier and Pare 2015). In particular, in the 1970s and early 1980s, scientists started to draw attention to the systematic steps needed to minimize bias and random errors in reviews of research (Light and Smith 1971, Glass 1976, Rosenthal 1978, Jackson 1980, Cooper 1982). In this context, scientists appealed to systematic reviews as a useful tool to navigate through complex bodies of literature and summarize them in a way that reduces bias. In the huge literature on systematic approaches to literature review, the adjective ‘systematic’ is typically contrasted with ‘haphazard study selection procedures’ or even ‘arbitrary study selection procedures’ (Slavin 1986, 6).

Systematic reviews typically address a question formulated in the Participants (Population), Intervention, Comparisons, Outcome (PICO) format. The question identifies a population, the intervention being investigated, a comparison point or points to the intervention, and the outcome of interest (Higgins and Green, 2011, section 5.1.1). For example, a researcher might ask, “for older adults with musculoskeletal disorders, is home-based rehabilitation more effective than inpatient rehabilitation in relation to function, cognition and quality of life?” (Stolee et al. 2011). Equal emphasis on each component of PICO is not necessary. For example, Shumway-Cook et al. (1997) address “the effect of multidimensional exercises on balance, mobility, and fall risk in community-dwelling older adults”, referring to a population, outcome and effect, but not stating a comparison in their question.

On the framework of systematic approaches to review, the criteria used to select studies for inclusion should be clearly stated, alongside the bibliographic databases searched, the dates and periods searched and any constraints, such as language. In addition to describing the search strategy, selection and data collection process, systematic reviews should also clearly discuss their objectives. Notably, questions of systematic reviews might be quite diverse in their nature, and in the medical field they might for instance seek to explore possible harm (e.g., will there be any negative effects?), prognosis (e.g., what is the likely outcome of this problem?), or etiology (e.g., what causes this problem) of a particular effect. Further, systematic reviews may ask broader or narrower questions, and it is generally important to strike a balance between comprehensiveness and precision when developing a search strategy for a question. In brief, increasing the comprehensiveness of a search might result in reducing its precision and retrieving more irrelevant articles. It is also important to distinguish systematic reviews from meta-analyses: only when results are mathematically combined (a process sometimes referred to as pooling), this is referred to as meta-analysis. As the Cochrane Collaboration Handbook points out, in the case of systematic reviews “statistical methods (meta-analysis) may or may not be used to analyse and summarise the results of the included studies” (Higgins and Green 2011).<sup>4</sup>

The appeal of systematic reviews varies from field to field. In the health sciences, systematic reviews have now become a standard and are well understood by all contemporary practitioners. Three decades ago, Mulrow et al. lamented that “medical reviews are often subjective, unsound and inefficient”, and that “strategies for identifying and selecting information are rarely defined” (1987, 485). But things changed significantly in the following years (Bracken et al. 2001). In other fields, instead, systematic reviews are still not mainstream. For example, consider the field of psychology. Narrative methods of review have been dominant and widely taught for long time. Some researchers encouraged authors of reviews “to take a point of view based on theory and offer a point of view that integrates the review” (Sternberg 1991, 3). Currently, some journals, like *Psychological Bulletin*, increasingly are publishing systematic reviews, while others, such as *Trends in Cognitive Sciences*, still publish narrative reviews only, as does the *Annual Review of Psychology* too.

---

<sup>4</sup> Hence, objections to systematic reviews which critique their application of statistical methods for evidence aggregation will not be discussed here (e.g., Stegenga 2011), as the use of meta-analytic tools is not essential to the argument made here.

Further, different protocols for systematic reviews are typically followed in different disciplines. Methods in environmental sciences are outlined by the Collaboration for Environmental Evidence, in social sciences by the Campbell Collaboration and in medicine by the Cochrane Collaboration. Still, in all these different fields, systematic review methods often use peer-reviewed and published protocols to lay out the methods for a review, and then searches for studies, articles screening for relevance and quality, and data extraction and synthesis are typically undertaken according to a predetermined strategy. Importantly, irrespective of the specific layout of the systematic review, systematic reviews seem to differ from traditional narrative reviews by virtue of being:

*Systematic/organized:* Systematic reviews are conducted according to a system or method that is designed in relation to and specifically to address the question the review is setting out to answer.

*Transparent/explicit:* The method used in the review is explicitly stated.

*Replicable/updatable:* As with many forms of primary research, the method and the way it is reported should be sufficiently detailed and clear such that other researchers can repeat the review, repeat it with modifications or update it.

*Synthesize/summarize:* Systematic reviews pull together in a structured and organized way the results of the review in order to summarize the evidence relating to the review question.

