



# Philosophers' linguistic expertise: a psycholinguistic approach to the expertise objection against experimental philosophy

Eugen Fischer<sup>1</sup> · Paul E. Engelhardt<sup>2</sup> · Aurélie Herbelot<sup>3</sup>

Received: 3 April 2021 / Accepted: 2 November 2021  
© The Author(s) 2022

## Abstract

Philosophers are often credited with particularly well-developed conceptual skills. The 'expertise objection' to experimental philosophy builds on this assumption to challenge inferences from findings about laypeople to conclusions about philosophers. We draw on psycholinguistics to develop and assess this objection. We examine whether philosophers are less or differently susceptible than laypersons to cognitive biases that affect how people understand verbal case descriptions and judge the cases described. We examine two possible sources of difference: Philosophers could be better at deploying concepts, and this could make them less susceptible to comprehension biases ('linguistic expertise objection'). Alternatively, exposure to different patterns of linguistic usage could render philosophers vulnerable to a fundamental comprehension bias, the linguistic salience bias, at different points ('linguistic usage objection'). Together, these objections mount a novel 'master argument' against experimental philosophy. To develop and empirically assess this argument, we employ corpus analysis and distributional semantic analysis and elicit plausibility ratings from academic philosophers and psychology undergraduates. Our findings suggest philosophers are better at deploying concepts than laypeople but are susceptible to the linguistic salience bias to a similar extent and at similar points. We identify methodological consequences for experimental philosophy and for philosophical thought experiments.

**Keywords** Experimental philosophy · Expertise objection · Philosophical method · Thought experiments · Comprehension inferences · Linguistic salience bias

---

✉ Eugen Fischer  
e.fischer@uea.ac.uk

<sup>1</sup> School of Politics, Philosophy, Language and Communication Studies, University of East Anglia, Norwich, UK

<sup>2</sup> School of Psychology, University of East Anglia, Norwich, UK

<sup>3</sup> Center for Mind/Brain Sciences/Department of Information Engineering and Computer Science, University of Trento, Rovereto (Trentino), Italy

## 1 Introduction

This paper will redevelop and assess the ‘expertise objection’ to experimental philosophy, by drawing on methods and findings from psycholinguistics. Experimental philosophy focuses on the empirical investigation of philosophically relevant intuitions. According to the expertise objection, experimental philosophers go wrong already at the first step of their empirical studies: they recruit the wrong participants.<sup>1</sup> Experimental philosophers typically recruit convenience samples without philosophical training: M-Turkers, psychology undergraduates, etc. But philosophical training and expertise improve thinkers’ conceptual competencies and, thereby, their intuitive case judgments. Findings about the intuitions of ‘laypeople’ are therefore irrelevant for philosophical research.

This objection has been initially directed at the ‘negative’, ‘restrictionist’, and ‘evidential’ strands of experimental philosophy. These strands seek to assess the evidentiary value of philosophically relevant intuitions and examine intuitions elicited by verbal case descriptions in philosophical thought experiments (reviews: Machery, 2017; Mallon, 2016). Empirical findings about laypeople’s intuitions about X—specifically, that they are sensitive to irrelevant factors or cognitive biases—are meant to support the conclusion that professional philosophers should not treat (all or some of) their own intuitions about X as evidence for philosophical theories. These methodological arguments rely on the inductive ‘*lay-expert inference*’ from experimental findings about laypeople to the conclusion that also professional philosophers’ intuitions will be influenced by the irrelevant factors and biases found to affect lay participants. The *expertise objection* challenges this inference: The objection assumes that professional philosophers have a methodological or conceptual expertise that laypeople possess to a lesser extent; it suggests that this expertise makes philosophers less vulnerable to the irrelevant factors and biases that affect laypeople’s case judgments; and it infers that philosophers’ intuitions are more stable and accurate (reviews: Machery, 2017, pp. 158–169; Nado, 2014). This has consequences also for straightforwardly ‘positive’ experimental philosophy (e.g., for ‘conceptual analysis 2.0’; Machery, 2017, pp. 208–244): If necessary at all, experimental implementations of the method of cases should recruit philosophers as participants.

The empirical assessment of this objection simultaneously promises to contribute to elucidating the nature of philosophical expertise. To assess the objection, the ‘direct strategy’ conducts experiments with laypeople and philosophers that examine whether irrelevant factors or biases affect the two groups’ intuitions about philosophically relevant cases differently. Only few studies to date have clearly executed this compelling strategy, with a strong focus on moral intuitions (see Sect. 2). Our paper will range further and dig deeper: We turn from intuitions about specific kinds of cases to comprehension inferences which determine how case descriptions are interpreted and thereby shape judgments about the cases described, in any area of philosophy. This move will allow us to redevelop the expertise objection by drawing on psycholinguistics.

---

<sup>1</sup> The complementary ‘reflection objection’ charges that most experimental philosophy studies the wrong judgments—spontaneous, rather than reflective (review: Machery, 2017, pp. 155–158). For its empirical assessment, see, e.g., de Bruin (2021), Kneer et al., 2021; Schwitzgebel and Cushman (2015).

An experiment will employ the direct strategy to examine whether academic philosophers are better than psychology undergraduates at deploying conceptual information and whether philosophers are less susceptible to cognitive biases affecting the interpretation of case descriptions.

Section 2 distinguishes different versions of the expertise objection and reviews extant evidence to identify the most promising version or objection. Section 3 draws on findings from psycholinguistics and experimental philosophy to develop this *linguistic expertise objection (LEO)*, complement it with the new *linguistic usage objection (LUO)*, and outline how these two objections jointly provide a ‘*master argument*’ against experimental philosophy’s lay-expert inference. Sections 4 and 5 empirically examine these two objections. Section 6 discusses the findings’—productive—consequences for both experimental philosophy and the methodology of philosophical thought experiments.

## 2 Expertise objections

The expertise objection is commonly motivated by an analogy: Like members of other academic disciplines, philosophers have specific professional expertise. Analytic philosophers arguably ‘are experts in the analysis of folk concepts’ (Horvath, 2010, p.465). Such analysis involves thought experiments that elicit intuitions about the applicability of concepts in hypothetical cases. While philosophers’ professional expertise will extend considerably further, it should therefore encompass an ‘intuitive expertise’: Like, e.g., the mathematical intuitions of mathematicians, philosophers’ intuitions about the applicability of concepts to hypothetical cases will be more reliable than those of non-experts (e.g., Hales, 2006, p. 171; Williamson, 2011, p. 220). This undermines experimental philosophy’s lay-expert inference (Devitt, 2011; Hales, 2006; Horvath, 2010; Kauppinen, 2007; Ludwig, 2007; Williamson, 2007, 2011).

The ‘intuitive expertise’ is taken to arise from philosophers’ superior ability to ‘apply general concepts to specific examples with careful attention to the relevant subtleties’ (Williamson, 2007, p. 191; cf. Ludwig, 2007, p. 138; Horvath, 2010, pp. 466–467). This superior conceptual competence can be due to different kinds of professional expertise that philosophers could credibly claim as a result of training or selection effects. Plausibly, philosophers are better versed in the methods of philosophical thought experimentation. Weinberg et al. (2010, p. 336) distilled from the debate the further suggestions that philosophers could benefit from better conceptual schemata or domain theories, or from better cognitive skills than laypeople. That is, philosophers could possess better relevant conceptual or world knowledge, or could be better at deploying their knowledge in making their judgments.

We thus obtain three distinct versions of the expertise objection that have been advanced, often in tandem:

- According to the ‘*methodological expertise objection*’, philosophers have more experience with the method of cases. This makes them better at interpreting the task and taking into account precisely the task-relevant information in vignettes (Ludwig, 2007, p. 153; Williamson, 2011, p. 216).

- According to the ‘*epistemological expertise objection*’, philosophical training and research lead philosophers to develop more extensive or better structured representations of conceptual and other knowledge about the domain of their philosophical theorizing. This makes their case judgments better informed and more sensitive to relevant information (Devitt, 2011, p. 426; cf. Ludwig, 2007, p. 153; Weinberg et al., 2010, pp. 335–336).
- According to the ‘*linguistic expertise objection*’, philosophers are better at *deploying* semantic or conceptual knowledge: In judgment and reasoning about verbally described cases, they are generally better at contextualizing conceptual information (Williamson, 2011, p. 216); i.e., they are better at taking into account also contextual information and background knowledge, e.g., in disambiguating ambiguous expressions and enriching sketchy case descriptions (Horvath, 2010, p. 467).

All objections claim that philosophers possess a certain expertise or skill to a higher extent than laypeople, assume that this expertise or skill renders intuitive case judgments more reliable, and conclude that philosophers’ case judgments are *more stable*, i.e., less susceptible to irrelevant factors and cognitive biases, and more accurate than laypersons’ intuitions.<sup>2</sup>

Philosophers’ intuitions can only be more stable and accurate than laypeople’s if they are different. The ‘*direct strategy*’ (Schulz et al., 2011, p. 1724) therefore assesses empirically (1) whether philosophers’ intuitive case judgments about a domain differ from lay judgments. It further assesses (2) whether the philosophers’ judgments are more stable. The assessment (3) of their relative accuracy is difficult since there are no uncontroversial ways of telling which philosophically relevant intuitions are accurate. Experimentalists have examined, instead, whether philosophers’ intuitions are more internally coherent (Löhr, 2019) or closer to a textbook consensus (Horvath & Wiegmann, 2016; Schindler & Saint-Germier, forthcoming).

Eight studies to date clearly execute the first two steps. All examine ethically relevant intuitions. All show that philosophers’ intuitions are influenced by irrelevant factors or biases. Four studies on ethically relevant intuitions (about hedonism, free will/moral responsibility, and moral dilemmas) did not ensure that philosophical participants had high levels of relevant topical expertise (Löhr, 2019; Schulz et al., 2011; Tobia, Buckwalter, et al., 2013a, Tobia, Chapman, et al., 2013b). Even so, these participants will have been proficient with the method of cases that is used across different areas of philosophy. These studies can therefore be regarded as addressing (only) the methodological expertise objection—and finding against it.<sup>3</sup> Four further studies simultaneously addressed also the epistemological expertise objection, by recruiting expert ethicists for an investigation of moral intuitions: They compared the moral permissibility judgments laypeople and expert ethicists make about trolley or related cases. Both groups’ judgments were subject to order effects of the same size, reduced

<sup>2</sup> Egler and Ross (2020) clarify the structure of these familiar objections. Schindler and Saint-Germier (forthcoming) propose a novel ‘informed expertise objection’ that suggests philosophers’ case judgments are more likely than laypeople’s judgments to be based on precisely the relevant information about the case and therefore more likely to address the thought experimentalist’s research question.

<sup>3</sup> This objection may merely call for improved experimental materials: Materials that render task-relevant information more salient can align the judgments of lay participants with philosophical textbook consensus, even where they previously diverged from it (Turri, 2013).

neither by reflection prompts nor self-reported expertise on the specific issues in question (Schwitzgebel & Cushman, 2012, 2015); experts' intuitions were no less sensitive to order effects and an irrelevant factor (inclusion of irrelevant response options) (Wiegmann et al., 2020) and were susceptible to almost as many psychologically distinct framing effects as laypersons' (Horvath & Wiegmann, 2021). These studies speak against the epistemological expertise objection: either philosophical ethicists do not have more extensive or better structured moral knowledge than laypeople, or such 'philosophically improved' moral knowledge does not render people's moral case judgments notably less susceptible to irrelevant factors and biases.

