Stereotype-Based Intuitions: A Psycholinguistic Approach to Experimental Philosophy’s ‘Sources Project’

Eugen Fischer (E.Fischer@uea.ac.uk)
School of Politics, Philosophy, Language and Communication Studies, University of East Anglia
Norwich, NR4 7TJ, UK

Paul E. Engelhardt (p.engelhardt@uea.ac.uk)
School of Psychology, University of East Anglia
Norwich Research Park, Norwich, NR4 7TJ, UK

Abstract

Experimental philosophy’s ‘sources project’ seeks to develop psychological explanations of philosophically relevant intuitions which help us assess their evidentiary value. This paper develops a psycholinguistic explanation of intuitions prompted by brief philosophical case-descriptions. For proof of concept, we target intuitions underlying a classic paradox about perception (‘argument from hallucination’). We trace them to stereotype-driven inferences automatically executed in verb comprehension. We employ a forced-choice plausibility-ranking task to show that contextually inappropriate stereotypical inferences are made from less salient uses of the verb “to see”. This yields a debunking explanation which resolves the philosophical paradox.

Keywords: Experimental philosophy; Sources Project; stereotype-driven inference; graded salience.

Introduction

Philosophical Research Context: This paper presents work that pioneers the use of psycholinguistic methods in experimental philosophy’s ‘Sources Project’. Experimental philosophy is a currently much-discussed and potentially transformative movement that imports methods from psychology into philosophy (for reviews see, Alexander, 2012; Knobe & Nichols, 2014). Philosophers frequently conduct thought experiments which revolve around the consideration of hypothetical cases. Intuitive judgments about such cases are elicited by verbal case-descriptions and play key roles in philosophy: A ‘standard justificatory procedure’ uses such intuitions as evidence for/against philosophical theories (Bealer, 1996; Cath, in press); clashes of intuitions with each other or background beliefs generate influential philosophical paradoxes and problems (Fischer, 2011; Papineau, 2009). Experimental philosophers employ surveys and experimental methods from psychology to elicit, explain, or assess such intuitions.

The strand of the movement that currently attracts most philosophical attention is the ‘Warrant Project’ (for a review see, Stich & Tobia, 2015). This project seeks to assess the evidentiary value of philosophically relevant intuitions and philosophers’ warrant for accepting them. Up to now, the project mainly employs questionnaire-based surveys to study the sensitivity of intuitions to demographic parameters like gender, age, personality type, or cultural or socio-economic background (Weinberg et al., 2001), as well as questionnaire-based experiments to investigate order and framing effects (e.g., Petrinovich & O’Neill, 1996, Weinberg et al., 2012). The standard approach has experimental philosophers make inferences from such sensitivity or effects to lack of evidentiary value. Many of these inferences are philosophically problematic (for a review see, Fischer & Collins, 2015), and various studies have failed to replicate previous results about sensitivity to demographic factors (e.g., Nagel et al., 2013; Seyedsayamdost, 2015a,b). It is therefore timely to pursue the philosophically crucial aims of the Warrant Project with fresh approaches.

The most ambitious strand of the Warrant Project, known as ‘Sources Project’ (Pust, 2012) or ‘cognitive epistemology’ (Fischer et al., 2015), explores such approaches: It seeks to develop and experimentally test psychological explanations of intuitions that help us assess their evidentiary value. One promising approach is to look for explanations that trace intuitions back to largely automatic cognitive processes that are generally reliable but predictably engender cognitive illusions, under specific circumstances. The most prominent line of such research traces intuitive knowledge attributions back to a ‘mind-reading’ capacity subject to specific biases (Alexander et al., 2015; Gerken & Beebee, 2016; Turri, 2015). Our research breaks new ground by examining whether philosophically relevant intuitions are generated or influenced by routine language processes, viz., by automatic inference processes that routinely occur in text comprehension and production. This approach seems worthwhile for the many philosophical relevant intuitions that are elicited by verbal descriptions of hypothetical cases. Elsewhere (Fischer et al., 2015; Fischer & Engelhardt 2016), we have argued that the routine process of stereotype-driven amplification (Levinson, 2000; cf. Garrett & Harnish, 2007) is generally reliable but engenders cognitive illusions, under specific vitiating circumstances. This paper explores the role of this process in generating intuitions at the root of influential philosophical paradoxes about perception, known as ‘arguments from hallucination’. We show that vitiating circumstances obtain in their formulation. This finding debunks the intuitions and helps resolve the paradoxes.
Philosophical Application: Together with ‘arguments from illusion’, arguments from hallucination engender the classic ‘problem of perception’ (Smith, 2002), which has again become a focus of debate (Brewer, 2011; Crane, 2015; Fish, 2009; Robinson, 2001). Both arguments lead to the conclusion that when people use their five senses, they are (directly) aware only of subjective perceptions or sense-data. The philosophical problem is that of reconciling this conclusion with the common-sense conviction that we see and otherwise perceive physical objects and public events.