As it turns out, whilst systematic reviews might follow different protocols and focus on somewhat different questions, they are nevertheless supposed to incorporate a set of key principles of scientific methodology and depart from traditional narrative approaches to literature review. For instance, Cooper nicely expresses the spirit behind the systematic review movement in the introduction to his book, *Synthesizing Research*:

The approach to research synthesis presented in this book represents a significant departure from how reviews had been conducted just 20 years ago. Instead of a subjective, narrative approach, this book presents an objective systematic approach. Here, the reader will learn how to carry out an integration of research according to scientific principles and rules. The intended result is a research synthesis that can be replicated by others, can create consensus among scholars, and can focus debate in a constructive fashion. (Cooper 1998, xi)

On one hand, it is obvious that systematic methodologies do not completely eliminate subjectivity from the process of search review. After all, when the researchers try to operationalize a research question, they are still called to make some decisions. For instance, one still needs to define what counts as older population. On the other hand, systematic literature reviews are undertaken according to strict guidelines to minimize

subjectivity, maximize transparency and replicability, and are supposed to provide a highly reliable review of evidence pertaining to a specific topic. The scientific method has the invaluable benefit of affording a systematic and unbiased investigation, and systematic reviews apply it to the practice of literature search of review. Systematic methods aim at making literature search and review objective: the reasoning is that subjectivity is a source of bias, and one that can and must be minimized by developing a clear protocol, making all the steps and the criteria explicit, following these steps and documenting all the relevant activity. By so doing, one is likely to maximize the chances of producing valid conclusions, and also makes the review replicable. It should be noted, here, that the use of the concept “objective” is eminently complicated, as also recent philosophical (e.g., Douglas 2004) and historical (e.g., Daston and Galison, 2010) analyses demonstrate. But in general the objectivity of results is thought to be a consequence of the method being objective.

Another important aspect to highlight is that systematic reviews are typically conducted in a team. Ensuring that tasks such as selection of studies for inclusion and data extraction can be performed by at least two people independently may increase the chance that errors and biases be detected. Importantly, at least when considering biases like confirmation, there is evidence suggesting that groups perform better than single individuals. More precisely, although groups in some contexts do fall prey to some of the errors made by the single individual, for many important biases such as confirmation, groups outperform individuals (Maciekoskj et al. 2013), and there is not evidence that single individuals outperform groups.

Still, some have expressed qualms about using these methodologies. Whilst it is frequently argued that the “the use of explicit, systematic methods in reviews limits bias (systematic errors) and reduces chance effects, thus providing more reliable results upon which to draw conclusions and make decisions (Antman 1992, Oxman 1993)” (Higgins and Green 2011), one possible objection is that peer review would in any case wash out researchers’ biases, eventually leading to reliable surveys of the literature. This, however, seems to rely on too romantic a view of peer review, and one with several problems. Whilst peer review typically brings a measure of rigor and trust to scholarly communication, the reliability of peer-review is far from optimal, and several biases in peer review have also been identified (Lee et al. 2013; Lee 2015; Shalvi et al. 2010). There are well known cases of so-called Mendel syndrome, mentioned after Gregor Mendel, whose discoveries in plant genetics were so unprecedented that it took thirty-four years for the scientific community

to catch up to it (Van Raan 2004; Gorry and Ragouet 2016). Moreover, obvious failures of peer review have also been clearly documented (Hawkes 2013). Even more importantly, the very fact that systematic reviews and narrative reviews have, at least in some cases, been shown to deliver results that are at odds suggests that pointing to peer review as a silver bullet might be an unwarranted move (e.g., Cipriani and Geddes 2003; Filippin and Crosetto 2016). For instance, De Dreu and Weingart (2003) show in a systematic review that the relationship between task conflict, team performance and team satisfaction is largely negative even though both academic papers and textbooks regularly report that task conflict has a generally positive effect. Moreover, consider also that whilst it is typically claimed in the literature that reasoning errors have severe impact in terms of wealth<sup>5</sup>, recent systematic literature searches show limited evidence that these errors would cause actual harm in terms of less wealth (Arkes et al. 2015).

Other critics of systematic reviews argue that a major threat to systematic reviews is dissemination bias, often referred to as publication bias, and which describes the selective publication and dissemination of results. In this situation, published studies constitute a biased sample leading to spurious conclusions. Published research can then be shaped by file-drawer effects (Rosenthal 1979). Again, this does not read as a knockdown objection, and at least for two reasons. First, whilst systematic reviews might not solve these problems, narrative reviews do not seem to be obviously better positioned at dealing with them. An argument would be needed to support this claim. Second, it actually seems that the methods of systematic reviews can also be applied to the grey literature as well. For instance, these methods can be applied to doctoral dissertations as well as conference proceedings. In fact, Cochrane systematic reviews use very comprehensive search strategies and include both published and unpublished studies. Overall, publication bias is clearly an important problem that the research community and research gatekeepers need to address, but also one which is orthogonal to the debate over the merits of narrative and systematic reviews.

