



Experimental ordinary language philosophy: a cross-linguistic study of defeasible default inferences

Eugen Fischer¹ · Paul E. Engelhardt² · Joachim Horvath³ · Hiroshi Ohtani⁴

Received: 29 May 2018 / Accepted: 2 January 2019
© The Author(s) 2019

Abstract

This paper provides new tools for philosophical argument analysis and fresh empirical foundations for ‘critical’ ordinary language philosophy. Language comprehension routinely involves stereotypical inferences with contextual defeaters. J.L. Austin’s *Sense and Sensibilia* first mooted the idea that contextually inappropriate stereotypical inferences from verbal case-descriptions drive some philosophical paradoxes; these engender philosophical problems that can be resolved by exposing the underlying fallacies. We build on psycholinguistic research on salience effects to explain when and why even perfectly competent speakers cannot help making stereotypical inferences which are contextually inappropriate. We analyse a classical paradox about perception (‘argument from illusion’), suggest it relies on contextually inappropriate stereotypical inferences from appearance-verbs, and show that the conditions we identified as leading to contextually inappropriate stereotypical inferences are met in formulations of the paradox. Three experiments use a forced-choice plausibility-ranking task to document the predicted inappropriate inferences, in English, German, and Japanese. The cross-linguistic study allows us to assess the wider relevance of the proposed analysis. Our findings open up new perspectives for ‘evidential’ experimental philosophy.

Keywords Experimental philosophy · Ordinary language philosophy · Salience effects · Crosscultural psycholinguistics · Naturalised argument analysis · Argument from illusion

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11229-019-02081-4>) contains supplementary material, which is available to authorized users.

✉ Eugen Fischer
e.fischer@uea.ac.uk

¹ School of Politics, Philosophy, Language and Communication Studies, University of East Anglia, Norwich NR4 7TJ, UK

² School of Psychology, University of East Anglia, Norwich NR4 7TJ, UK

³ Institute for Philosophy II, Ruhr University Bochum, 44780 Bochum, Germany

⁴ Faculty of Human Sciences, Musashino University, Tokyo 135-8181, Japan

1 Introduction

Some words bewitch. So do some senses. Among these are dominant senses of polysemous words, namely when they are (in ways to be explained) ‘functional’ for the interpretation of more rarefied uses: Even in competent speakers, such less salient uses then predictably prompt stereotypical inferences licensed by the dominant sense or use but not by the less salient sense or use demanded by the context. Such contextually inappropriate stereotypical inferences are at the root of some philosophical paradoxes and problems. These two hypotheses, one psycholinguistic, the other metaphilosophical, promise to jointly provide fresh foundations for the best known critical project in ordinary language philosophy, initiated by J.L. Austin (1962), which addresses philosophical problems by disentangling contextually inappropriate default inferences. We will examine the two hypotheses through experiments on inferences from appearance verbs and by following up the suggestion that the contextually inappropriate inferences we document are at the root of a classical paradox about perception, known as ‘argument from illusion’. This philosophically motivated exercise in experimental pragmatics will forge fresh connections between experimental philosophy and ordinary language philosophy.

The rise of experimental philosophy has contributed to renewed interest in ordinary language philosophy (OLP) (Fischer 2011; Baz 2012, 2016; Garvey 2014; Gustafsson and Sørli 2011; Hansen 2014, 2018; Laugier 2013). OLP was analytic philosophy’s first attempt to overcome limitations of armchair reflection through the use of (informal) experiments (Hansen and Chemla 2015), (peer-based) focus groups (Urmson 1969), and empirical surveys (Murphy 2014).¹ Forging fresh connections between experimental philosophy and its historical precursor, this paper will draw on inspiration from OLP to recruit psycholinguistic methods and findings for further development of one main strand of experimental philosophy’s ‘evidential’ research programme (Sytsma and Livengood 2016, pp. 40–42). We will explore how ideas pioneered by OLP’s ‘critical’ Austinian strand, exemplified by *Sense and Sensibilia* (Austin 1962),² can help develop the prominently discussed ‘restrictionist’ strand of evidential experimental philosophy (reviews: Mallon 2016; Stich and Tobia 2016), into an ‘experimental philosophy 2.0’ (Nado 2016).