Two studies examining accuracy rather than stability suggest that the difficulties documented for the epistemological and methodological expertise objections are not restricted to the domain of moral philosophy. Horvath and Wiegmann (2016) found the intuitive knowledge attributions of expert epistemologists were only partially consistent with the textbook consensus. A recent study speaks to the methodological expertise objection: Schindler and Saint-Germier (forthcoming) compared philosophers' and laypersons' judgments about six cases pertaining to thought experiments from across theoretical philosophy and found philosophers' judgments were significantly closer to the textbook consensus for—only—half the cases.<sup>4</sup>

While these first two expertise objections require further investigation, extant findings motivate turning to the remaining linguistic expertise objection. Our study is the first to develop and assess this objection—and to execute all three steps of the direct strategy. We now set out this empirically neglected objection, explain why it matters, and how we propose to render it empirically tractable.

According to the *linguistic expertise objection (LEO)*, philosophers are better than laypeople at *deploying* conceptual information (even when they possess the same conceptual information as laypeople); this deployment competence makes their judgments about verbally described cases more stable and accurate. This objection considers the process that leads, in philosophical thought experiments, from verbal case descriptions to intuitive judgments about the cases described. Properly understood, LEO addresses the first stage of the process: the interpretation of the verbal case description. Psycholinguistic research (to be reviewed in Sect. 3.1) reveals that the interpretation readers place on texts is built up from 'conceptual' information that is automatically activated by words, by default, as we read them. The interpretation process involves integrating information that gets sequentially activated, as we read through the text: we need to *integrate* information activated by words we read now with information activated by words we read previously; we need to *complement* information activated by individual words with information activated only by larger chunks of text (e.g., combinations of words) or wider discourse context, and with background knowledge; and we need to *suppress* initially activated information that subsequently turns out to be irrelevant in the given context. Being better at deploying conceptual information thus amounts to being better at contextualizing conceptual information in these ways.

<sup>4</sup> Further studies that address the expertise objection less directly include Beebe and Monaghan (2018), Carter et al. (2016, 2019), Hitchcock and Knobe (2009), Machery (2012), Schindler and Saint-Germier (2020), Starmans and Friedman (2020), Sytma and Machery (2010), and Vaesen et al. (2013).

As developed in the light of these empirical findings, LEO assumes that:

- (1) Philosophers are better than laypeople at contextualizing conceptual information, that is, at complementing and suppressing default information, as appropriate. LEO further assumes that.
- (2) Better contextualization (complementation and suppression) ability renders philosophers' interpretations of vignettes less susceptible to comprehension biases and, thereby, less sensitive to irrelevant factors (like verbal differences between equivalent formulations or order of presentation). Philosophical vignettes are crafted to include the information to be taken into account in making the judgments of interest to the thought experimentalist. This motivates the third assumption:
- (3) Improved ability to take sentence and discourse context into account through complementation and suppression of default information will better align readers' interpretations with the intended interpretation.

Improved contextualization ability thus renders philosophers' *interpretations* of vignettes more stable and accurate. Since these interpretations shape the intuitive judgments people make about the cases described, LEO infers that also philosophers' intuitive case *judgments* are more stable and accurate than those of laypeople.

LEO challenges experimental philosophy's lay-expert inference for many important philosophical thought experiments: The default information activated by words includes mainly information about typical properties of objects, people, and events (see Sect. 3.1). However, to address their research questions, philosophical thought experiments frequently need to consider unusual cases that pull apart features that typically go together (Machery, 2017, pp. 111–118). To accurately interpret descriptions of such cases, people need to either complement the default information with further contextual information or to suppress some of the default information that is stipulated not to apply to the case. For example, to correctly interpret Gettier cases, people need to complement the information that the protagonist has a justified true belief with the further information that they are right by chance (which is atypical for cases of justified true belief)—and need to take both into account in their case judgments (Turri, 2013). Similarly, to correctly interpret zombie scenarios, people need to disambiguate the polysemous term 'zombie' and suppress the default information that zombies have rotting bodies and attack and eat humans, to take into account that the 'philosophical zombies' at issue are physico-behaviorally indistinguishable from us (Fischer & Sytsma, 2021). It is therefore *prima facie* plausible to suggest that pronounced differences in the ability to complement and suppress default information can translate into different judgments in many important philosophical thought experiments.

We propose to go beyond extant studies not only in examining this empirically neglected expertise objection, but also in drilling down deeper. To contribute to the gradual elucidation of how different cognitive skills are involved in philosophical expertise, intuitive or other, we drill down to the level of specific cognitive skills, as captured by empirically valid psychological constructs. Above, we distinguished three relevant kinds of expertise and detailed how extant studies found against expertise objections based on two of them. In turning to the remaining expertise of interest, we employ a '*specific skills approach*': We consider specific cognitive skills that

underwrite the expertise, and ask whether philosophers possess a particular skill to a higher extent than laypeople (as per assumption 1 above), and whether this renders philosophers' judgments more stable (as per 2) and more accurate (as per 3).<sup>5</sup> With this approach, we examine suppression or 'inhibition' (a focus motivated in Sect. 3.1 below), investigate susceptibility to the comprehension bias from which higher inhibition is most likely to shield participants (Sect. 3.2), and study its influence on interpretation accuracy (see Sect. 6.1). The novel approach also motivates the use, in the main study (Sect. 5), of simple (one-sentence) items, whose interpretation does not stand to benefit from familiarity with philosophical thought experimentation or expert background knowledge. This allows for targeted examination of the linguistic expertise objection, without confounds pertaining to the methodological or epistemological expertise objections.

### 3 Two complementary objections

According to the linguistic expertise objection (LEO), philosophers are better than laypeople at deploying conceptual information and this makes their judgments about verbally described cases more stable. We now draw on research from psycholinguistics in order to translate this objection into empirically testable hypotheses. To do so, we spell out what 'deploying conceptual information' amounts to (Sect. 3.1) and identify a philosophically relevant bias that better 'deployment competence' should shield philosophers from (Sect. 3.2). These two steps will translate the objection's first two assumptions—(1) and (2) above, respectively—into testable hypotheses. Appreciation of the bias will simultaneously motivate a new 'linguistic usage objection' (Sect. 3.3).

#### 3.1 Conceptual information and its deployment

What is psychologically real 'conceptual information'? Cognitive science draws the distinction between conceptual and other information in processing terms and typically conceives of '*concepts*' as bodies of information stored in long-term memory and retrieved by default, in the exercise of higher cognitive competencies including language comprehension, perceptual categorization, and inductive learning (review: Machery, 2009). Conceptual information thus is information that is retrieved *by default*, i.e., rapidly retrieved (e.g., in response to a verbal stimulus), either in every context (such as any textual context) (Machery, 2017) or outside all context (as in single word priming experiments) (Fischer, 2020), by an automatic process (Bargh et al., 2012).

The information that qualifies as 'conceptual' in virtue of default retrieval mostly is information about the world that philosophers consider 'empirical': Information is retrieved automatically through activation of representations including stereotypes (a.k.a. 'prototypes' or 'schemas'). *Stereotypes* are built up through observation of co-occurrences in the physical environment and through extraction of co-occurrence

<sup>5</sup> Parallel work by Schindler and Saint-Germier (forthcoming) (which we only learned about at the revision stage) pursues a similar (if less directly psychologically informed) approach.



information from linguistic discourse (McRae & Jones, 2013). They encode statistical information about typical and diagnostic properties of category members (Hampton, 2006). More complex stereotypes (situation schemas) encode information about typical features of events or actions, agents, ‘patients’ acted on, and typical relations between them (Ferretti et al., 2001; Hare et al., 2009; McRae et al., 1997). Dependency networks in complex schemas encode causal, functional, and nomological information (Sloman et al., 1998). Much of this ‘world knowledge’ qualifies as conceptual information, due to default activation: Many stereotypes are associated with nouns and verbs which rapidly activate them in single-word priming experiments (Lucas, 2000).

Activated stereotypes support defeasible default inferences about what (else) is (also) true of the situation talked about (e.g., unless indicated otherwise, the ‘tomato’ is red; Levinson, 2000).<sup>6</sup> ‘Conceptual’ information in cognitive science’s sense, namely, statistical world knowledge encoded by stereotypes, thus provides an initial basis for utterance interpretation (Elman, 2009). For present purposes, the most relevant utterances are the case descriptions philosophers consider in thought experiments—and typically encounter through reading, like participants in experimental-philosophy studies. In reading comprehension, relevant conceptual knowledge and further world knowledge need to be integrated into the *situation model*: the mental representation of the situation described by the text, which provides the basis for further judgements and reasoning about that situation (Kintsch, 1988; Zwaan, 2016). To facilitate accurate judgment and reasoning about specific situations, we need to contextualize our default inferences. In this setting, the competence of ‘*deploying conceptual information*’ consists in a twofold ability to manage the information that individual words activate by default, as we read them: the ability to *suppress* from the situation model the conclusions of default inferences that are contextually irrelevant (Faust & Gernsbacher, 1996), and to *complement* relevant default information with further world knowledge that is contextually relevant but is activated only by combinations of words rather than any single word (Bicknell et al., 2010; Matsuki et al., 2011), in the sentence or wider discourse context (Metusalem et al., 2012).

Competence at these tasks is modulated by two different forms of intelligence (Cattell, 1987): ‘fluid intelligence’ only minimally depends upon prior learning; ‘crystallized intelligence’ reflects cultural learning and includes both world or domain knowledge and lexical knowledge. Better domain knowledge helps readers to complement conceptual knowledge, to arrive at utterance interpretations that are positive, stereotypical, and specific (Levinson, 2000, pp. 114–115; Garrett & Harnish, 2007). Better domain knowledge also cancels stereotypical inferences that less knowledgeable readers regard as relevant. Similarly, richer lexical knowledge supports both complementation and suppression, namely, by facilitating pragmatic inferences from oppositions between authors’ chosen words and informationally stronger and weaker expressions (Levinson, 2000, pp. 75–104) and from authors’ preferences of marked

<sup>6</sup> These automatic inferences are instrumental in facilitating effective communication in the face of the ‘articulation bottleneck’: Normal speech conveys information at a slow rate of under 100 bits per second (Levinson, 2000, p. 28). Pre-articulation processes in speech production are 3–4 times faster (Wheeldon & Levelt, 1995), as are parsing processes and comprehension inferences (Mehler et al., 1993). Default inferences that deploy our statistical knowledge about the world allow hearers to rapidly fill in detail. Anticipating such inferences allows speakers to skip mention of typical features and use fewer words.



expressions over shorter, more frequent, or neutral words (pp. 136–137). These pragmatic inferences can complement or defeat stereotypical inferences (pp. 157–158). At the level of fluid intelligence, low-level cognitive abilities conceptualized as ‘*executive functions*’ (Miyake et al., 2000) modulate the exercise of several cognitive competencies, including reading comprehension (review: Butterfuss & Kendeou, 2018). For our purposes, the key function is *inhibition* (Miyake & Friedman, 2012; cf. Dempster, 1991): the ability to manage the activation of irrelevant information and to actively inhibit or suppress prepotent responses to stimuli—such as default inferences from verbal stimuli, where they are contextually irrelevant.

Better domain knowledge is claimed for philosophers by the epistemological expertise objection which did not stand up well to empirical scrutiny (Sect. 2). By contrast, it is a priori plausible that, due to training and selection effects, academic philosophers should benefit (1) from better lexical knowledge, which correlates with years in formal education (Engelhardt et al., 2008) and extent of reading (Stanovich, 1993), and (2) from higher inhibition, which correlates with verbal intelligence in adolescents and adults (Friedman et al., 2006). On balance, these two factors favor suppression of irrelevant default information more than complementation with relevant further information. In developing the linguistic expertise objection (LEO), we therefore focus on suppression ability: to test LEO’s first assumption, that philosophers are better than laypeople at contextualizing conceptual information, we’ll examine

**H<sub>1</sub>** Academic philosophers are better than laypeople (e.g., psychology undergraduates) at suppressing default inferences that are contextually irrelevant.