Where does all this lead us? The thrust of the section is not to argue that narrative reviews should be replaced by systematic reviews *tout court*. Arguably, narrative reviews have important benefits, including a broad overview of relevant information tempered by years of knowledge from an experienced author. It is also true that the narrative thread can be

---

<sup>5</sup> For example, Milkman et al. write that “errors induced by biases in judgment lead decision makers to undersave for retirement [...]. Given the massive costs that can result from suboptimal decision-making, it is critical for our field to focus increased effort on improving our knowledge about strategies that can lead to better decisions” (2009, 379).

lost in the strict rules of systematic review, which might hinder the piece's readability. As it turns out, the benefits of appealing to a particular approach might depend on the specific situation. For instance, Paluck and Green point out that in some cases:

A narrative rather than a meta-analytic review suits this purpose, in the interest of presenting a richer description of the prejudice-reduction literature. Moreover, the methods, interventions, and dependent variables are so diverse that meta-analysis is potentially meaningless (Baumeister and Leary 1997; see also Hafer and Begue 2005), especially given that many of the research designs used in this literature are prone to bias, rendering their findings unsuitable for meta-analysis. (2008, 342)

The point that this section seeks to drive home is that systematic approaches to review should at least be seen as important complements to traditional methods of literature search and review, as the former are better placed at reducing bias and increasing reproducibility. Although there is often some tension between the users of the two methods, and some experts who favour systematic analyses disdain narrative approaches as obsolete, both methods could actually have a valuable place in science.

### **5. Systematic reviews for empirical philosophers**

It has been shown that systematic reviews are methodologies widely employed in natural, life and health sciences, and that they offer important tools to minimize bias and increase transparency and reproducibility. Since empirical philosophers are also likely to be affected by critical biases in the process of literature search and review, it seems tempting to conclude that they should also carefully consider these tools. This might seem too ambitious a goal to be taken seriously, but the adoption of systematic tools would fit just nicely with some recent trends in philosophy to employ new tools to improve their appraisal of scholarly literature. A number of philosophers have already adopted conceptual and technical tools from digital humanities (e.g., Andow 2015; Betti and Van Der Berg 2016) and bibliometrics (Knobe 2015; Machery and Cohen 2012; Wray 2010) to improve the reliability of their assessments of the literature, yet the value of systematic reviews has been surprisingly neglected. As it turns out, embracing systematic methodologies would just be a natural extension of this laudable trend. It should also be mentioned that systematic methodologies for review have already been introduced in some areas of the humanities, speaking in favor of the feasibility of the suggestions offered in this paper. Consider some recent trends in the field of bioethics. It has been argued there that bioethics would benefit from “the corrective effect of review articles that are designed to reduce the potential for bias in the conclusions drawn” (McCullough et al.

2007, 66). More precisely, in bioethics systematic methods have been applied to both empirical (Strech et al. 2007) and reason-based (Stretch and Sofaer 2012) literature.

Here, a possible objection is that empirical philosophers should let scientists review empirical literature systematically, and simply rely on the results of their systematic analyses. Again, there are some problems with this claim, though. First, this rejoinder would still acknowledge that philosophers should pay close attention to results obtained via systematic methods, and just rejects the claim that philosophers should actively deploy such methods. In other words, philosophers should still appreciate that systematic reviews would constitute a privileged source of evidence to use in their philosophical work. Second, it also seems that by engaging directly with systematic methodologies, philosophers would disengage in part from the agenda of particular sciences, and would be able to contribute to redirect it towards topics that are especially important to philosophers. More precisely, there are plenty of topics that are of great interest to philosophers and that could be target of systematic methodologies. Imagine an empirical philosopher interested in exploring whether conscious decision-making leads to better outcomes than non-conscious one: being able to properly apply systematic tools would greatly help her in her philosophical work. Instead, by relying passively on the syntheses provided by scientists, philosophers would risk failing to adequately answer questions that have been traditionally central in their disciplines. Further, experimental philosophers have also produced empirical work themselves, and they should be able to review their findings systematically, instead of expecting non-philosophers to accomplish the task for them.

The suggested shift towards systematic methodologies would offer important benefits for empirical philosophers. For instance, consider a paper recently published in *Philosophical Psychology*, where Smithdeal (2016) reviewed empirical evidence allegedly suggesting that belief in free will is beneficial. His review was narrative in character: he appealed to some sources suggesting that belief in free will is a valuable support for prosocial behavior (Vohs and Schooler 2008; Baumeister et al. 2009), and then criticised some studies pointing to possible detrimental effects of belief in free will (Nadelhoffer and Tocchetto 2013). However, narrative reviews, as mentioned above, are hardly updatable and reproducible. It turns out that, besides the referenced sources suggesting that disbelief in freewill is linked with a decreased willingness to help others, some other studies examining how free will beliefs influence true self-knowledge (Seto and Hicks 2016) seem highly relevant to the research question addressed by Smithdeal's paper. It is hard to tell,

however, if the author was aware of the existence of this study, and in case he was, why or how such sources were excluded. Further, relevant studies have been published after the publication of Smithdeal's paper. Notably, Caspar et al. (2017) provided further evidence that disbelief in free will had a positive impact on the morality of decisions toward others. Would this paper feature in Smithdeal's analysis, should he be writing this paper today? It is hard to answer this question without an idea of any criteria of inclusion and exclusion. Instead, a systematic review would here give readers clear answers as to why some seemingly relevant papers were excluded as well as indications on the extent of the literature covered, in terms of period, databases and journals considered. Choices by authors of systematic reviews can obviously be questioned, but the transparency of the methodology, alongside stating the limits of the time window considered and the relevant inclusion criteria, all seem to constitute important virtues that would improve the prospects of empirical philosophy.