Restrictionism seeks to debunk intuitions adduced as evidence for philosophical claims and theories (reviews: Alexander 2012, pp. 70–89; Horvath 2010) and—more recently—to ‘dissolve’ certain philosophical problems (Weinberg 2017, p. 179). A first generation of contributions sought to assess the evidentiary value of philosophically relevant intuitions. These studies examined the sensitivity of such intuitions to presumably irrelevant parameters, and inferred lack of evidentiary value from observed sensitivity to demographic parameters, order and framing effects. This research met

¹ ‘OLP’ is a family resemblance concept (*cf.* Baz 2016, p. 112) associated with several related paradigms. These include two precursors of experimental philosophy: J.L. Austin (e.g., 1957, 1962) and Arne Naess (e.g., 1956, 1961). This paper develops Austinian ideas. For Naess, and methodological debates between the two camps, see Murphy (2014) and *cf.* Hansen (2017).

² OLP’s other, ‘constructive’ main strand, exemplified by Austin (1957), is continued, e.g., by experimental work on epistemic contextualism (Gerken and Beebe 2016; Grindrod et al. 2018; Hansen and Chemla 2013; Schaffer and Knobe 2012).

with empirical and theoretical challenges: Studies on demographic parameters (age, gender, etc.) encountered replication difficulties (review: Machery 2017, ch. 2) exceeding those of experimental philosophy as a whole (Cova 2018); the inference from apparent order effects to lack of evidentiary value was forcefully questioned (Horne and Livengood 2017); and critics maintained that philosophers do not rely on intuitions as evidence in the way restrictionism presupposes (Cappelen 2012; Deutsch 2015; Williamson 2007). Partially in response to these difficulties, recent calls for an ‘experimental philosophy 2.0’ (Nado 2016) suggest that evidential experimental philosophers should, instead, (1) examine cognitive processes that underpin philosophical thought (paradigm: Nichols and Knobe 2007), (2) seek to develop epistemological profiles of such processes, which indicate under which conditions we may (not) trust their outputs (Weinberg 2015, 2016), and (3) assess a wider range of outputs: not only intuitive judgments but also inferences in arguments (Fischer and Engelhardt 2017).

At this point in the development of evidential experimental philosophy, OLP’s critical Austinian strand provides fresh inspiration: Austin (1962) considered defeasible default inferences which are shaped by ordinary uses of words, and sought to ‘dissolve’ philosophical problems by disentangling such inferences. Defeasible default inferences continuously occur in language comprehension and production—e.g., whenever thinkers read or state verbal case-descriptions or premises of arguments. This paper conceptualises relevant inferences, in a neo-Gricean framework, as part-and-parcel of ‘stereotypical enrichment’ (Levinson 2000). In the spirit of experimental philosophy 2.0, we (1) experimentally examine this key cognitive process which can shape thought in *any* area of philosophy, in order to (2) contribute towards an epistemological profile of stereotypical enrichment and, on this basis, (3) assess how stereotypical inferences influence philosophical argument, for better or worse. In the spirit of critical OLP, we explore to what extent our findings can contribute to resolving a philosophical paradox and ‘dissolving’ a problem it engenders. For this purpose, we chose the classical paradox about perception (‘argument from illusion’) targeted by Austin’s *Sense and Sensibilia* (1962). Together with related paradoxes, it engenders the ‘problem of perception’, a renewed focus of debate (Crane and French 2015; Fish 2009; Robinson 2001; Smith 2002). We thus seek to trial the application of psycholinguistic methods and findings in philosophical argument analysis and to provide proof of concept for experimental implementation of Austin’s critical project: an ‘experimental ordinary language philosophy (2.0)’.

We now empirically substantiate Austinian ideas about defeasible default inferences from words by reviewing psycholinguistic findings about stereotypical inferences and their contextual integration (Sect. 2). We then draw on psycholinguistic research on salience effects (Fein et al. 2015; Giora 2003) to contribute to an epistemological profile of stereotypical enrichment and identify vitiating conditions under which the process leads to inappropriate inferences (Sect. 3). Three experiments on inferences from appearance verbs evaluate our psycholinguistic hypothesis that, under the conditions identified, even competent language users make and accept contextually inappropriate stereotypical inferences. To show that our findings are not merely reflective of idiosyncrasies in English and to assess their philosophical relevance, we follow up an English study (Sect. 4) with replications in languages with different sentence positions for verbs, viz., German (Sect. 5) and Japanese (Sect. 6). Finally, we deploy

these empirical findings to develop the metaphilosophical hypothesis that contextually inappropriate stereotypical inferences from appearance verbs are at the root of the ‘argument from illusion’, and explore how psycholinguistic findings can contribute to ‘dissolving’ the ‘problem of perception’ (Sect. 7). We thus aim to provide proof of concept for a more widely applicable approach that addresses certain kinds of problems by developing and empirically investigating psycholinguistic (and metaphilosophical) hypotheses which are also interesting in their own right.