Philosophers typically formulate arguments from hallucination without world-knowledge about the phenomenon, assume merely that it is possible that ‘one perceives’ a physical object which is not there at all’ (Ayer, 1956/1990, pg. 90), and proceed from brief descriptions of hypothetical cases. Analytic philosophers formulating the argument typically distinguish between different senses of perception-verbs like see: The relevant case-descriptions, they explain, use these verbs in a purely ‘phenomenal’ sense which serves merely to describe people’s experiences and which lacks the existential and spatial implications see, etc. ordinarily have. Here is a classic statement:

‘Let us take as an example Macbeth’s visionary dagger: since we are concerned only with what is possible, the fact that this episode may be fictitious does not matter. There is an obvious [ordinary] sense in which Macbeth did not see the dagger; he did not see the dagger for the sufficient reason that there was no dagger there for him to see. There is another [viz., phenomenal] sense, however, in which it may quite properly be said that he did see a dagger; to say that he saw a dagger is quite a natural way of describing his experience. But still not a real dagger; not a physical object; not even the look of a physical object, if looks are open to all to see. If we are to say that he saw anything, it must have been something that was accessible to him alone, something that existed only so long as this experience lasted; in short a sense-datum.’ (Ayer, 1956, pg. 90)

The second half of the argument then postulates that in this possible scenario the subject’s experience is qualitatively indistinguishable from the experience of seeing a physical object; it assumes that qualitatively indistinguishable experiences involve awareness of the same kind of object; and it concludes that all cases of perception involve awareness of sense-data. This second half has been widely criticized (for a review see, Smith, 2002).

Already the argument’s first half, however, involves a curious mistake: The verb see is explicitly used here in a phenomenal sense: He saw a dagger is to mean ‘He had an experience as of / like that of / seeing a dagger’. An experience is being described by comparing it to that of seeing a certain physical object. This does not require that the object ‘seen’ be around, when the subject has the experience thus described. So we can say that Macbeth saw a physical dagger if his experience ‘is like that of seeing a solid, physical dagger’ (rather than, say, like that of seeing a strangely translucent dagger-image). The case-description explicitly makes this postulate (Ayer 1956, pg. 90), and the second half of the argument crucially depends upon it (above). Of course, only Macbeth ‘sees’ the dagger at this point, and he ‘sees’ it only in the phenomenal sense. But what he ‘sees’ in this sense is still ‘a real dagger’—e.g., the very dagger now sticking in the king’s corpse next door.

What is ‘special’ is the sense in which the argument uses the verb see, but not the object ‘seen’. So why did generations of competent philosophers (e.g., Price 1932, pg. 28-9; Ayer 1956, pg. 90; Smith 2002, pg. 194-5), including philosophers who stress that the use of see and its cognates is special (viz., phenomenal), conclude that the object seen is special (namely, non-physical)?

One key factor, we submit, is a spontaneous inference from ‘Macbeth sees a dagger’ to ‘there is something for Macbeth to see in his vicinity, before his eyes’. This Intuition (I) has thinkers take for granted that the ‘something’ must be around for Macbeth to see, disregard that Macbeth is (in the phenomenal sense) aware of a real dagger (which currently is not in his physical vicinity), and think that the only way of avoiding the ‘preposterous … claim that a hallucinating person is aware of nothing whatever’ (Smith, 2002, 195) is to ‘recognize a non-normal object of awareness’ (ibid.), which is in the subject’s range of vision, before his eyes: Thinking in terms of an intuitive dichotomy between external and internal perception (introspection), proponents of the argument immediately conclude that the thing seen must be before the subject’s inner eye, in his mind. I.e.: We submit the argument rests on intuition (I), and (I), in turn, on inappropriate spatial inferences from the phenomenal use of see.

Psycholinguistic explanation: Both nouns (Hare et al., 2009) and verbs (Ferretti et al., 2001; Harmon-Vukic et al., 2009) are associated with stereotypes. Verbs can be associated with typical features of events, agents, and patients, which jointly form structured stereotypes (a.k.a. ‘generalized situation schemas’). E.g., manipulators are typically cunning and shrewd, and their victims naïve and gullible, so the verb manipulate is associated with ‘cunning’ and ‘shrewdness’ as agent-properties, and ‘naïveté’ and ‘gullibility’ as patient-properties. Such associations guide spontaneous inferences from those expressions. Within a neo-Gricean framework, their interplay with integration processes (see below) is captured by the I-heuristic (Levinson, 2000; cf. Garrett & Harnish, 2007): In the absence of explicit indications to the contrary, infer that situations talked about conform to stereotypes associated with the nouns and verbs used.