Further, among the reasons why empirical philosophers should appreciate the value of such systematic tools is also that systematic reviews have already delivered results at odds with a few theses widely accepted in empirical philosophy. As a case in point, consider that philosophers have widely embraced the assumption that emotions and affective processes cause moral judgment. For instance, philosopher Joshua May writes that "scientists have apparently amassed converging evidence that emotions play a substantial role in the production of most, if not all, of our moral judgments" (2014, 125). But Landy and Goodwin's (2015; see, however, also Schnall et al. 2015) offer a systematic review in which they consider both published and unpublished studies, eventually arguing "against some claims about the role of affect in moral judgments" (518). In addition, while it is typically accepted in the philosophical literature that "psychological life occurs through nonconscious means (Bargh and Chartrand 1999)" (Fridland 2015, 4), recent systematic reviews cast doubt on much research on social priming on which these claims often are based (Kuehberger et al. 2016). Exploring whether conclusions typically reached via narrative literature reviews would stand in light of the application of systematic tools seems an important task for empirical philosophers, and one in line with the mission of critically appraising scientific projects. After all, many effects that have attracted the attention of philosophers, from power posing (Carney et al. 2010) to the deliberation without attention effect (Dijksterhuis et al. 2006), have been followed by plenty of exchanges in the scientific literature, featuring failed replications and contradictory findings.

As it turns out, many questions that have attracted the attention of philosophers, such as whether psychopathy increases propensity to engage in immoral behavior compared to subjects without it, or whether intelligence increases the likelihood of achieving good life outcomes in healthy subjects, can be properly tackled using the framework of systematic reviews. Clearly, several possible complications should not be overlooked or dismissed too quickly.

One possible objection here is that systematic reviews would be too strict for the purpose of philosophical analysis. I take this practical hurdle to be an important one. For instance, some empirically minded philosophers might be looking for interviews rather than experimental studies to inform their analyses, and yet the PICO model described above does not fit questions that deal with interview research. In brief, the PICO model could be applied if philosophers used experimental methods that use comparisons and focused on specific outcomes, but would fit less nicely interview research. Still, this is not a knockdown objection. For instance, Strech et al. (2008) tried to offer a model of systematic reviews for empirical bioethics, which is a field that heavily relies on interviews studies. Strech et al. advocated a model based on Methodology, Issue, Participants for review questions over the PICO format (2008, 473). He proposed a model based on 7-steps for systematic reviews of empirical bioethics: (1) careful definition of review question; (2) selection of relevant databases; (3) application of ancillary search strategies; (4) development of search algorithms; (5) relevance assessment of the retrieved references; (6) quality assessment of included studies; and (7) data analysis and presentation. What this suggests is that it might indeed be possible to pursue systematic reviews even in cases where the PICO model appears to be too narrow.

Needless to say, other changes may be required in order to adapt the systematic approach to the field of philosophy, which is also interested in subject areas that are characterized by less rigid terminology than that used in biomedicine (making comprehensive searching more challenging). More precisely, traditional systematic reviews prefer databases that include a wide range of publications of clinical trials, such as MEDLINE and EMBASE, and usually deal with issues (such as specific diseases and interventions) and study designs (such as randomized controlled trials) that correspond well to the controlled vocabulary of such databases. Research of relevance to empirical philosophers is often indexed in databases other than MEDLINE and EMBASE, and because of the heterogeneity of the search terms that are relevant for empirical philosophy and are used by different databases, search algorithms for systematic reviews by empirical philosophers have to be

adapted to the databases' vocabulary to enhance the sensitivity and specificity of literature searches. Defining the best search strategies and most suitable databases are important practical aspects of the suggested methodological shift, and they certainly should not (and do not) only concern philosophers. In several disciplines there are ongoing discussions about the most suitable databases for systematic reviews (Bramer et al. 2013; 2016; Sampson et al. 2016; Gehanno et al. 2013; Martin-Martin et al. 2017; Vassar et al. 2015; 2017). Most likely, the ideal combination of databases that empirical philosophers should use will depend on the specific questions they are trying to address. For instance, empirical papers published in philosophy journals and articles published in social sciences journals might not be displayed in databases such as *PubMed* and *Scopus*. But it seems that the systematic approach has already proven flexible enough to be applied to a broad range of fields, and quite recently the approach has been introduced even in non-empirical, reason-based, literature (Stretch and Sofaer 2012). This suggests that there do not seem to be solid grounds for skepticism about the prospects of introducing systematic methodologies in empirical philosophy.