2 Stereotypical enrichment

Much of *Sense and Sensibilia* discusses default inferences from words which have subtle contextual defeaters. Without the benefit of the conceptual apparatus available today, Austin sought to clarify ‘the root ideas behind the uses of’, e.g., appearance verbs ‘look’, ‘appear’, and ‘seem’ (Austin 1962, p. 37) which are employed in the initial premises of the paradox he targeted in that work. He considered example sentences and ‘in just what circumstances we would say which, and why’ (p. 36), e.g., (*cf.* Price 1932, p. 28):

1. ‘The hill looks steep’—[it has the look of a steep hill];
2. ‘The hill appears steep’—when you look at it from down here;
3. ‘The hill seems steep’—to judge by the fact that we had to change gear twice.

Such examples suggest that while ‘looks’ is used simply to comment on the look of things, ‘appears’ ‘would typically be used with reference to certain *special circumstances*’ affecting judgment, and ‘seems’ ‘makes an implicit reference to certain [inconclusive] *evidence*’ supporting judgment (Austin 1962, pp. 36–37). If this is correct, hearers will tend to infer *S is inclined to judge/think that X is F* from ‘X appears F to S’ and ‘X seems F to S’, but not from ‘X looks F to S’ (Fischer 2014a).

Simultaneously, Austin stresses the radical context-sensitivity of inferences made or anticipated in language-comprehension and -production:

‘If I say that petrol looks like water, I am simply commenting on the way petrol looks; I am under no temptation to think, nor do I imply, that perhaps petrol *is* water. [...] But ‘This looks like water’ ... may be a different matter; if I don’t already know what ‘this’ is, I *may* be taking the fact that it looks like water as a ground for thinking it *is* water’ (Austin 1962, pp. 40–41).

Therefore, ‘just what is meant and what can be inferred (if anything) can be decided only by examining the full circumstances in which the words are used’ (p. 41): Under some circumstances, ‘X looks F (to S)’ may prompt and warrant a doxastic inference to *S is inclined to think that X is F*, which otherwise is the preserve of ‘seem’ and ‘appear’.

Psycholinguistic research has since substantiated both general suggestions: that the use of particular words is associated with ‘root ideas’ which facilitate defeasible default inferences; and that inferences—appropriately—made from uses of words depend upon sentence- and, indeed, utterance-context. In today’s terms, Austin’s ‘root

ideas' are stereotypes; and the 'root ideas' associated with the verbs of interest are generalised situation schemas.

As traditionally conceived, *stereotypes* are sets of features (categories, properties, relations, etc.) which come to mind first, and are easiest to process, when we hear a noun, verb, or idiomatic expression. In the simplest examples, relevant features can be elicited through listing or sentence-completion tasks ('Tomatoes are__') (e.g., McRae et al. 1997). Single words activate features rapidly (within 250 ms), as shown by priming experiments (Balota and Lorch 1986; Ferretti et al. 2001; Hare et al. 2009; Lupker 1984; Welke et al. 2015).³ According to standard accounts of semantic memory (McRae and Jones 2013; Neely and Kahan 2001), stereotypical associations between things and their features (properties, relations, etc.), parts and wholes, causes and effects are built up through observed co-occurrences in the physical environment and linguistic representations. Strength of stereotypical association thus encodes information about the world.

Stereotypical associations do not determine the extension of words (Hampton and Passanisi 2016), but support automatic default inferences from words to features stereotypically associated with them. Such *stereotypical inferences* have been studied through reading times (Garrod and Sanford 1981; McKoon and Ratcliff 1980; O'Brien and Albrecht 1992), eye movements (Patson and Warren 2010; Rayner 1998), and event-related brain potentials (ERPs) (Kutas and Federmeier 2000, 2011). In these studies, participants read sentences where the expression of interest is followed by a sequel inconsistent with a hypothesised inference. Conflicts lead readers to slow down and make more backwards eye-movements; they also prompt signature electrophysiological responses (known as 'N400s'). For example, when reading 'sewing', people rapidly infer the agent used a needle—and slow down when the text continues '...the job would be easier if Carol had a needle' (Harmon-Vukic et al. 2009).