### 3.2 A philosophically relevant cognitive bias

According to LEO’s second assumption, higher levels of conceptual competence shield philosophers from comprehension biases. The stronger competence claimed by **H<sub>1</sub>** should shield them at least against biases that promote contextually irrelevant stereotypical inferences.

One such bias is the linguistic salience bias that affects polysemy processing. Many words (over 40% in English) are *polysemous*, i.e., have several distinct, but related senses (Byrd et al., 1987). Subordinate senses can sometimes be generated by rules (as in metonymy) and sometimes not (as in metaphor) and are processed accordingly (reviews: Eddington & Tokowicz, 2015; Vicente, 2018). Different senses of ‘irregular’ polysemes do not activate distinct semantic representations (Klepousniotou et al., 2012; MacGregor et al., 2015), but a ‘unitary representation’ that consists of overlapping feature clusters (stereotypes) (Brocher et al., 2016). The interpretation of specific uses involves suppressing component features that are not shared by different senses and irrelevant in the given utterance context (cf. Giora, 2003; Giora et al., 2007). E.g., the verb ‘to see’ activates a schema with agent features including *S looks at X*, *S knows X is there*, and *S knows what X is*, and patient features including *X is in front of S* and *X is near S*. To interpret a purely epistemic use (‘Mary saw the possibilities’), the hearer needs to suppress all features except the epistemic agent features, to obtain the intended interpretation (Mary knew there were possibilities and knew what they were).

Such suppression becomes difficult where one sense exceeds all others in linguistic salience. The *linguistic salience* of a sense is a function of exposure frequency (of how often the hearer encounters the word in one sense, rather than another), modulated by prototypicality (how good examples of the relevant category—say, *seeing*—the word is deemed to stand for in this sense) (Giora, 2003). The feature cluster associated with more frequently encountered senses are activated more strongly (Brocher et al., 2018), and clusters constitutive of more prototypical sub-categories are activated more strongly (Hampton, 2006). Accordingly, features associated with the most salient sense are activated most strongly. Frequently co-occurring component features of an activated stereotype exchange lateral cross-activation (Hare et al., 2009; McRae et al., 2005). Where such cross-activation complements strong initial activation due to high linguistic salience, feature suppression becomes difficult. Irrelevant component features of the dominant stereotype then remain partially activated and support inappropriate inferences (from ‘Mary saw the possibilities’ to *the possibilities were in front of Mary*), as per the linguistic salience bias hypothesis (SBH) (Fischer & Engelhardt, 2019, 2020): When

- (i) one sense of an irregular polyseme is much more salient than all others,
- (ii) interpretation of utterances using a subordinate sense requires suppression of features associated with that dominant sense, and
- (iii) some, but not all, of the features strongly associated with the dominant sense are contextually relevant

then

- (1) contextually irrelevant stereotypical inferences supported by the dominant sense will be triggered by the subordinate use as well, and
- (2) these automatic inferences will influence further judgment and reasoning.

This bias matters for philosophy: Philosophers often employ familiar words in new, but related senses, so that conditions (i) and (ii) are met (Fischer et al., 2021a, 2021b). Philosophical thought experiments often pull apart features that typically go together (Machery, 2017, pp. 116–18), so that (iii) is met. In such thought experiments and related arguments, case descriptions will trigger contextually inappropriate inferences whose conclusions will enter the situation model on which judgments and reasoning about the described case are based. For example, Fischer and Engelhardt (2020) suggested that the ‘argument from hallucination’ relies on contextually inappropriate default inferences from phenomenal uses of perception verbs (‘Macbeth saw a dagger’) to factive and spatial conclusions (*There was a dagger in front of Macbeth*) that are cancelled by the context but, even so, presupposed in further reasoning (from ‘There was no physical dagger before Macbeth’ to ‘There was a non-physical dagger’).

Empirically studied examples include inappropriate inferences from appearance- and perception-verbs in arguments from illusion and hallucination (Fischer & Engelhardt, 2016, 2017, 2019, 2020; Fischer et al., 2021a, 2021b), from ‘zombie’ in the eponymous argument (Fischer & Sytsma, 2021), and from purely descriptive uses of the verb ‘cause’ in morally valenced cases (Livengood & Sytsma, 2020; Livengood et al., 2017).

As proponents of the expertise objection have plausibly assumed (e.g., Horvath, 2010, p. 471), analytic philosophers are well trained in distinguishing, explaining,

and reasoning with, different senses of words. Polysemy processing therefore is an arena in which it is particularly plausible to expect analytic philosophers to be better at deploying conceptual knowledge than laypeople. The linguistic salience bias that affects polysemy processing is a cognitive bias from which better suppression ability (as per  $H_1$ ) seems most apt to shield philosophers. This bias is therefore ideally suited to put LEO to the test: we will examine the objection's second assumption, that better contextualization ability renders philosophers less susceptible to comprehension biases, by investigating

$H_2$  Professional analytic philosophers are less susceptible to linguistic salience bias than laypeople.

We thus obtain this *empirically developed LEO*: Philosophers are better than laypeople at suppressing contextually irrelevant default inferences (as per  $H_1$ ). They are therefore less susceptible to cognitive biases including the linguistic salience bias (as per  $H_2$ ). As a result, their interpretations of verbally described cases will be more stable and based on more coherent situation models—and, therefore more accurate (as per LEO's remaining third assumption). This will render philosophers' case judgments more stable and accurate.

### 3.3 LUO: the linguistic usage objection

The linguistic salience bias hypothesis simultaneously motivates a new, alternative objection to the lay-expert inference: If (*pace*  $H_2$ ) philosophers *are* equally susceptible to this bias, they will be susceptible to it at different points. Specialists may use a word much more frequently in a technical sense than laypeople do. In this case, an ordinary sense that stands out in salience for laypeople will not stand out so much for specialists. Even if linguistic salience bias leads laypeople to make contextually inappropriate stereotypical inferences that are supported by that word's ordinarily dominant sense, the bias will not lead the specialists to make these inferences. Conversely, specialist discourse may make a sense that is not dominant in ordinary discourse clearly stand out in linguistic salience for specialists. This salience imbalance may lead specialists to make inappropriate stereotypical inferences that laypeople avoid. Either way, different inferences will feed into the situation models that ground laypeople's and philosophers' judgments about verbally described cases, and their responses will differ, as will the soundness of these responses. This *linguistic usage objection (LUO)* translates into two hypotheses:

$H_f$  [**Frequency Hypothesis**] Different senses or uses of some familiar words have notably different relative exposure frequencies for laypeople and expert philosophers.

$H_3$  These differences make expert philosophers and laypeople vulnerable to linguistic salience bias when encountering different words.

Together, LEO and LUO promise to add up to a '*master argument*' against experimental philosophy's lay-expert inference: If academic philosophers are better at suppressing irrelevant default inferences (as per  $H_1$ ) and therefore less susceptible to linguistic salience bias than laypeople (as per  $H_2$ ), the philosophers will make

fewer mistakes. If they are equally susceptible to the bias (and their linguistic diet differs from laypersons', as per  $H_F$ ), philosophers will make *different* mistakes (as per  $H_3$ ). Either way, experimental findings about lay responses to verbally described cases will not simply carry over to expert philosophers. This argument challenges the lay-expert inference where philosophers use irregular polysemes from ordinary discourse in special senses, to talk about unusual cases that pull apart what typically goes together. To empirically assess this argument, we conducted corpus analyses and an experiment.

## 4 Corpus analyses

Exposure frequencies are commonly inferred from occurrence frequencies in corpora. To examine the frequency hypothesis  $H_F$  and derive empirically testable predictions from the competing hypotheses  $H_2$  and  $H_3$ , we conducted three manual corpus studies (Sect. 4.1) and distributional semantic analysis (Sect. 4.2). To support  $H_F$  and test  $H_3$ , we need to identify polysemous words of philosophical interest that display pronounced salience imbalances in ordinary discourse which are absent or reversed in specialist philosophical discourse (so that similar susceptibility to linguistic salience bias will lead laypeople and philosophers to make different inappropriate inferences). To test  $H_2$ , we need words where salience imbalances are preserved in philosophical discourse (so that different propensities to make inappropriate inferences from them will be indicative of different susceptibility to the bias). The best experimental evidence for linguistic salience bias of philosophical interest (see Sect. 3.2) comes from two perception-verbs (Fischer & Engelhardt, 2017, 2019, 2020). We examined these verbs, to ascertain whether they provide what we need.

### 4.1 Corpus analyses

We examined the use of the verbs 'see' and 'be aware of' in samples of at least 1000 sentences randomly drawn from three corpora roughly representative of ordinary discourse, academic philosophy, and a specific sub-area, respectively: (1) the *British National Corpus (BNC)*, (2) a topically generic philosophy corpus compiled from two philosophy encyclopedias (*Stanford Encyclopedia of Philosophy* and *International Encyclopedia of Philosophy (SEP/IEP)*), and (3) a philosophy of perception corpus (*PHILO-P*) comprised of ten monographs that shaped philosophical debates about sense-data, (challenges to) naïve and direct realism, and the resulting 'problem of perception'. We classified the occurrences of 'see' and 'aware of' as perceptual or non-perceptual and assigned uses of 'see' to one of twelve dictionary-attested senses. Methods and results are detailed in Online Appendix A.

Headline findings (Table 1) provide evidence of pronounced salience imbalances in ordinary discourse that, in specialist philosophical discourse, are roughly preserved for 'see' and reversed for 'aware of'. In ordinary discourse (BNC), perceptual uses (where the agent perceives by sense the object of sight or awareness) are clearly dominant for 'see' and clearly subordinate for 'aware of'. Slight changes in usage patterns across

**Table 1** Perceptual uses as percentage of classifiable uses in random samples from corpora

	BNC	SEP-IEP	PHILO-P
<b>‘see’</b>	71%	59%	60%
<b>‘aware of’</b>	21%	40%	91%

corpora for ‘see’ are driven mainly by an increase of purely epistemic uses of ‘see’ (‘know/understand something’ or ‘find out’ without using one’s eyes), from 12% of classifiable occurrences in the BNC sample to 23% in the SEP-IEP sample and 36% in PHILO-P (see Online Appendix A). For ‘aware of’, we observe a dominance reversal between ordinary discourse, where the purely epistemic use (‘know about a fact or situation’) dominates, and specialist discourse (PHILO-P), where the perceptual use is dominant. The two verbs seem to provide what we need.

## 4.2 Distributional semantic analysis

To extend our analysis, we built a computational model of the uses of ‘see’ and ‘aware of’ across our three corpora. Methods and results are detailed in Online Appendix B.

We constructed distributional semantics representations of each occurrence of either verb in our annotated samples. We used those representations to train a classifier which classifies a given occurrence as perceptual or non-perceptual. We had already annotated manually all uses of ‘see’ and ‘aware of’ in the smaller PHILO-P corpus but deployed the classifier to classify all their uses in the larger corpora—and considered separately their use in the academic section of the BNC (ACPROSE) and the remainder of this corpus (Table 2).

In the whole BNC we observed a still dominant, but lower proportion of perceptual uses of ‘see’ than in our random sample. This demonstrates the usefulness of automatic classification to correct potential sampling biases. For the BNC, we now observe an almost identical proportion as for SEP-IEP and PHILO-P. The markedly lower proportion in ACPROSE (42%) suggests that academic philosophers may be professionally exposed to perceptual uses of ‘see’ less frequently than the philosophy corpora suggest. Even so, differences in exposure frequencies between academic philosophers and laypersons seem bound to remain minor for ‘see’. Distributional semantic analysis thus confirms that perceptual uses of ‘see’ will be roughly equally salient for academic philosophers and laypeople, so that any differences in judgment and reasoning will be due to different susceptibility to linguistic salience bias (as per

**Table 2** Perceptual uses as percentage of classifiable uses in different corpora

	BNC	BNC without ACPROSE	ACPROSE	SEP-IEP
<b>‘see’</b>	62%	65%	42%	61%
<b>‘aware of’</b>	27%	31%	10%	45%

**Table 3** Classification of (A) ‘see’, (B) ‘aware’

	BNC (%)	SEP-IEP (%)	PHILO-P (%)
(A) ‘see’			
Baseline	71	59	60
Accuracy	90	98	96
(B) ‘aware’			
Baseline	79	60	91
Accuracy	90	98	92

**H<sub>2</sub>**. Findings for ‘aware of’ confirm the dominance reversal in the philosophy of perception, as reflected in PHILO-P: In all other corpora, the verb’s perceptual use is subordinate. This dominance reversal (as per **H<sub>F</sub>**) suggests comparisons of philosophers of perception with other philosophers and laypeople will allow us to assess **H<sub>3</sub>**.