There are for sure other important questions that empirical philosophers will face and need to address. For instance, how will a systematic review include or exclude studies based on quality? Would these choices be based on the studies' sample size or on reported *p* values? These are important and burning questions; for instance, in contemporary debates in psychology it has been stressed the need to improve traditional statistical standards (Johnson et al. 2017). What matters here is that whilst there can obviously be disagreements about the choices made by authors with regard to these issues, by making their criteria explicit authors will also make themselves more easily accountable for their choices and favor greater methodological transparency and awareness of the importance of carefully assessing the quality of the studies rather than accepting empirical conclusions at face value. Rather than impossible challenges, these seem to be important opportunities for the development of greater and more mature engagement with the empirical literature within the philosophical community.

The upshot of this section is that there are no obvious reasons why empirical philosophers should not adopt systematic approaches to literature search and review. Naturalist philosophers should be highly interested in systematic reviews, as this is a key tool used in many scientific disciplines. Further, empirically minded philosophers and philosophers of science may also notice that the suggestion to expand their methodological toolkit to include systematic approaches fits quite nicely with some other proposals already put

forward by philosophers to minimize bias and error and improve our understanding of the literature.

## **6. Conclusion**

In summary, this essay has attempted to highlight and discuss some overlooked problems with the methodology of naturalistic philosophy and to point to solutions that might help overcome them. More precisely, it has firstly been stressed that empirical philosophers have not adequately reflected on the obvious and yet important fact that literature search and review are likely to be affected by widespread and systematic biases. This has been shown to be highly worrying, as empirical philosophers do not typically philosophize in a vacuum, and in fact seem to rely on literature search and review in a number of ways and for several purposes. The suggested solution to tackle these problems comes from scientific disciplines. Whilst naturalistic philosophers have recently started to look at methods and tools from the sciences to expand their methodological toolkit and offer philosophy better chances of accomplishing its goals, it turns out that the tradition of systematic reviews of literature from scientific disciplines has been unduly neglected. But systematic reviews are important tools that minimize bias and allow for reproducibility and transparency. The upshot of this investigation is that, if naturalistic philosophers wish to reduce bias in philosophy, as it is here assumed that they should, they should consider ways to supplement their traditional tools for literature search and review by including systematic reviews.

## **Acknowledgements:**

The author acknowledges the support of the European Research Council under the ERC Consolidator Grant agreement number 616358 for a project called Pragmatic and Epistemic Role of Factually Erroneous Cognitions and Thoughts (PERFECT) and of the Italian Academy for Advanced Studies in America at the Columbia University.

## References

- Alexander, J. (2012). *Experimental philosophy: An introduction*. Polity.
- Andow, J. (2016). Thin, fine and with sensitivity: a metamethodology of intuitions. *Review of Philosophy and Psychology*, 7(1), 105-125.
- Andow, J. (2015). How “Intuition” Exploded. *Metaphilosophy*, 46(2), 189-212.
- Arkes, H. R., Gigerenzer, G., and Hertwig, R. How Bad Is Incoherence? (2016). *Decision*, 3(1), 20-39.
- Arkes, H. R., Gonzalez-Vallejo, C., Bonham, A. J., Kung, Y-H., and Bailey, N. (2010). Assessing the merits and faults of holistic and disaggregated judgments. *Journal of Behavioral Decision Making*, 23, 250-270.
- Arkes, H. R., Shaffer, V. A., & Dawes, R. M. (2006). Comparing holistic and disaggregated ratings in the evaluation of scientific presentations. *Journal of Behavioral Decision Making*, 19, 429-439.
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452-454.
- Baldi, S. (1998). Normative versus social constructivist processes in the allocation of citations: A network-analytic model. *American Sociological Review*, 829-846.
- Baumeister, R. F., Masicampo, E. J., and DeWall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, 35, 260–268.
- Bazerman, M.H., G. Loewenstein, and D.A. Moore. (2002). Why good accountants do bad audits. *Harvard Business Review* 80(11): 96–103.
- Betti, A., and van den Berg, H. (2016). Towards a Computational History of Ideas. In L. Wieneke, C. Jones, M. Düring, F. Armaselu, and R. Leboutte (Eds.), *Proceedings of the Third Conference on Digital Humanities in Luxembourg with a Special Focus on Reading Historical Sources in the Digital Age*. CEUR Workshop Proceedings, CEUR-WS.org.
- Baumeister, R. F. (2013). Writing a literature review. In *The Portable Mentor* (pp. 119-132). Springer New York.
- Baumeister, R. F., and Leary, M. R. (1997). Writing narrative literature reviews. *Review of general psychology*, 1(3), 311.
- Bracken, M. B. (2001). Commentary: toward systematic reviews in epidemiology. *International journal of epidemiology*, 30(5), 954-957.
- Bramer, W. M., Giustini, D., and Kramer, B. M. (2016). Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. *Systematic reviews*, 5(1), 39.