Event nouns (Hare et al. 2009) and verbs (e.g., Ferretti et al. 2001) can be associated with complex stereotypes: Where the actions or events denoted typically involve particular (kinds of) agents, patients acted on, instruments, or relations between them (Tanenhaus et al. 1989), associated stereotypes include typical features of these role-fillers. For example, 'frighten' immediately brings to mind agent-properties *mean*, *ugly*, and *big*, as well as patient properties including *small* and *weak* (McRae et al. 1997). In incremental language comprehension, these complex stereotypes are deployed in a structured manner: Sentence fragments ('She was arrested by the __') activate typical agents (*cop*) in post-verbal position only when they leave the agent role blank (as above), not when they leave open the patient ('She arrested the __') (Ferretti et al. 2001).⁴ These complex, structured stereotypes are known as *generalised*

³ Participants are presented with a 'prime' word or short text and then a 'probe' word or letter string, and have to, e.g., read out the word or decide whether the string forms a word. That the prime *activates* the probe concept, i.e., makes it more accessible and likely to be used by cognitive processes (from word recognition to forward-inferencing), is inferred from shorter response times (Lucas 2000).

⁴ Similarly, while ERP studies show the verb is expected (as indicated by reduced N400 amplitude), where preceded by subject and object ('The restaurant owner forgot which *customer* the *waitress* had served...'), even where their typical roles are reversed ('... which *waitress* the *customer* had served') (Chow et al. 2016), this reversal prompts signature electrophysiological responses to syntactic violations (known as 'enhanced P600'). This suggests that participants expected the verb in the passive voice ('which waitress the customer

situation schemas (Rumelhart 1978; Tanenhaus et al. 1989). If Austin is right, the situation schemas associated with ‘X appears F to S’ and ‘X seems F to S’ include the patient-feature *S is inclined to judge that X is F*.

The radical context-sensitivity of comprehension inferences that Austin noted partially arises from the fact that inferences made from the verb can take into account previous agent and instrument nouns. Self-paced reading-time studies found that participants read the remainder of the sentence more slowly when subject and verb were followed by a patient atypical for that particular agent-action pairing (‘The *mechanic/journalist* checked the spelling of his latest report’) (Bicknell et al. 2010). A similar finding was made for instruments (‘Susan used the *saw/scissors* to cut the expensive paper...’), despite the absence of single-word priming of typical patients (e.g., ‘scissors’-*paper*) (Matsuki et al. 2011). These findings suggest that reading activates not only knowledge about the typical features of, say, journalists and mechanics, or of checking events, but also more specific knowledge about what mechanics check that does not get activated by single words. ERP and eye-tracking studies suggest that inferences supported by activation of such specific knowledge were made at the earliest possible moment, i.e., right after the verb (Bicknell et al. 2010; Kamide et al. 2003). In incremental comprehension, hearers/readers immediately employ knowledge encoded in event schemas of varying degrees of complexity and specificity, the moment their relevance becomes apparent. These schemas go beyond schemas associated with specific words and include generalised situation schemas which encode empirical knowledge about events, but are not associated with any one word.

In accordance with the *I-heuristic* (‘What is expressed simply is stereotypically exemplified’, Levinson 2000, p. 37), hearers deploy such schemas and speakers anticipate their use, to devise or facilitate interpretations that are positive, stereotypical, and highly specific (op.cit. pp. 114–115), in the process of *stereotypical enrichment*:

- (I-speaker) Skip mention of stereotypical features but make deviations from stereotypes explicit.
- (I-hearer) In the absence of such explicit indications to the contrary, assume that the situation talked about conforms to the relevant schemas, deploy the most specific schemas relevant, and fill in detail in line with this knowledge about situations of the kind at issue.

The research reviewed supports what we call a ‘*cued schemas account*’ of language comprehension and production: Articulation of speech proceeds at a slower pace than pre-articulation in speech production (Wheeldon and Levelt 1995) or parsing- and inference-processes in comprehension (Mehler et al. 1993). Inferences based on cued schemas (‘stereotypical inferences’, in a wider sense) help mitigate this ‘communication bottleneck’: Words and syntactic constructions (Goldberg 2003; along with verb aspect: Ferretti et al. 2007; Kehler et al. 2008) are used as complementary cues for indicating and accessing relevant empirical knowledge in incremental language comprehension and production (Elman 2009). Relevant knowledge is encoded by

Footnote 4 continued
had been served by’), consistent with assignments of agent and patient-roles typical for the verb (Kim et al. 2016; cf. Kim and Osterhout 2005).