Further relevant findings emerge from prior validation of our classifier. Classifiers are validated by assessing their verdicts against human annotations and showing that they perform better than a simple chance heuristic (which classifies all occurrences of a word as instances of its dominant use in the corpus). We observed major improvements on this baseline, and accuracy over 90% (Table 3). This indicates that context words (without even syntactic parsing) provide enough information to identify non-perceptual uses of ‘see’ and ‘aware of’, whose interpretation requires suppression of initially activated schema components (see Sect. 5.1). We infer that no specialist knowledge is required to identify need for suppression; laypeople should perform as well at the task as philosophers.

To follow up this suggestion and determine whether differences in linguistic diet might make it difficult for laypeople to identify perceptual vs non-perceptual uses in unfamiliar discourse settings like philosophical vignettes, we performed cross-domain classification: We trained our classifier on one domain’s annotation (e.g., BNC) and tested its accuracy on an annotated sample from another domain (e.g., SEP-IEP). Results (Table 4) still show considerable improvements over baseline.

**Table 4** Classification of (A) ‘see’ cross-domain, (B) ‘aware’ cross-domain

	BNC (%)	SEP-IEP (%)	PHILO-P (%)
(A) ‘see’			
BNC		96%	83%
SEP	87%		94%
PHILO-P	81%	97%	
(B) ‘aware’			
BNC		85%	88%
SEP	89%		92%
PHILO-P	75%	80%	

Moderate drops in performance are observed in specific directions, especially when training on PHILO-P, but accuracy remains over 80% in nearly all cases—even though classifications are based only on information about word co-occurrences. Humans, who can take into account also syntactic information, wider context, and world knowledge, should have little trouble identifying non-perceptual uses (or need for suppression) in unfamiliar discourse settings. Validation and cross-domain classification findings jointly suggest that laypeople are no less able than philosophers to identify subordinate uses of our target words—and the need to suppress components of initially activated schemas—when reading experimental vignettes.

Responses to texts using the two verbs therefore allow us to study to what extent laypeople and philosophers differ in their ability to act on this insight—where exposure frequencies are similar ('see') (to test  $H_2$  and LEO) and where they differ ('aware of') (to test  $H_3$  and LUO).

## 5 Experiment

To examine the competing objections (LEO and LUO), we employed a plausibility rating task, and compared responses from psychology undergraduates, philosophers of perception ('*PoPs*'), and '*Other Philosophers*'.

### 5.1 Predictions

We used the psycholinguistic cancellation paradigm to examine spatial inferences from visual and purely epistemic uses of 'S sees X' and 'S is aware of X' to *X is in front of S*.<sup>7</sup> Participants read sentences with concrete and abstract objects, intended to invite visual and purely epistemic readings of the verb, respectively.

(1a/b) Matt sees / is aware of the spot on the wall facing him. (*s-consistent visual*)

(2a/b) Joe sees / is aware of the problems facing him. (*s-consistent epistemic*)

Half the items were inconsistent with the 'see'-stereotype ('*s-inconsistent*') and placed the object behind the agent:

(3a/b) Chuck sees/is aware of the spot on the wall behind him. (*s-inconsistent visual*)

(4a/b) Jack sees/is aware of the problems that lie behind him. (*s-inconsistent epistemic*)

Arguably due to embodiment effects (Glenberg & Kaschak, 2002), both 'see' and 'aware of' initially activate a schema that sees the agent looking at an object of sight/awareness before them (Fischer & Engelhardt, 2019, pp. 71–72, 81, 2020, p. 428).

<sup>7</sup> In this paradigm, participants read or hear sentences where the expression of interest is followed by a sequel that is inconsistent with (or 'cancels') a hypothesised inference from that expression. If the automatic inference is triggered, its clash with the sequel will engender comprehension difficulties requiring cognitive effort. If the inference is not suppressed, the perceived clash will persist and lower the sentence's plausibility. Effort is picked up by eye-tracking measures including pupil dilations and longer 'late' reading times, and by signature electrophysiological responses ('N400s'); plausibility is assessed with rating tasks (see Fischer & Engelhardt, 2019, for a review).



Both verbs thus trigger spatial default inferences, which clash with s-inconsistent sequels.<sup>8</sup> In response to such conflicts, stereotypical inferences can be completely suppressed within one second and fail to influence subsequent unspeeded plausibility judgments (Fischer & Engelhardt, 2017).

The linguistic salience bias hypothesis (**SBH**) maintains that participants are unable to completely suppress such contextually inappropriate inferences where the stereotype supporting them is associated with the dominant use of the verb. For ‘see’, the perceptual use that supports spatial inferences is dominant in ordinary and philosophical discourse (Sect. 4). The **SBH** hence predicts that spatial inferences will influence plausibility judgments of laypeople and philosophers even where ‘see’ is ostensibly used in a purely epistemic sense (including s-inconsistent items like 4a). The dictionary-attested purely epistemic sense of ‘see’ (‘know/understand something’) and familiar spatial time metaphors (whereby ahead = in the future; behind = in the past) facilitate purely metaphorical interpretations of these items (*Joe knows what problems he will have in the future* and *Jack knows what problems he had in the past*). To obtain these intended interpretations, participants need to completely suppress initial spatial inferences. But what if participants cannot suppress spatial inferences from ‘see’? The space–time metaphors in our items give rise to embodied cognition effects (Boroditsky & Ramscar, 2002; Bottini et al., 2015) and support spatial reasoning about temporal relations (Casasanto & Boroditsky, 2008; Gentner et al., 2002). Persistent spatial inferences from ‘see’ will prevent purely metaphorical interpretation also of the space–time metaphors, engage spatial reasoning, and create the impression of a conflict, in s-inconsistent ‘see’-items. Prevention of purely metaphorical interpretation can result in persistent ‘visual’ interpretation that identifies, e.g., the problems seen with visible objects (*Mountaineer Jack sees the difficult-to-cross crevice that lies behind him*). Even where the object (say, problem) is not identified as visual, the impression of a conflict between spatial implications from ‘see’ and the sequel will make s-inconsistent ‘see’-sentences feel ‘weird’ and lower their plausibility.

By contrast, the perceptual use of ‘aware of’ is clearly subordinate at least for laypeople and Other Philosophers (Sect. 4). Linguistic salience bias will therefore impede their suppression of initially triggered spatial inferences only from ‘see’, but not from ‘aware of’, and will not prevent purely metaphorical interpretation of s-inconsistent epistemic items with ‘aware’. Since s-inconsistent epistemic items are more plausible on the purely metaphorical interpretation that (according to the **SBH**) is unobtainable for ‘see’-sentences, laypeople and Other Philosophers will rate s-inconsistent epistemic ‘see’-items *less* plausible than their ‘aware’-counterparts—even though, on the contextually appropriate metaphorical interpretation, both mean the same (*Jack knows what problems he had in the past*). This sameness of meaning makes the effect size of this comparison a potential measure of the strength of linguistic salience bias.

The competing hypotheses making up LEO and LUO, respectively, make different predictions about cross-group comparisons. LEO’s first component, **H**<sub>1</sub>, claims

<sup>8</sup> The studies cited used the cancellation paradigm to document these inferences and provide evidence from pupillometry (Fischer & Engelhardt, 2020) and reading times (Fischer & Engelhardt, 2019). They also exclude various confounds (e.g., appropriate factive, rather than inappropriate spatial inferences from epistemic uses).

philosophers are better than laypeople at suppressing contextually cancelled default inferences. We can assess this claim without complications from linguistic salience bias by considering spatial inferences from ‘aware of’. These are cancelled by s-inconsistent sequels. The purely epistemic use of ‘aware of’ that is dominant in ordinary discourse facilitates ‘non-visual’ interpretations, which do not require current visual contact: Chuck is aware of the spot on the wall behind him because he has seen it earlier or been told about it. Jack is aware of the problems that lie behind him because he keeps being reminded of them. Etc. To the extent to which initial spatial inferences are suppressed, readers can adopt these ‘non-visual’ interpretations and feel no conflict. **H<sub>1</sub>** thus predicts that philosophers will deem s-inconsistent ‘aware’-items more plausible than undergraduates. The dominance of the perceptual use in specialist discourse in philosophy of perception could make suppression of spatial inferences and ‘non-visual’ interpretation more difficult for PoPs. This motivates restricting this prediction to Other Philosophers.

LEO’s **H<sub>2</sub>** claims that professional philosophers will be less susceptible to linguistic salience bias than undergraduates. Hence philosophers will be better able than undergraduates to suppress spatial inferences from epistemic uses of ‘see’. As a result, **H<sub>2</sub>** predicts, philosophers will deem s-inconsistent epistemic ‘see’-sentences more plausible than undergraduates, and the plausibility differential between these sentences and corresponding ‘aware’-sentences (as reflected by the effect size for this comparison) will be smaller for philosophers than undergraduates.

LUO’s **H<sub>3</sub>** claims that exposure to different usage patterns renders specialists susceptible to linguistic salience bias at different points and leads to non-suppression of inappropriate inferences from different words. For high-frequency words, differences between specialists and others will arise only from outright dominance reversals in specialist discourse. We observe a clear reversal for ‘aware of’, whose perceptual use is clearly dominant in the philosophy of perception corpus. By **H<sub>3</sub>**, this renders PoPs less able to suppress initial spatial inferences from epistemic uses of the verb and leads PoPs to find s-inconsistent items with these uses less plausible. This reduction in plausibility should show up in comparisons between groups benefiting from similar levels of suppression ability, that is, PoPs and Other Philosophers. **H<sub>3</sub>** thus predicts PoPs will judge s-inconsistent ‘aware’-items with abstract objects less plausible than Other Philosophers. Predictions are summed up in Table 5.

**Table 5** Predictions: hypothesis, relevant condition(s), predicted patterns of plausibility ratings

<b>H1</b>	S-INCON AWARE VISUAL S-INCON AWARE EPISTEMIC	Other Philosophers > UGs
<b>H2</b>	S-INCON SEE EPISTEMIC	Philosophers (PoP and Other) > UGs
<b>H3</b>	S-INCON AWARE EPISTEMIC	Other Philosophers > PoPs
<b>SBH</b>	S-INCON EPISTEMIC	At least Other Philosophers and UGs: AWARE > SEE

## 5.2 Methods

*Participants:* All participants self-identified as native speakers of English (that is, all other participants were excluded from analyses).

92 *undergraduate psychology students* (first and second year) from the University of East Anglia participated for course credit. Their mean age was 19.6 (SD = 2.89). 11 were male, 81 female.<sup>9</sup>

Academic philosophers were recruited through an electronic mailing list, a blog announcement, and personal emails to members of 14 UK philosophy departments and to individually targeted experts who had made sustained contributions, including recent contributions, to the pertinent debates in the philosophy of perception (namely, to the debates captured by our PHILO-P corpus) (see Online Appendix C for details).

72 academic philosophers with a PhD in philosophy were assigned to the group of *Other Philosophers* because they reported no research or teaching in philosophy of perception. Mean reported age was 43.6 (SD = 9.94). 49 were male, 23 female.