- Bramer, W. M., Giustini, D., Kramer, B. M., and Anderson, P. F. (2013). The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. *Systematic reviews*, 2(1), 115.
- Brembs, B., Button, K., and Munafo, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience*, 7, 291.
- Callaway, E. (2015). Oldest stone tools raise questions about their creators: the 3.3-million-year-old implements predate the first members of the Homo genus. *Nature*, 520(7548), 421-422.
- Cappelen, H. (2012). *Philosophy without intuitions*. Oxford University Press.
- Carmel, D. (2011). Experimental philosophy: Surveys alone won't fly. *Science*, 332(6035), 1262-1262.
- Carney, D. R., Cuddy, A. J., and Yap, A. J. (2010). Power posing brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, 21(10), 1363-1368.
- Caspar, E. A., Vuillaume, L., Magalhaes De Saldanha Da Gama, P. A., and Cleeremans, A. (2017). The influence of (dis) belief in free will on immoral behaviour. *Frontiers in Psychology*, 8, 20.
- Choi, S. J., and A. C. Pritchard. 2003. Behavioral economics and the SEC. *Stanford Law Review*, 56: 1-73.
- Christensen-Szalanski, J. J., and Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist*, 39(1), 75.
- Cipriani, A., and Geddes, J. (2003). Comparison of systematic and narrative reviews: the example of the atypical antipsychotics. *Epidemiology and psychiatric sciences*, 12(3), 146-153.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181-204.
- Hawkes N. (2013). Spoof research paper is accepted by 157 journals. *BMJ*; 347: f5975.
- Higgins, J. P., and Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 4). John Wiley & Sons.
- Cooper, H. M. (1998). *Synthesizing research: A guide for literature reviews* (Vol. 2). Sage.
- Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of educational research*, 52(2), 291-302.
- Cova, F., Garcia, A., and Liao, S. Y. (2015). Experimental Philosophy of Aesthetics. *Philosophy Compass*, 10(12), 927-939.
- Daston, Lorraine J. and Galison, P. (2010). *Objectivity*. Zone Books.
- Dawes, R. M., Faust, D., and Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668-1674.

- De Dreu, C. K., and Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis. *Journal of applied Psychology, 88*(4), 741.
- Demaree-Cotton, J. (2016). Do framing effects make moral intuitions unreliable?. *Philosophical Psychology, 29*(1), 1-22.
- de Solla Price, D. J. (1986). *Little science, big science... and beyond* (p. 301). New York: Columbia University Press.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F., and Van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*(5763), 1005-1007.
- Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese, 138*(3), 453-473.
- Erdt, M., Nagarajan, A., Sin, S. C. J., and Theng, Y. L. (2016). Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics, 109*(2), 1117-1166.
- Feltz, A., and Cova, F. (2014). Moral responsibility and free will: A meta-analysis. *Consciousness and cognition, 30*, 234-246.
- Fox, C.R., B.A. Rogers, and A. Tversky. (1996). Options traders exhibit subadditive decision weights. *Journal of Risk and Uncertainty, 13*(1): 5-17.
- Filippin, A., and Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Management Science, 62*(11), 3138-3160.
- Fridland, E. (2015). Automatically minded. *Synthese, 1-27*.
- Gehanno, J. F., Rollin, L., and Darmoni, S. (2013). Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC medical informatics and decision making, 13*(1), 7.
- Génova, G., Astudillo, H., and Fraga, A. (2016). The scientometric bubble considered harmful. *Science and engineering ethics, 22*(1), 227-235.
- Gilovich, T. (1990). Differential construal and the false consensus effect. *Journal of personality and social psychology, 59*(4), 623.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher, 5*(10), 3-8.
- Gorry, P., and Ragouet, P. (2016). "Sleeping beauty" and her restless sleep: Charles Dotter and the birth of interventional radiology. *Scientometrics, 107*(2), 773-784.
- Grant, M. J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91-108.
- Hastie, R., and Dawes, R. (2010). *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*. Thousand Oaks: Sage.