Stereotypical inferences are supported by automatic activation processes in semantic memory (McRae & Jones, 2013). According to the well-supported graded salience hypothesis (Fein et al., 2015; Giora, 2003), a linguistic stimulus activates all semantic and stereotypical features
associated with the expression, in any of its uses or senses. The speed and strength of initial activation depends upon the ‘salience’ of the sense or use. This is a function of familiarity, conventionality, frequency, and prototypicality. Features associated with the expression’s most salient use are activated regardless of context. E.g., the ambiguous stimulus mint activates the probe ‘candy’ rapidly and strongly, even where it is clearly used in a less frequent sense (prime: All buildings collapsed except the mint) (Simpson & Burgess, 1985; Till et al., 1988). This facilitates contextually inappropriate inferences from less salient uses of words. Such inferences need not go through, since processes including reinforcement and decay (Oden & Spira, 1983), and suppression (Faust & Gernsbacher 1996) mitigate initial preferential activation of contextually inappropriate stereotypical associates, namely, in the light of contextual cues, explicit indications of deviation from relevant stereotypes, and explicit marking of less salient uses (Givoni et al., 2013). Where one use is very much more salient than another, and the stereotypical association with certain features particularly strong, inappropriate stereotypical inferences may, however, go through, even in the face of explicit indications to the contrary (Giora 2003).

We hypothesize that this happens in the argument from hallucination. The verb see is used in at least three senses: an ordinary, literal, visual sense (Bob saw the builders fixing the road), a metaphorical, epistemic sense (Jane saw Joe’s point), and an extended phenomenal sense (Hitting his head, Jack saw stars). According to our first hypothesis,

$$H_1$$ The literal, visual sense of see is much more salient than the metaphorical, epistemic sense; this, in turn, is much more salient than the phenomenal sense.

This non-salient sense is employed in the case-descriptions that serve as first premises of the argument from hallucination. We hypothesize further that proponents of the argument then make contextually inappropriate stereotype-driven inferences from those premises: We assume that the most salient visual sense of see has a strong stereotypical association with spatial patient-properties: S sees X is strongly associated with ‘X is in front of S’ and ‘X is before S’s eyes’ (directional implications) as well as with ‘X is around (S) to be seen’ and ‘X is within S’s range of vision’ (proximity implications). On this basis we hypothesize that

$$H_2$$ Competent speakers infer spatial patient-properties stereotypically associated with the visual sense of see also from less salient epistemic and phenomenal uses.

Such a stereotypical inference, we submit, leads proponents of the argument from hallucination from such initial premises as ‘When hallucinating, Macbeth sees a dagger’ to the intuitive judgment (I) that there is a dagger around for Macbeth to see, before his eyes (see above). $H_1$ and $H_2$ could jointly explain the intuition (I) we identified as the intuitive source of the argument from hallucination.

In following up $H_1$, we bear in mind that salience is a function of familiarity, conventionality, frequency, and prototypicality. We think it unlikely that the visual and epistemic uses of see will attract different familiarity- or conventionality-ratings, and suggest salience differences are due to differences in frequency and prototypicality. A corpus study to establish frequency is still ongoing. First evidence for prototypicality differences is provided by a production study. 13 participants were asked to provide up to 10 written completions of 4 sentence stems containing either see or aware (e.g. Jane sees____; Bob saw____). To discourage mono-topical response strategies, participants were instructed to ‘try to give varied responses’. We coded responses as either perceptual (e.g. the handsome American) or non-perceptual (e.g. fresh opportunities). Results showed that completions of see were 94% perceptual use (by contrast, 51% for aware). This suggests that visual cases of ‘seeing’ are by far the most prototypical of this category, and that the literal, visual sense of see is by far the most salient. Non-perceptual completions involved mainly epistemic uses. Phenomenal uses were absent. This is consistent with $H_1$. We then used a forced-choice plausibility-ranking task to examine $H_2$ and, specifically, explore spatial (directional and proximal) inferences from metaphorical/epistemic uses of see.

**Methods**

**Participants**

Fifty undergraduate psychology students from the University of East Anglia were recruited through the participant pool and received course credit for participating.