22 academic philosophers holding a PhD in philosophy were assigned to the *Philosophy of Perception (PoP)* group because they reported philosophy of perception as ‘main’ or ‘primary research area’ and at least ‘some’ teaching in the area. The small size of this sample (otherwise typical of studies in clinical psychology that examine rare mental health conditions) reflects the small size of the highly specialized population targeted. Mean reported age was 46.8 (SD = 11.33). 18 were male, 4 female.

*Materials:* We used 48 critical items: six for each of the eight conditions (illustrated by examples 1a–4b, Sect. 5.1). S-inconsistent epistemic items (like 4a/b) employed the cancellation phrases ‘that lie(s) behind him/her’ and ‘[that] s/he has turned from’, in equal number. There were 24 filler items. ‘See’ and ‘aware’ versions of items were rotated across two lists of materials, with approximately half of the participants completing each list. Participant instructions and critical items are provided by Online Appendix C.

*Design and Procedure:* In a  $2 \times 2 \times 2 \times 3$  design, context (s-consistent/s-inconsistent), verb (see/aware), and object (visual/epistemic) were manipulated within subject. Group (UG/PoP/ Other Philosophers) was between subject.

Participants read items online via Qualtrics and rated their plausibility on a 5-point Likert scale anchored at 1 with ‘very implausible’, at 3 with ‘neutral (neither plausible nor implausible)’, and at 5 with ‘very plausible’. Items were presented to each participant in random order. The main task was followed by demographic questions. In our analyses, we applied the Bonferroni-Holm correction to control for multiple comparisons associated with the several simple effects *t*-tests conducted (Armstrong, 2014; Cabin & Mitchell, 2000; Holm, 1979). The corrected significance thresholds are reported in square brackets after the relevant *p* values.

<sup>9</sup> The gender and age imbalances within and across our participant samples motivated correlational analyses of demographic factors, which excluded them as confounds (see Online Appendix D).

### 5.3 Results

To preview findings, results bore out predictions from the linguistic salience bias hypothesis **SBH** and from **H<sub>1</sub>**, but not from **H<sub>2</sub>** and **H<sub>3</sub>**. The most striking finding is that philosophers are better at deploying conceptual information than undergraduates (as per **H<sub>1</sub>**)—but this does not render them less susceptible even to the cognitive bias from which this ability seems most apt to shield them (*pace* **H<sub>2</sub>**). The findings speak against the linguistic expertise objection and fail to support the linguistic usage objection.

**SBH** predicts differences between conditions, within groups. Our key hypotheses **H<sub>1</sub>**–**H<sub>3</sub>** predict differences between groups. We report first global analyses that provide the statistical justification for comparisons between conditions and between groups—and first evidence pertaining to our hypotheses. We then report comparisons between conditions, within groups, and finally comparisons between groups that directly assess our key hypotheses.

#### 5.3.1 Global analyses

A  $2 \times 2 \times 2 \times 3$  (context  $\times$  verb  $\times$  object  $\times$  group) mixed-model ANOVA showed a significant four-way interaction  $F(2,183) = 4.03$ ,  $p = 0.019$ ,  $\eta^2 = 0.042$  and revealed main effects of context  $F(1,183) = 606.90$ ,  $p < 0.001$ ,  $\eta^2 = 0.76$ , verb  $F(1,183) = 142.11$ ,  $p < 0.001$ ,  $\eta^2 = 0.437$ , object  $F(1,183) = 20.82$ ,  $p < 0.001$ ,  $\eta^2 = 0.102$ , and group  $F(2,183) = 6.01$ ,  $p = 0.003$ ,  $\eta^2 = 0.062$ . See Fig. 1.

Three-way repeated measures ANOVAs confirmed significant three-way interactions for each group (psychology UGs:  $F(1,91) = 22.17$ ,  $p < 0.001$ ,  $\eta^2 = 0.192$ ; Other Philosophers:  $F(1,71) = 53.57$ ,  $p < 0.001$ ,  $\eta^2 = 0.43$ ; philosophers of perception:  $F(1,21) = 43.61$ ,  $p < 0.001$ ,  $\eta^2 = 0.68$ ) as well as main effects of context, verb, and object (see Table 6). Whereas plausibility ratings of all three groups were equally sensitive to the context manipulation, academic philosophers (of perception and others) were more sensitive than psychology undergraduates to differences in verb ('see' vs 'aware') and kind of object (visual vs epistemic), as evidenced by main effects of verb and object that are, respectively, three times and twice as large for philosophers than for psychology undergraduates. This is consistent with **H<sub>1</sub>** (cf. below).

#### 5.3.2 Comparisons between conditions

To decompose the interactions, we considered responses to items with visual and epistemic objects separately. Table 7 presents the results of these analyses, and the subsequent paired comparisons. Across all three groups, we observe the same pattern of significant differences across the board, including in the epistemic conditions where linguistic salience bias may assert itself: even though all three groups rated s-consistent epistemic items with 'see' and 'aware' equally plausible, all three groups deemed s-inconsistent epistemic items with 'see' *less* plausible than such items with 'aware' (for more detailed analyses, see Online Appendix D). This is evidence of linguistic salience bias (as per **SBH**) across all three groups. For philosophers (PoPs and others),

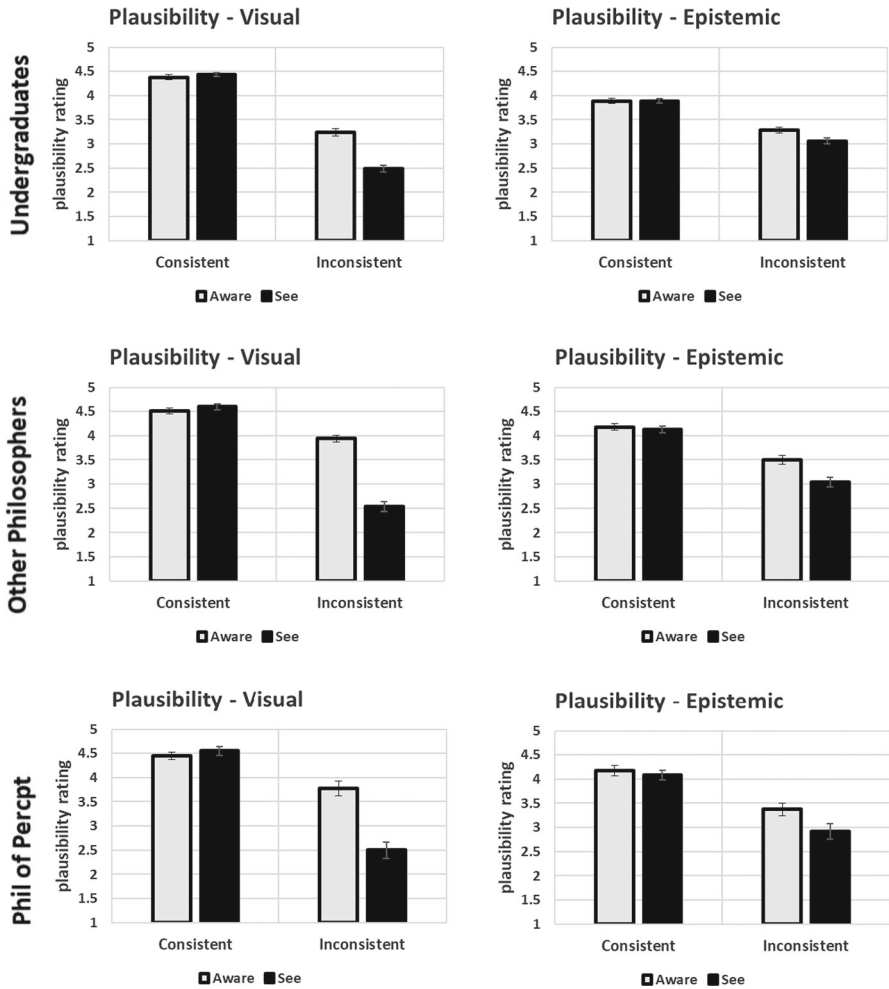


Fig. 1 Mean plausibility ratings per condition and group: psychology undergraduates (top), Other Philosophers (middle), and Philosophers of Perception (bottom). Error bars show the standard error of the mean

Table 6 Main effects per group

	Psychology UGs	Other philosophers	Philosophers of perception
Context	$F(1,91) = 449.52$ $p < .001, \eta^2 = .832$	$F(1,71) = 268.36$ $p < .001, \eta^2 = .791$	$F(1,21) = 221.41$ $p < .001, \eta^2 = .913$
Verb	$F(1,91) = 27.81$ $p < .001, \eta^2 = .234$	$F(1,71) = 188.76,$ $p < .001, \eta^2 = .727$	$F(1,21) = 60.34$ $p < .001, \eta^2 = .742$
Object	$F(1,91) = 8.08$ $p = .006, \eta^2 = .082$	$F(1,71) = 13.12,$ $p = .001, \eta^2 = .156$	$F(1,21) = 4.73$ $p = .041, \eta^2 = .184$

**Table 7** Inferential analysis with Holm threshold (in square brackets) and effect sizes (in parentheses:  $\eta^2$  for interactions, Cohen's  $d$  for  $t$  tests)

	Visual objects	Epistemic objects
<b>Psychology students</b>		
Interaction	$F(1,91) = 44.61, p < .001 (.67)$	$F(1,91) = 11.95, p = .001 (.12)$
Aware-Con. versus Aware-Incon	$t(91) = 12.81, p < .001 [.002]$	$t(91) = 10.00, p < .001 [.002]$
See-Con. versus See-Incon	$t(91) = 22.46, p < .001 [.002]$	$t(91) = 10.32, p < .001 [.002]$
Aware-Con. versus See-Con	$t(91) = -1.82, p = .072 [.01]$	$t(91) = -.17, p = .87 [.05]$
Aware-Incon. versus See-Incon	$t(91) = 6.50, p < .001 [.002] (.67)$	$t(91) = 3.81, p < .001 [.002] (.40)$
<b>Other philosophers</b>		
Interaction	$F(1,71) = 173.02, p < .001 (.78)$	$F(1,71) = 19.16, p < .001 (.21)$
Aware-Con. versus Aware-Incon	$t(71) = 6.97, p < .001 [.002]$	$t(71) = 7.19, p < .001 [.002]$
See-Con. versus See-Incon	$t(71) = 16.54, p < .001 [.002]$	$t(71) = 9.91, p < .001 [.002]$
Aware-Con. versus See-Con	$t(71) = -2.64, p = .01 [.008]$	$t(71) = 1.32, p = .19 [.025]$
Aware-Incon. versus See-Incon	$t(71) = 13.30, p < .001 [.002] (1.56)$	$t(71) = 5.85, p < .001 [.02] (.69)$
<b>Philosophers of perception</b>		
Interaction	$F(1,21) = 74.86, p < .001 (.78)$	$F(1,21) = 6.66, p < .001 (.24)$
Aware-Con. versus Aware-Incon	$t(21) = 5.24, p < .001 [.002]$	$t(21) = 5.36, p < .001 [.002]$
See-Con. versus See-Incon	$t(21) = 11.67, p < .001 [.002]$	$t(21) = 6.35, p < .001 [.002]$
Aware-Con. versus See-Con	$t(21) = -1.89, p = .073 [.013]$	$t(21) = 1.39, p = .18 [.017]$
Aware-Incon. versus See-Incon	$t(21) = 9.13, p < .001 [.002] (1.95)$	$t(21) = 3.60, p < .001 [.002] (.78)$

we observed a medium effect of the verb manipulation ('see' vs 'aware') in the s-inconsistent epistemic condition. This effect was larger (rather than smaller) than for undergraduates. These two findings provide first evidence against  $H_2$ . For philosophers (PoPs and others), we further observed a large effect of the verb manipulation in the s-inconsistent visual condition. This effect was larger than the (medium) effect for undergraduates, due to higher ratings for 'aware'-items (Fig. 1). These findings are consistent with  $H_1$ .