- Higgins, J. P., and Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 4). John Wiley & Sons.
- Jackson, G. B. (1980). Methods for integrative reviews. *Review of educational research*, 50(3), 438-460.
- Johnson, V. E., Payne, R. D., Wang, T., Asher, A., and Mandal, S. (2017). On the reproducibility of psychological science. *Journal of the American Statistical Association*, 112(517), 1-10.
- Johnston, L. (1996). Resisting change: Information-seeking and stereotype change. *European Journal of Social Psychology*, 26, 799-825.
- Kahneman, D., and Tversky, A. (1982). The psychology of preferences. *Scientific American*.
- S. Kitsiou, G. Pare, M. Jaana. (2013). Systematic reviews and meta-analyses of home telemonitoring interventions for patients with chronic diseases: a critical assessment of their methodological quality, *Journal of Medical Internet Research*, 15 (7), e150.
- Knobe, J. (2015). Philosophers are doing something different now: Quantitative data. *Cognition*, 135, 36-38.
- Knobe, J., and Nichols, S. (Eds.). (2008). *Experimental philosophy* (Vol. 2). Oxford University Press.
- Kuehberger, A., Scherndl, T., Ludwig, B., and Simon, D. M. (2016). Comparative Evaluation of Narrative Reviews and Meta-Analyses. *Zeitschrift für Psychologie*.
- Landy, J. F., and Goodwin, G. P. (2015). Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. *Perspectives on Psychological Science*, 10(4), 518-536.
- Lee, C. J. (2015). Commensuration Bias in Peer Review. *Philosophy of Science*, 82(5), 1272-1283.
- Lee, C. J., Sugimoto, C. R., Zhang, G., and Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2-17.
- Light, R., and Smith, P. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard educational review*, 41(4), 429-471.
- Lincoln, A. E., Pincus, S., Koster, J. B., & Leboy, P. S. (2012). The Matilda Effect in science: Awards and prizes in the US, 1990s and 2000s. *Social studies of science*, 42(2), 307-320.
- Livengood, J., Sytsma, J., Feltz, A., Scheines, R., and Machery, E. (2010). Philosophical temperament. *Philosophical Psychology*, 23, 313-330.
- Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology*, 1(1), 65-82.
- Lundgren, S. R., and Prislin, R. (1998). Motivated cognitive processing and attitude change. *Personality and Social Psychology Bulletin*, 24, 715— 726.

- Machery, E., and Cohen, K. (2012). An Evidence-Based Study of the Evolutionary Behavioral Sciences. *British journal for the philosophy of science*, 63(1), 177-226.
- Machery, E., and O'Neill, E. (2014). *Current controversies in experimental philosophy*. Routledge.
- Maciejovsky, B., Sutter, M., Budescu, D. V. and Bernau, P. (2013) Teams make you smarter: Learning and knowledge transfer in auctions and markets by teams and individuals. *Management Science*. 59:1255–70.
- Martin-Martin, A., Orduna-Malea, E., Harzing, A. W., and López-Cózar, E. D. (2017). Can we use Google Scholar to identify highly-cited documents?. *Journal of Informetrics*, 11(1), 152-163.
- May, J. (2014). Does disgust influence moral judgment? *Australasian Journal of Philosophy*, 92(1), 125-141.
- McCullough, L. B., Coverdale, J. H., and Chervenak, F. A. (2007). Constructing a systematic review for argument-based clinical ethics literature: the example of concealed medications. *Journal of Medicine and Philosophy*, 32(1), 65-76.
- McNeil, B.J., S.G. Pauker, H.C. Sox Jr., and A. Tversky. (1982). On the elicitation of preferences for alternative therapies. *The New England Journal of Medicine*, 306(21): 1259-1262.
- Mercier, H., and Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and brain sciences*, 34(02), 57-74.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56-63.
- Milkman, K. L., Chugh, D., and Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science*, 4: 379-383.
- Moustafa, K. (2014). The disaster of the impact factor. *Science and Engineering Ethics*, 21(1), 139-142.
- Mulrow, C. D. (1987). The medical review article: state of the science. *Annals of Internal Medicine*, 106(3), 485-488.
- Nadelhoffer, T., and Tocchetto, D. G. (2013). The potential dark side of believing in free will (and related concepts). In G. Caruso (Ed.), *Exploring the illusion of free will and moral responsibility* (121-141). Lanham, MD: Lexington Books.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Paluck, E. L., and Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual review of psychology*, 60, 339-367.

- Pashler, H., and Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531-536.
- Pinkley, R. L., Griffith, T. L., and Northcraft, G. B. (1995). "Fixed pie" a la mode: Information availability, information processing, and the negotiation of suboptimal agreements. *Organizational Behavior and Human Decision Processes*, 62, 101-112.
- Polonioli, A. (2016). New Issues for New Methods: Ethical and Editorial Challenges for an Experimental Philosophy. *Science and Engineering Ethics*, 1-26.
- Polonioli, A. (2014). Blame It on the Norm: The Challenge from "Adaptive Rationality". *Philosophy of the Social Sciences*, 44(2), 131-150.
- Potthoff, M., and Zimmermann, F. Is there a gender-based fragmentation of communication science? An investigation of the reasons for the apparent gender homophily in citations. *Scientometrics*, 1-17.
- Prinz, J. (2008) Empirical Philosophy and Experimental Philosophy. In: Knobe J, Nichols S (eds) *Experimental Philosophy*. Oxford University Press, Oxford, 189-208.
- Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.
- Quine, W.O. (1952), 'Two Dogmas of Empiricism', in *From a Logical Point of View*, ed. W.O. Quine, Cambridge, MA: Harvard University Press, pp. 20-46.
- Robins, R. W., and Craik, K. H. (1993). Is there a citation bias in the judgment and decision literature?. *Organizational Behavior and Human Decision Processes*, 54(2), 225-244.
- Rose, D., and Danks, D. (2013). In defense of a broad conception of experimental philosophy. *Metaphilosophy*, 44(4), 512-532.
- Rose, D., and Danks, D. (2012). Causation: Empirical Trends and Future Directions. *Philosophy Compass*, 7(9), 643-653.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, 86(3), 638.
- Rosenthal, R. (1978). Combining results of independent studies. *Psychological bulletin*, 85(1), 185-193.
- Rossiter, M. W. (1993). The Matthew Matilda effect in science. *Social studies of science*, 23(2), 325-341.
- Roumbanis, L. (2016). Academic judgments under uncertainty: A study of collective anchoring effects in Swedish Research Council panel groups. *Social Studies of Science*, 0306312716659789.
- Bramer, W. M., Giustini, D., Kramer, B. M., and Anderson, P. F. (2013). The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. *Systematic reviews*, 2(1), 115.