**Materials**

We administered a paper-and-pencil questionnaire that had 120 minimal pairs. There were 48 critical items and 72 fillers. Twenty-four critical items contrasted see and aware. (Aware can be used in both perceptual and non-perceptual contexts and lacks directional implications in either.) The other 24 critical items contrasted see and think of. (Think of lacks proximity implications in all uses). The see-aware items consisted of single sentences, the see-think of items consisted of pairs of sentences (see Tables 1 and 2).

<table>
<thead>
<tr>
<th>Table 1: Example items for testing spatial inferences and specifically, directional inferences (e.g. in front of).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual, Stereotype-Consistent</td>
</tr>
<tr>
<td>1a. Mona sees the drivers ahead of her in the queue.</td>
</tr>
<tr>
<td>1b. Mona is aware of the drivers ahead of her in the queue.</td>
</tr>
<tr>
<td>Visual, Stereotype-Inconsistent</td>
</tr>
<tr>
<td>2a. Ben sees the friend walking right behind him.</td>
</tr>
<tr>
<td>2b. Ben is aware of the friend walking right behind him.</td>
</tr>
<tr>
<td>Epistemic, Stereotype-Consistent</td>
</tr>
<tr>
<td>3a. Emma sees the challenges facing her.</td>
</tr>
<tr>
<td>3b. Emma is aware of the challenges facing her.</td>
</tr>
</tbody>
</table>

528
Epistemic, Stereotype-Inconsistent
4a. Jim sees the scheming going on behind his back.
4b. Jim is aware of the scheming going on behind his back.

Table 2: Example items for testing proximity inferences (i.e. around to be see, within the range of vision).

Visual, Stereotype-Consistent
1. Megan spent the week at home. She saw/thought of her parents a lot.
Visual, Stereotype-Inconsistent
2. Sitting in the office, John sees/thinks of Peter. Peter is on holiday in Turkey.
Epistemic, Stereotype-Consistent
3. Joan sees/thinks of fresh opportunities. Some opportunities are around.
Epistemic, Stereotype-Inconsistent
4. Joe sees/thinks of ways to solve the problem. There are no solutions around.

There were two manipulations. The first concerns the use of the verb see (literal/visual vs. metaphorical/epistemic). More specifically, it turns on whether the direct object of the main verb is a visible physical object (e.g. drivers, friends, etc., in the literal/visual condition) or an abstract object (e.g. challenges, scheming, etc., in the metaphorical/epistemic condition). The second variable manipulated was whether the remainder of the sentence was either consistent or inconsistent with the hypothesized stereotypical inferences, viz. directional and proximity inferences: In the stereotype-consistent condition, the patient of see and aware is said or clearly implied to be in front of the agent or before the agent’s eyes; in the stereotype-inconsistent condition, the patient is placed behind the agent. Similarly, the patient of see and think of was placed in the vicinity or visual field of the agent, in the stereotype-consistent condition, and outside this field, in the stereotype-inconsistent condition. The critical verb in the see/think of items could occur in either first or second sentence (see Table 2).

Design and Procedure
The design of the study was 2 × 2 (visual/epistemic × s-consistent/s-inconsistent). The visual/epistemic variable refers to the direct object (visual vs. abstract) of the critical verb. S-consistent/s-inconsistent refers to the ‘consistency/inconsistency with spatial (directional and proximity) implications from literal/visual use of see’ and was contained in further contextual information in the sentence. Separate repeated-measures ANOVAs (with follow-up t-tests) were run on see/aware and see/think of separately. In line with our hypothesis H2, we expected see to be preferred with visual direct objects in s-consistent contexts, and aware and think of to be preferred with both visual and abstract direct objects in s-inconsistent contexts. The critical condition occurred when the direct object was abstract and invoked epistemic sense and the context was s-inconsistent with directional and proximity inferences. For graphical purposes, we coded see responses as 1 and aware and think of responses as 0. Thus, high bars indicate greater preference for see.

Participants were presented with pairs of sentences or short two-sentence pairs, which differ only in one critical word (i.e. minimal pairs). Participants were asked to indicate which of the two versions they thought was more plausible, and to make a judgment even if they did not have a clear preference. By noting how often participants preferred one version over the other, we can measure the consistency of plausibility judgments, and which of the two verbs have stronger stereotypical associations with the targeted spatial properties.

Results
See vs. Aware
Results showed that both main effects were significant (visual/epistemic F(1,48) = 171.75, p < .001 and s-consistent/s-inconsistent F(1,48) = 662.78, p < .001). The interaction was also significant F(1,48) = 226.76, p < .001 (see Figure 1). With visual direct objects, see was preferred over aware 96% of the time, when contexts were s-consistent, but only 5% of the time when contexts were s-inconsistent. With epistemic direct objects, see was preferred 44% of the time when contexts were s-consistent but only 8% of the time when contexts were s-inconsistent.