### 5.3.3 Comparisons between groups

We finally made comparisons between groups. There was little variability across groups in the s-consistent conditions. To assess the key predictions from  $H_1$ – $H_3$  (summed up in Table 5), we examined the s-inconsistent conditions. Figure 2 displays the means for ease of comparisons.

*S-inconsistent visual condition.* A  $2 \times 3$  mixed model (verb  $\times$  group) ANOVA showed a significant interaction  $F(2,183) = 9.15, p < 0.001, \eta^2 = 0.09$ . Consistent with  $H_1$ , follow-up independent-samples t-tests revealed that philosophers deemed ‘aware’-items more plausible than psychology undergraduates (Other Philosophers:  $t(162) = -6.07, p < 0.001$  [0.0083]; PoPs:  $t(112) = -2.97, p = 0.004$  [0.01]). There were no significant differences between the two philosophy groups  $t(92) = 0.95, p = 0.345$  [0.0125]. There were also no significant differences between the three groups’ plausibility judgments concerning ‘see’-items (Psychology UGs vs. Other Philosophers:  $t(162) = -0.397, p = 0.692$  [0.0167]; Psychology UGs vs. PoPs:

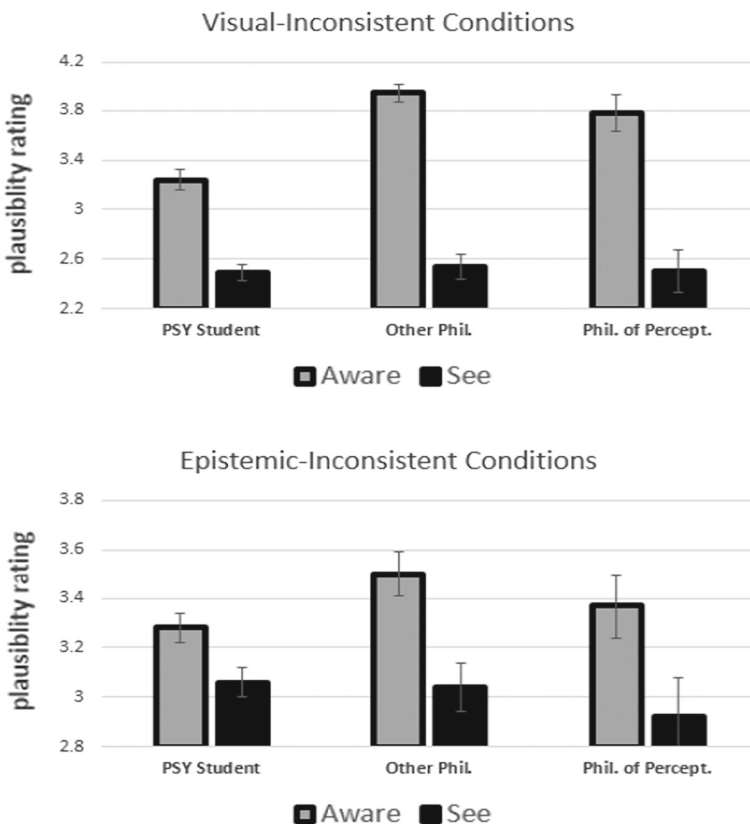


Fig. 2 Mean plausibility ratings per group in s-inconsistent conditions. Error bars show the standard error of the mean



$t(112) = -0.062, p = 0.95 [0.05]$ ; Other Philosophers vs. PoPs:  $t(92) = 0.178, p = 0.859 [0.025]$ ).

*S-inconsistent epistemic condition.* A  $2 \times 3$  mixed model (verb  $\times$  group) ANOVA showed a significant interaction  $F(2,183) = 3.63, p = 0.029, \eta^2 = 0.038$ . Independent samples t-tests examined whether our three groups gave different ratings to ‘see’- and ‘aware’-items, respectively. *Pace H<sub>2</sub>*, there were no significant group differences in ratings of ‘see’-items (UGs vs Other Philosophers:  $t(162) = 0.170, p = 0.856 [0.05]$ ; UGs vs PoPs:  $t(112) = 0.867, p = 0.388 [0.01]$ ; PoPs vs Other Philosophers:  $t(92) = 0.591, p = 0.556 [0.025]$ ). *Pace H<sub>3</sub>*, philosophers of perception did not significantly differ from other philosophers in their ratings of ‘aware’-items  $t(92) = 0.717, p = 0.475 [0.0125]$ . Nor did they differ significantly from psychology undergraduates  $t(112) = -0.582, p = 0.562 [0.0167]$ . Qualifying the above evidence for *H<sub>1</sub>*, the difference in ‘aware’-ratings between undergraduates and other philosophers remained shy of even marginal significance upon correction for multiple comparisons  $t(162) = -1.97, p = 0.051 [0.0083]$ .

To sum up: For all three groups, we found response patterns predicted by the linguistic salience bias hypothesis *SBH*. For all three groups, we thus replicated findings from previous studies with undergraduate participants (Fischer & Engelhardt, 2017, 2019, 2020). Previous studies combined eye tracking with plausibility ratings, in a laboratory setting; replication with a new online delivery format strengthens support for the *SBH*. Distributional semantic analysis (Sect. 4.2) provided further evidence, suggesting that all participants should be able to identify the need for suppression in epistemic contexts, so that present findings evidence the inability to suppress contextually inappropriate default inferences that the *SBH* predicts. Online Appendix D provides further analyses assessing this hypothesis. We now discuss *H<sub>1</sub>–H<sub>3</sub>* in connection with the competing objections they motivate.

## 6 Philosophical conclusions

### 6.1 Assessing the linguistic expertise objection (LEO)

The linguistic expertise objection makes three assumptions (Sect. 2): (1) Philosophers are better than laypeople at contextualizing conceptual information, that is, at complementing and suppressing default information, as appropriate. (2) Better contextualization ability renders philosophers’ interpretations of case descriptions less susceptible to comprehension biases and, thereby, less sensitive to irrelevant factors (e.g., framing and order effects). (3) This also makes philosophers’ interpretation of case descriptions more accurate. LEO infers from these assumptions that philosophers’ intuitive judgments about verbally described cases are more stable (i.e., less susceptible to biases and irrelevant factors) and more accurate.

We put assumption (1) to the test by examining the hypothesis *H<sub>1</sub>* that philosophers are better at suppressing default inferences, where these are contextually irrelevant. Our findings were largely consistent with *H<sub>1</sub>*: Academic philosophers’ item ratings were more sensitive than psychology undergraduates’ to differences in verb (‘see’ vs ‘aware’) and object (visual vs epistemic) (Table 6), suggesting better ability to

integrate default information activated by verb and object-noun. The best test for  $H_1$  is provided by cross-group comparisons of ratings for s-inconsistent ‘aware’-items, which require suppression of default inferences but involve no complications from linguistic salience bias (Sect. 5.1). Academic philosophers gave higher ratings than psychology undergraduates to s-inconsistent ‘aware’-items with visual objects. This suggests they were better at winning through to a ‘purely epistemic’ interpretation of such items. However, the predicted difference between Other Philosophers and undergraduates in ratings for s-inconsistent ‘aware’-items with epistemic objects, while numerically notable, remained shy of even marginal significance upon correction for multiple comparisons. These findings offer qualified support for the hypothesis  $H_1$  that philosophers are better than laypeople at suppressing contextually irrelevant default inferences.

We put LEO’s assumption (2) to the test by examining the hypothesis  $H_2$  that academic philosophers differ from psychology undergraduates in being less susceptible to the linguistic salience bias. This comprehension bias is most apt to be mitigated by better suppression ability and affects polysemy interpretation—at which analytic philosophers can be plausibly thought to excel (Sect. 3.2). A potential measure of susceptibility to the bias is the effect size of the comparison between ratings for ‘aware’- and ‘see’-items in the s-inconsistent epistemic condition: These sentences mean the same, on the intended purely metaphorical interpretation. Even so, we observed a medium-sized effect for philosophers, which was almost twice as large as for undergraduates (Table 7). However, the larger effect size for philosophers is primarily due to philosophers giving higher ‘aware’ ratings, rather than lower ‘see’ ratings. Higher ratings for s-inconsistent ‘aware’ items are promoted by pragmatic inferences (Manner inferences with the M-heuristic, see Levinson, 2000, pp. 136–137): Preference of the marked expression ‘aware of’ over the simpler alternative ‘see’, rendered salient by our materials, suggests that the situation talked about deviates from the *seeing*-stereotype associated with the simpler alternative. This inference supports suppression of contextually inappropriate default inferences from ‘aware’ (see Sect. 3.1). We interpret philosophers’ larger effect size as indicative of better pragmatic inferencing skills, rather than worse inhibition. On this interpretation, our findings do not show that philosophers are more susceptible to linguistic salience bias than undergraduates. However, our groups’ equally low ‘see’ ratings do show that philosophers are no less susceptible to the bias.

To test LEO’s third assumption, that better suppression ability will render philosophers’ interpretations of case descriptions more accurate, we consider ratings for items with epistemic objects, whose interpretation requires suppression of default inferences from either of our two verbs. The intended interpretation of these items is made explicit by knowledge attributions like, e.g., ‘Jack knows what problems he had in the past’. An earlier study (Fischer & Engelhardt, 2020, Online Appendix A) elicited plausibility ratings for these knowledge attributions from psychology undergraduates. The attributions were rated distinctly plausible (mean rating 4.03, SD = 0.37). We can use this mean rating as a norm of accuracy, to assess present ratings in the relevant (s-inconsistent epistemic) conditions (Fig. 2). Ratings for ‘see’-sentences did not differ between groups, all means were neutral (not significantly above mid-point 3; see Online Appendix D), and thus equally inaccurate. Mean ratings for ‘aware’-sentences

did not significantly differ between groups, either. They were significantly above mid-point ‘3’ for undergraduates and Other Philosophers (while the small sample size prevented ratings from Philosophers of Perception to take this hurdle, if by a whisker, upon correction for multiple comparisons; see Online Appendix D). Philosophers’ mean ratings were merely numerically closer to 4. All groups made the same judgments about these items (deemed them plausible)—philosophers just did so slightly more emphatically, getting closer to our accuracy norm.

This largely refutes LEO: Philosophers do seem better at suppressing inappropriate default inferences (as per  $H_1$ ), at any rate where these inferences are not supported by linguistic salience bias. As a result, philosophers do make some slightly more accurate plausibility assessments. However (*pace*  $H_2$ ), philosophers are no less susceptible to the linguistic salience bias than undergraduates, and the judgments affected by this bias are not more accurate when coming from philosophers rather than undergraduates. While  $H_1$  would benefit from further support, present findings suggest a striking conclusion: Philosophers’ likely better ability to deploy conceptual information does not render them less susceptible even to the cognitive bias from which this ability seems most apt to shield them.

## 6.2 Assessing the linguistic usage objection (LUO)

The key finding, that (*pace*  $H_2$ ) philosophers are no less susceptible to the linguistic salience bias than laypeople, secures the starting point of the new linguistic usage objection (Sect. 3.3). The finding entails that pronounced salience imbalances arising from the dominant use of ordinarily subordinate uses of words, in specialist discourse, can lead specialist philosophers to go along with inappropriate default inferences from those words, which laypeople (and philosophers with other specializations) avoid (as per  $H_3$ ).