- Samuels, R, Stich, S., and Bishop, M. (2002). "Ending the Rationality Wars." In *Common Sense, Reasoning, and Rationality*, edited by R. Elio, 236-68. Oxford: Oxford University Press.
- Saul, J. (2013). Implicit bias, stereotype threat, and women in philosophy. *Women in philosophy: What needs to change*, 39-60.
- Schnall, S., Haidt, J., Clore, G. L., and Jordan, A. H. (2015). Landy and Goodwin (2015) confirmed most of our findings then drew the wrong conclusions. *Perspectives on Psychological Science*, 10(4), 537-538.
- Schwitzgebel, E., and Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, 27(2), 135-153.
- Schwitzgebel, E., and Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition*, 141, 127-137.
- Schwitzgebel, E., and Ellis, J. (2016). Rationalization in Moral and Philosophical Thought. In *Moral Inferences*, ed. J. F. Bonnefon and B. Tremoliere (Psychology Press).
- Seyedsayamdost, H. (2015). On gender and philosophical intuition: Failure of replication and other negative results. *Philosophical Psychology*, 28(5), 642-673.
- Seto, E., and Hicks, J. A. (2016). Disassociating the Agent From the Self: Undermining Belief in Free Will Diminishes True Self-Knowledge. *Social Psychological and Personality Science*, 7(7), 726-734.
- Shalvi, S., Baas, M., Handgraaf, M. J., and De Dreu, C. K. (2010). Write when hot-Submit when not: Seasonal bias in peer review or acceptance?. *Learned Publishing*, 23(2), 117-123.
- Shumway-Cook, A., Gruber, W., Baldwin, M., and Liao, S. (1997). The effect of multidimensional exercises on balance, mobility, and fall risk in community-dwelling older adults. *Physical therapy*, 77(1), 46.
- Simon, H. A. (1990) Invariants of human behavior, *Annual Review of Psychology*, 41, 1-19.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational researcher*, 15(9), 5-11.
- Smithdeal, M. (2016). Belief in free will as an adaptive, ungrounded belief. *Philosophical Psychology*, 29(8), 1241-1252.
- Stanovich, K. E, and West. R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94, 672-695.
- Stegenga, J. (2011). Is meta-analysis the platinum standard of evidence? *Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences*, 42(4), 497-507.

- Steinpreis, R. E., Anders, K. A., and Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex roles*, 41(7-8), 509-528.
- Sternberg, R. J. (1991). Editorial. *Psychological Bulletin*, 109: 3-4.
- Stolee, P., Lim, S. N., Wilson, L., and Glenny, C. (2012). Inpatient versus home-based rehabilitation for older adults with musculoskeletal disorders: a systematic review. *Clinical rehabilitation*, 26(5), 387-402.
- Strech, D., and Sofaer, N. (2012). How to write a systematic review of reasons. *Journal of medical ethics*, 38(2), 121-126.
- Strech, D., Synofzik, M., and Marckmann, G. (2008). Systematic reviews of empirical bioethics. *Journal of Medical Ethics*, 34(6), 472-477.
- Sytsma, J., and Buckwalter, W. (Eds.). (2016). *A Companion to Experimental Philosophy*. John Wiley & Sons.
- Templier, M. and Parè, G. (2015) A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems* 37, 6.
- Tversky, A., and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.
- Van Raan, A. F. (2004). Sleeping beauties in science. *Scientometrics*, 59(3), 467-472.
- Vassar, M., Yerokhin, V., Sinnott, P. M., Weiher, M., Muckelrath, H., Carr, B., and Cook, G. (2017). Database selection in systematic reviews: an insight through clinical neurology. *Health Information & Libraries Journal*, 34(2), 156-164.
- Vassar, M., Carr, B., Kash-Holley, M., DeWitt, E., Koller, C., Day, J., ... and Holzmann, M. (2015). Database choices in endocrine systematic reviews. *Journal of the Medical Library Association: JMLA*, 103(4), 189.
- Vohs, K. D., and Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, 19, 49-54.
- Wray, K. B. (2010). Philosophy of science: What are the key journals in the field?. *Erkenntnis*, 72(3), 423-430.