The comparison of s-consistent and s-inconsistent was significant for both visual (t(48) = 45.35, p < .001) and epistemic (t(48) = 9.26, p < .001) senses. The preference for aware in the epistemic/s-consistent condition was not significantly different from chance (t(48) = 1.90, p = .06).

Figure 1: Results showing proportion of “see” responses with visual and epistemic objects.

See vs. Think of
Results showed that both main effects were significant (visual/epistemic F(1,48) = 92.93, p < .001 and s-consistent/s-inconsistent F(1,48) = 817.33, p < .001). The interaction was also significant F(1,48) = 93.53, p < .001.
(see Figure 2). With visual objects, see was preferred over think of 95% of the time when contexts were s-consistent, but only 2% of the time when contexts were s-inconsistent. With epistemic objects, see was preferred 58% of the time when contexts were s-consistent, but only 5% of the time when contexts were s-inconsistent. The comparison of s-consistent vs s-inconsistent was significant for both visual t(48) = 43.83, p < .001 and epistemic t(48) = 12.75, p < .001 senses. The preference for see in the epistemic/s-inconsistent condition was significantly above chance t(48) = 1.99, p = .053. Thus, there was a marginal see preference.

![Figure 2: Results showing proportion of “see” responses with visual and epistemic objects.](image)

**Discussion**

These results are consistent with our key hypothesis H2 that competent speakers infer spatial patient-properties stereotypically associated with the visual sense of see also from epistemic and phenomenal uses. H2 implies that in epistemic contexts which are s-inconsistent, competent speakers will reject see in favor of verbs that work equally well in epistemic contexts but lack spatial implications. This is precisely what we found: As long as items were s-consistent, participants had no pronounced preference for aware or think of over the evidently metaphorical use of see – all three verbs were deemed to work roughly equally well in these contexts. By contrast, participants had a very pronounced preference for aware and think of in epistemic contexts which were inconsistent with the (evidently irrelevant) spatial implications of see. This suggests that participants made spatial inferences from the verb even in these inappropriate contexts, resulting in perceived inconsistencies with the context, which in turn lower subjective plausibility.

In conjunction with previous evidence that (H1) visual uses of see have higher salience than epistemic and phenomenal senses, this is also consistent with the graded salience hypothesis. This hypothesis could explain our key finding (see Introduction) and let us extend it: Speakers/hearers are the more likely to infer features stereotypically associated with the most salient use of a word also from its less salient uses, the greater the difference in salience is (Simpson & Burgess, 1985). Since the phenomenal use of see is even less salient than the epistemic use (see Introduction), it stands to reason that competent speakers will be even more prone to make inappropriate spatial inferences from phenomenal than from epistemic uses.

This conclusion can help resolve the targeted philosophical paradox (argument from hallucination), by debunking the intuition at its root, viz., that when Macbeth sees the dagger, there is something around for him to see, before his eyes. Our conclusion suggests that stereotype-driven inferences from the most salient visual use of see have philosophers leap to this intuitive judgment from prior case-descriptions which use see in a phenomenal sense. The intuition is thus due to a contextually inappropriate stereotypical inference. More generally, we have identified a vitiating circumstance: The generally reliable process of stereotype-driven amplification (Levinson, 2000) misfires where a word with strong stereotypical associates of a highly salient use is employed in a much less salient sense.

**Future directions:** We intend to garner further evidence for H2 through a computer-based rating experiment asking participants to rate ‘how good examples of seeing are afforded by the following cases’, whose descriptions include the three uses distinguished. To collect further evidence for H2 we intend to replicate the current results with a less artificial task, using pupillometry (Engelhardt et al., 2010). A more comprehensive explanation of the targeted intuitions will take further linguistic theories into account, such as head-driven phrase structure grammar (Pollard & Sag, 1994): When accessing a verb, we retrieve all associated syntactic information, including the grammatical roles (e.g., see requires a patient-role) and selection-restrictions on role-fillers (e.g., physical objects and public events). This may account for the intuition that Macbeth sees something. This paper explains how this ‘something’ gets endowed with spatial features – which exclude any absent physical objects as patients of see. The situation from which the present intuition arises is not rare in philosophy: Philosophers often give familiar words special uses which are less salient than those in ordinary language. It bears investigating to what extent these uses prompt contextually inappropriate inferences and unwarranted intuitions.

**Conclusion**

Psycholinguistic methods and findings can be profitably used to identify and explain automatic inferences that generate philosophically relevant intuitions. For proof of concept, this paper showed how psycholinguistic data can contribute to debunking explanations of intuitions that engender philosophical paradoxes.

**References**