Our study examined this possibility by considering inferences from ‘aware of’, which is predominantly given a perceptual use in key debates in the philosophy of perception, while a non-perceptual, purely epistemic use is dominant in ordinary discourse. However, we did not find any evidence of persistent spatial inferences from non-perceptual uses, in the judgments of philosophers of perception who engage with the relevant debates extensively enough for their linguistic exposure patterns to be affected by them. The small size of this highly specialist population was reflected in the small size of our sample. This places a caveat on our findings. We did observe the pattern of numeric results predicted by  $H_3$ : Philosophers of perception gave s-inconsistent epistemic items with ‘aware’ mean ratings that were numerically lower than mean ratings from other philosophers with arguably equal conceptual competence but different linguistic diet; but the difference remained so far shy of significance that even a sample comprising the entire specialist population of interest is highly unlikely to produce a significant difference (Sect. 5.3).

The relevant PHILO-P corpus contained only 375 occurrences of ‘aware of’ among its 1 million words—less than a fifth as many occurrences as ‘see’. This suggests that despite its prominence in the targeted debates, ‘aware of’ may still be used too infrequently in the philosophy of perception for its use in this specialist discourse to

influence overall relative exposure frequencies of the high-frequency verb (Foraker & Murphy, 2012). Even philosophers of perception contributing to and teaching the relevant debates will be exposed to the word more often in ordinary or generic academic discourse, where the verb's purely epistemic use dominates. Despite the dominance reversal in specialist debates, they will overall encounter the word more in its purely epistemic use—like other philosophers and laypeople. Hence they are no worse at suppressing inappropriate perception-related (spatial) inferences from the verb.

Present findings suggest that common words must be used very frequently in specialist discourse, for even an outright dominance reversal to create new vulnerabilities to inappropriate default inferences. This considerably narrows the scope of the linguistic usage objection: For new vulnerabilities to be created, it is not enough that an irregular polyseme has a different dominant use in specialist discourse—specialists must also use the word very frequently, in such discourse. This restricts LUO to rather few plausible candidates, like the verb 'to know'.<sup>10</sup> By largely refuting LEO and mitigating LUO, our findings defang the 'master argument' against experimental philosophy's lay-expert inference (Sect. 3.3). They reduce principled objections to local difficulties.

### 6.3 Main findings and methodological consequences

In summary, we found that professional philosophers are better at deploying conceptual information than laypeople (psychology undergraduates): they are better at suppressing contextually irrelevant default inferences from words. Even so, philosophers are no less susceptible to the cognitive bias this competence seems most apt to shield them from, viz., the linguistic salience bias. This comprehension bias allows contextually inappropriate default inferences to influence utterance interpretation and further cognition. It does so under conditions which frequently recur in philosophy (Sect. 3.2): where unbalanced polysemous words are used in a subordinate sense, to talk about cases that pull apart features that go together, in the associated stereotype. Neither the observed difference in conceptual competence nor marked differences in linguistic usage between expert and ordinary discourse lead this bias to result in notable differences between lay and expert judgments. Since this comprehension bias is the bias most likely affected by the examined difference in conceptual competence, it seems unlikely that the observed difference in competence will render philosophers less susceptible than laypeople to *any* cognitive bias and result in markedly different case judgments. In a nutshell, philosophers' better conceptual competence does not make their judgments more stable or greatly more accurate than those of laypeople.

Present findings have productive methodological consequences for experimental philosophy. First, they support experimental philosophy's lay-expert inference in the face of linguistic expertise and usage objections—the arguably most promising versions of the expertise objection (see Sects. 2–3). Present findings refute these objections in a perhaps unexpected way. Expertise objections assume there is a difference in

<sup>10</sup> This verb displays marked frequency differences between different uses (Hansen et al., 2021), is used differently by philosophers than by laypeople and other academics (Horvath & Wiegmann, 2016; Starmans & Friedman, 2020), and stands to be frequently used in epistemology.

expertise or competence between philosophers and laypeople, and that this difference makes philosophers' case judgments less susceptible to cognitive biases and irrelevant factors (Sect. 2). Present findings provide some evidence of potentially relevant differences, but reveal these differences need not make a difference: We found some evidence of differences in conceptual competence between philosophers and laypeople, and documented a difference in linguistic diet; but these differences did not translate into different susceptibility to even the most pertinent cognitive bias, or render philosophers' judgments appreciably more accurate. This suggests that contributions to 'negative', 'restrictionist', or 'evidential' experimental philosophy can work with lay participants to assess claims about the stability and accuracy of philosophers' judgments.

Second, present findings open up new avenues for these related research programs. Contributions to 'negative' and 'restrictionist' experimental philosophy have elicited sensitivity to order and framing effects (reviews: Machery, 2017; Mallon, 2016). Linguistic salience bias explains contextually inappropriate inferences that lead to framing effects (like 'see' vs 'aware of'). For example, when laypeople are asked to imagine philosophical zombies that have bodies like ours and behave like us, but where 'all is dark inside', twice as many people accept that the imagined beings lack conscious experience when these beings are described as 'zombies', rather than 'duplicates', and this framing effect is explained by linguistic salience bias (Fischer & Sytsma, 2021). Indeed, given that the bias asserts itself under conditions that frequently recur in philosophy, it is arguably a major source of philosophically relevant framing effects. The advance from eliciting to explaining (some) framing effects facilitates a move from purely negative to more specific and constructive findings: The mere elicitation of such effects allows us to infer only that intuitive judgments about the topic at issue are unreliable (cf. Machery, 2017, pp. 77–85). By contrast, explanations of case judgments that invoke the linguistic salience bias allow us to adjudicate between judgments elicited by different frames, and identify biasing and non-biasing frames. The finding that the linguistic salience bias affects philosophers and laypeople equally means that psycholinguistic findings about this comprehension bias can be deployed for the restrictionist purpose of identifying conditions under which philosophers may (not) trust their intuitions (e.g., Weinberg, 2015).

Moreover, the finding allows evidential experimental philosophy to expand its philosophical remit, and assess not only case judgments in thought experiments but also verbal reasoning in philosophical argument. Psychological findings about how cognitive biases affect verbal reasoning help expose previously undetected fallacies. A number of studies with lay participants followed up the suggestion that linguistic salience bias leads to previously undetected fallacies of equivocation, for example, in philosophical arguments about perception: arguments 'from illusion' and 'from hallucination' rely on default inferences from special ('phenomenal') uses of appearance- and perception- verbs that are licensed only by their dominant sense and cancelled by the sentence or discourse context (Fischer & Engelhardt, 2016, 2017, 2020; Fischer et al., 2021a, 2021b). These and other philosophical arguments have been advanced mainly by professional philosophers. The finding that professional philosophers are no less susceptible to linguistic salience bias than laypeople provides the necessary empirical foundation for this extension of evidential experimental philosophy.

Present findings also have ultimately productive methodological consequences for philosophical thought experimentation. The debate about whether expertise renders philosophers' case judgments immune to factors that vitiate lay judgments have developed into a more wide-ranging debate about the soundness of the method of cases—specifically, about whether non-accidental features of this method systematically undermine the reliability of both lay and expert judgments. A focus of debate has been the 'esotericity' of the cases considered (e.g., Cappelen, 2012; Machery, 2017; Weinberg, 2015; Williamson, 2016). To test modal implications of philosophical theories, thought experiments must consider cases that are *unusual* (which we hardly, if ever, observe or read/hear about); to adjudicate between competing theories that agree about typical cases, they must consider cases that *pull apart features that typically go together* (Machery, 2017, pp. 113–120). Critics of the method suggest that these features promote unreliability in both lay and expert judgments (ibid.). In the only study to date to specifically address this suggestion, Schindler and Saint-Germier (2020) examined thought experiments from physics involving cases with these two 'disturbing' features. They found a clear majority of expert physicists and laypeople made correct judgments about five of six cases presented.<sup>11</sup> These first findings suggest that, to be viable, the criticism of the method of cases needs to be developed through causal hypotheses that propose specific links between the two disturbing features and unreliability.

Present findings motivate such hypotheses. To describe cases which pull apart features that typically go together, philosophers frequently fall back on familiar words associated with a stereotype that combines those typically co-occurring features (e.g., 'see' for cases of hallucination). Where they cannot fall back on an established subordinate sense of the word, philosophers will create a new special use. Either way, the interpretation of the word ('see') in the description of the case (hallucination) requires suppression of the automatically activated typical feature (*object of sight is in front of the viewer*) that has been 'pulled away' and cancelled by contextual information (e.g., the information that the protagonist hallucinates). Where this happens, linguistic salience bias is liable to arise (Sect. 3.2). Readers of the case description are then prone to only partially suppress the irrelevant feature and integrate it to some extent into the situation model that informs further judgement and reasoning about the case (cf. Sect.3.1). The case judgment of interest will be unduly influenced by the cancelled feature that judges are meant to set aside. Linguistic salience bias can thus affect judgments about cases that pull apart typically co-occurring features, and render the judgments unreliable. Our key finding reveals this problem arises to the same extent for laypeople and philosophers. The problem may be exacerbated where cases are also unusual (like hallucination): People (including philosophers) tend to know little about unusual cases. Suppression of contextually irrelevant default information is aided by integration with background knowledge (Fischer & Engelhardt, 2017) activated by discourse context (Metusalem et al., 2012). Where cases are unusual, paucity of background knowledge makes it more likely that irrelevant default information remains unsuppressed and unduly influences judgments of laypersons and experts alike.

<sup>11</sup> Physicists outperformed laypeople on half the cases (Schindler & Saint-Germier, 2020, p.2679), but, strikingly, not those requiring better knowledge of physics (pp.2683–4). Findings thus fail to support the expertise objection.

Insights into specific sources of the problem are productive, as they allow us to work around the problem. Where judgments are rendered unreliable by linguistic salience bias, we can rephrase case descriptions so that they do not trigger inappropriate default inferences we cannot suppress: In describing cases that pull apart typically co-occurring features, thought experimentalists need to avoid words whose dominant sense is associated with an ‘unhelpful’ stereotype that comprises the typically co-occurring features the thought experiment pulls apart (e.g., the zombie stereotype comprises both *lack of conscious experience* and *attacks and eats humans*; Fischer & Sytsma, 2021). Rather, they need to find descriptions that do not trigger contextually irrelevant inferences (e.g., ‘physical duplicate that lacks conscious experience’). Since the linguistic salience bias arises only where a polyseme has a clearly dominant sense (Sect. 3.2), it may occasionally also be viable to recruit a balanced polyseme whose main sense is associated with an unhelpful stereotype but is not clearly dominant.

As we have seen above, extensions of restrictionist experimental philosophy can lead to insights into the sources of judgment unreliability. Such insights allow thought experimentalists to avoid the pinpointed pitfalls by developing suitable case descriptions. Further empirical study of the sources of unreliability will reveal to what extent the method of cases remains viable. In any case, thought experiments will need to become more similar to psychological experiments: The development of suitable case descriptions requires preliminary work of the sort standardly involved in developing materials for psychology experiments. To guard against linguistic salience bias, for example, thought experimentalists need to explore word-related stereotypes (e.g., through listing and sentence completion tasks, cf. McRae et al., 1997) or examine relative occurrence frequencies of different senses (Fischer & Engelhardt, 2020, pp. 434–435). Present findings show that philosophers need to take these precautions also when developing case descriptions for their own benefit. Just like psychological experiments, philosophical thought experiments require some empirical preparation—also when conducted by expert philosophers.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11229-022-03487-3>.

**Acknowledgements** For helpful comments on previous versions, we are indebted to Boudewijn de Bruin, Alex Wiegmann, Joachim Horvath, two anonymous reviewers for this journal, and an audience at the European Online Conference in Experimental Philosophy, 2020. We thank Mark Curtis and Ian Hare for annotating the corpora of interest. We would also like to express our gratitude to the colleagues who participated in our study.

**Authors’ contributions** EF and PE contributed in equal parts to material development and interpretation of results for the main study. Engelhardt gathered and analyzed the data. AH analyzed the data from the corpus studies and conducted the distributional semantics analysis. Fischer undertook the philosophical and all remaining research and drafted the paper.

**Funding** This research was funded through QR funding by the University of East Anglia.

**Availability of data and materials** The authors undertake to make data and materials available upon request.

**Code availability** The code used for distributional semantics analysis is available at <https://github.com/minimalparts/Perception>.



## Declarations

**Ethical approval** The use of human research participants was approved by the Research Ethics Committee of the University of East Anglia's School of Psychology. The research conformed to the ethical standards for conducting research as outlined by the British Psychological Society.

**Conflict of interest** The authors are unaware of conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Psychological Optics*, *34*, 502–508.
- Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences*, *16*, 593–605.
- Beebe, J. R., & Monaghan, J. (2018). Epistemic closure in folk epistemology. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 2, pp. 38–70). Oxford: OUP.
- Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, *63*, 489–505.
- Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13*, 185–188.
- Bottini, R., Crepaldi, D., Casasanto, D., Crollen, V., & Collignon, O. (2015). Space and time in the sighted and blind. *Cognition*, *141*, 67–72.
- Brocher, A., Foraker, S., & Koenig, J. P. (2016). Processing of irregular polysemes in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1798–1813.
- Brocher, A., Koenig, J.-P., Mauner, G., & Foraker, S. (2018). About sharing and commitment: The retrieval of biased and balanced irregular polysemes. *Language, Cognition and Neuroscience*, *33*(4), 443–466.
- Butterfuss, R., & Kendeou, P. (2018). The role of executive functions in reading comprehension. *Educational Psychology Review*, *30*, 801–826.
- Byrd, R. J., et al. (1987). Tools and methods for computational lexicology. *Computational Linguistics*, *13*, 219–240.
- Cabin, R. J., & Mitchell, R. J. (2000). To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America*, *81*, 246–248.
- Cappelen, H. (2012). *Philosophy without intuitions*. Oxford: OUP.
- Carter, J. A., Peterson, M., & van Bezooijen, B. (2016). Not knowing a cat is a cat: Analyticity and knowledge ascriptions. *Review of Philosophy and Psychology*, *7*(2), 817–834.
- Carter, J. A., Pritchard, D., & Shepherd, J. (2019). Knowledge-how, understanding-why and epistemic luck: An experimental study. *Review of Philosophy and Psychology*, *10*(4), 701–734.
- Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*, 579–593.
- Cattell, R. B. (1987). *Intelligence: Its structure*. Elsevier.
- De Bruin, B. (2021). Saving the armchair by experiment: What works in economics doesn't work in philosophy. *Philosophical Studies*, *178*, 2483–2508.
- Dempster, F. N. (1991). Inhibitory processes: A neglected dimension of intelligence. *Intelligence*, *15*, 157–173.
- Devitt, M. (2011). Experimental semantics. *Philosophy and Phenomenological Research*, *82*, 418–435.

- Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: The current state of the literature. *Psychonomic Bulletin & Review*, 22(1), 13–37.
- Egler, M., & Ross, L. D. (2020). Philosophical expertise under the microscope. *Synthese*, 197, 1077–1098.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33, 547–582.
- Engelhardt, P. E., Nigg, J. T., Carr, L. A., & Ferreira, F. (2008). Cognitive inhibition and working memory in attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology*, 117, 591–605.
- Faust, M. E., & Gernsbacher, M. A. (1996). Cerebral mechanisms for suppression of inappropriate information during sentence comprehension. *Brain and Language*, 53, 234–259.
- Ferretti, T., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44, 516–547.
- Fischer, E. (2020). Conceptual control. On the feasibility of conceptual engineering. *Inquiry*, 2020, 1–29. <https://doi.org/10.1080/0020174X.2020.1773309>
- Fischer, E., & Engelhardt, P. E. (2016). Intuitions' linguistic sources: Stereotypes, intuitions, and illusions. *Mind and Language*, 31, 67–103.
- Fischer, E., & Engelhardt, P. E. (2017). Stereotypical inferences: Philosophical relevance and psycholinguistic toolkit. *Ratio*, 30, 411–442.
- Fischer, E., & Engelhardt, P. E. (2019). Eyes as windows to minds: Psycholinguistics for experimental philosophy. In E. Fischer & M. Curtis (Eds.), *Methodological advances in experimental philosophy* (pp. 43–100). Bloomsbury.
- Fischer, E., & Engelhardt, P. E. (2020). Lingering stereotypes: Salience bias in philosophical argument. *Mind and Language*, 35, 415–439.
- Fischer, E., Engelhardt, P. E., Horvath, J., & Ohtani, H. (2021a). Experimental ordinary language philosophy: A cross-linguistic study of defeasible default inferences. *Synthese*, 198, 1029–1070.
- Fischer, E., Engelhardt, P. E., & Sytsma, J. (2021b). Inappropriate stereotypical inferences? An adversarial collaboration in experimental ordinary language philosophy. *Synthese*, 198, 10127–10168.
- Fischer, E., & Sytsma, J. (2021). Zombie intuitions. *Cognition*, 215, e104807.
- Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language*, 67, 407–425.
- Friedman, N. P., et al. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172–179.
- Garrett, M., & Harnish, R. M. (2007). Experimental pragmatics: Testing for implicatures. *Pragmatics and Cognition*, 15, 65–90.
- Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space time metaphors. *Language and Cognitive Processes*, 17, 537–565.
- Giora, R. (2003). *On our mind. Salience, context, and figurative language*. Oxford: OUP.
- Giora, R., Fein, O., Aschkenazi, K., & Alkabets-Zlozover, I. (2007). Negation in context: A functional approach to suppression. *Discourse Processes*, 43, 153–172.
- Glenberg, A., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558–565.
- Hales, S. (2006). *Relativism and the foundations of philosophy*. Cambridge, Mass.: MIT Press.
- Hampton, J. (2006). Concepts as prototypes. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 79–113). Elsevier.
- Hansen, N., Porter, J. D., & Francis, K. (2021). A corpus study of “know”: On the verification of philosophers' frequency claims about language. *Episteme*, 18(2), 242–268. <https://doi.org/10.1017/epi.2019.15>.
- Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, 111, 151–167.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 106, 587–612.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Horvath, J. (2010). How (not) to react to experimental philosophy. *Philosophical Psychology*, 23, 447–480.
- Horvath, J., & Wiegmann, A. (2016). Intuitive expertise and intuitions about knowledge. *Philosophical Studies*, 10, 2701–2726.
- Horvath, J., & Wiegmann, A. (2021). Intuitive expertise in moral judgements. *Australasian Journal of Philosophy*. <https://doi.org/10.1080/00048402.2021.1890162>
- Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations*, 10, 95–118.

- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.
- Klepousniotou, E., Pike, B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. *Brain and Language*, *123*, 11–21.
- Kneer, M., Colaço, D., Alexander, J., & Machery, E. (2021). On second thought: A refutation of the reflection defense. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (vol. 4, pp. 257–296). Oxford: OUP.
- Levinson, S. C. (2000). *Presumptive meanings*. Cambridge, Mass.: MIT Press.
- Livengood, J., Sytma, J., & Rose, D. (2017). Following the FAD: Folk attributions and theories of actual causation. *Review of Philosophy and Psychology*, *8*, 274–294.
- Livengood, J., & Sytma, J. (2020). Actual causation and compositionality. *Philosophy of Science*, *87*, 43–69.
- Löhr, G. (2019). The experience machine and the expertise defense. *Philosophical Psychology*, *32*, 257–273.
- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*, 618–630.
- Ludwig, K. (2007). The epistemology of thought experiments: First-person versus third-person approaches. *Midwest Studies in Philosophy*, *31*(1), 128–159.
- MacGregor, L. J., Bouwsema, J., & Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. *Neuropsychologia*, *68*, 126–138.
- Machery, E. (2009). *Doing without concepts*. Oxford: OUP.
- Machery, E. (2012). Expertise and intuitions about reference. *Theoria*, *27*, 37–54.
- Machery, E. (2017). *Philosophy within its proper bounds*. Oxford: OUP.
- Mallon, R. (2016). Experimental philosophy. In H. Cappelen, T. Szabo Gendler, & J. Hawthorne (Eds.), *Oxford handbook of philosophical methodology* (pp. 410–433). Oxford: OUP.
- Matsuki, K., et al. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *37*, 913–934.
- McRae, K., Ferretti, T. R., & Amyote, I. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*, 137–176.
- McRae, K., Hare, M., Elman, J. L., & Ferretti, T. R. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, *33*, 1174–1184.
- McRae, K., & Jones, M. (2013). Semantic memory. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology*. Oxford: OUP.
- Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: The role of rhythm and meaning. *Annals of the New York Academy of Sciences*, *682*, 272–282.
- Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*, 545–567.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14.
- Miyake, A., et al. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- Nado, J. (2014). Philosophical expertise. *Philosophy Compass*, *9*, 631–641.
- Schindler, S., & Saint-Germier, P. (2020). Are thought experiments “‘disturbing’”? The case of armchair physics. *Philosophical Studies*, *177*, 2671–2695.
- Schindler, S., & Saint-Germier, P. (forthcoming). Philosophical expertise put to the test. *Australasian Journal of Philosophy*.
- Schulz, E., Cokely, E. T., & Feltz, A. (2011). Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. *Consciousness and Cognition*, *20*, 1722–1731.
- Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language*, *27*, 135–153.
- Schwitzgebel, E., & Cushman, F. (2015). Philosophers’ biased judgments persist despite training, expertise and reflection. *Cognition*, *141*, 127–137.
- Slooman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189–228.
- Stanovich, K. E. (1993). Does reading make you smarter: Literacy and the development of verbal intelligence. *Advances in Child Development and Behavior*, *24*(133), 180.

- Starmans, C., & Friedman, O. (2020). Expert or esoteric? Philosophers attribute knowledge differently than all other academics. *Cognitive Science*, *44*, e12850.
- Sytsma, J., & Machery, E. (2010). Two conceptions of subjective experience. *Philosophical Studies*, *151*, 299–327.
- Tobia, K., Buckwalter, W., & Stich, S. (2013). Moral intuitions: Are philosophers experts? *Philosophical Psychology*, *26*, 629–638.
- Tobia, K., Chapman, G., & Stich, S. (2013). Cleanliness is next to morality, even for philosophers. *Journal of Consciousness Studies*, *20*, 195–204.
- Turri, J. (2013). A conspicuous art: Putting Gettier to the test. *Philosophers' Imprint*, *13*, 1–16.
- Vaesen, K., Peterson, M., & van Bezooijen, B. (2013). The reliability of armchair intuitions. *Metaphilosophy*, *44*, 559–578.
- Vicente, A. (2018). Polysemy and word meaning: An account of lexical meaning for different kinds of content words. *Philosophical Studies*, *175*, 947–968.
- Weinberg, J. (2015). Humans as instruments, on the inevitability of experimental philosophy. In E. Fischer & J. Collins (Eds.), *Experimental philosophy, rationalism, and naturalism* (pp. 171–187). Routledge.
- Weinberg, J., Gonnerman, C., Buckner, C., & Alexander, J. (2010). Are philosophers expert intuiters? *Philosophical Psychology*, *23*, 331–355.
- Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*, 311–334.
- Wiegmann, A., Horvath, J., & Meyer, K. (2020). Intuitive expertise and irrelevant options. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford Studies in experimental philosophy* (Vol. 3, pp. 275–310). OUP.
- Williamson, T. (2007). *The philosophy of philosophy*. Blackwell.
- Williamson, T. (2011). Philosophical expertise and the burden of proof. *Metaphilosophy*, *42*, 215–229.
- Williamson, T. (2016). Philosophical criticisms of experimental philosophy. In J. Sytsma & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 22–36). Wiley.
- Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, *23*, 1028–1034.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.