stereotypes, in particular situation schemas. Increasingly specific schemas are activated by words and combinations of verbs and agent- or patient-nouns, as well as discourse context (Metusalem et al. 2012). Activated schemas then support a multitude of rapid, parallel stereotypical inferences. At each point, receivers use the most specific inferences to flesh out utterance content. The activation processes in semantic memory that support these inferences occur in both comprehension and production (Pickering and Garrod 2013; Stephens et al. 2010). Psycholinguistic findings therefore provide empirical support for Austin's (1962, pp. 40–41) suggestion that while words are associated with 'root ideas', how they are 'intended and taken' in ordinary discourse depends upon the utterance-context.

3 Contextually inappropriate inferences

Austin (1962, pp. 4–5) also moots the idea that some philosophical paradoxes and problems turn on contextually inappropriate default inferences, so they can be resolved by exposing such inappropriate inferences. This idea provides the metaphilosophical foundation of Austin's brand of critical ordinary language philosophy. This key idea, however, faces a serious challenge: As we have just seen, how competent speakers intend and interpret words is highly sensitive to utterance context. Philosophers formulating and addressing paradoxes are competent speakers. Austin's empirically confirmed positive point about the context-sensitivity of comprehension inferences thus seems to undercut the rationale of his critical project, and motivates the question: Exactly when (if ever) and why should competent speakers (like philosophers) make, or fall for, contextually inappropriate stereotypical inferences, in formulating their arguments or problems?

This question is rendered yet more pressing by the 'paradox of charity' (Lewinski 2012; cf. Adler 1994): Hermeneutic principles of charity constrain attributions of fallacies to competent thinkers. To warrant such attributions, any 'diagnostic' analysis that seeks to identify fallacies in philosophical paradoxes has to be supported by empirical explanations that let us understand when and why competent thinkers commit those fallacies (Thagard and Nisbett 1983). Austin's general approach as well as 'diagnostic' responses to specific philosophical problems require empirical validation.

To provide empirical foundations for Austin's general approach, and diagnostic argument analysis more generally, we will develop and experimentally test a psycholinguistic explanation that identifies one relevant set of conditions under which even competent speakers make contextually inappropriate stereotypical inferences (Sects. 3–6). Then we will explore to what extent our findings support a specific philosophical application, namely, a diagnostic reconstruction of Austin's chief target, the 'argument from illusion', which identifies inappropriate stereotypical inferences from appearance-verbs as the root of this paradox (Sect. 7). We now draw on psycholinguistic research on salience effects to explain when and why competent speakers go along with contextually inappropriate stereotypical inferences (Sect. 3.1). Then we review empirical evidence concerning appearance verbs (Sect. 3.2). This will allow us to derive from our explanation some word-specific hypotheses that are experimentally testable (Sect. 3.3).

3.1 Psycholinguistic explanation

Most words have more than one meaning or sense (Klein and Murphy 2001). Whenever we hear or read them, all their meanings or senses get initially activated (Fodor 1983; Simpson and Burgess 1985; Till et al. 1988). That is, a linguistic stimulus activates all (sets of) semantic and stereotypical features associated with the expression, in any of its senses. It does so regardless of contextual relevance. For example, the homophonous word ‘mint’ activates *candy* rapidly and strongly, even where used in a less frequent meaning (‘All buildings collapsed except the mint’) (Till et al. 1988).

Purely stimulus-driven activation processes run in parallel with context-sensitive processes (Giora 2003; Levinson 2000; Peleg and Giora 2011) that are driven, *inter alia*, by more specific situation schemas activated in incremental comprehension (Sect. 2). Their outputs are continuously integrated via processes such as reinforcement and decay (Oden and Spira 1983), and more effortful suppression (Faust and Gernsbacher 1996). Thus, initial activation is mitigated in the light of context (‘the secretary scratched his beard’) (Sturt 2003) and explicit indications of deviation from relevant stereotypes (‘male secretary’) (Osterhout et al. 1997), including situation schemas and scripts (Traxler et al. 2000). We now build up towards conditions under which initially activated stereotypical features remain activated in inappropriate contexts, even so, and influence further cognitive processing.

According to the *Graded Salience Hypothesis* (Fein et al. 2015; Giora 2003), initial activation is ordered by ‘salience’ (where this label is applied to a magnitude that is insensitive to immediate discourse context): The (non-contextual) *salience* of a sense or use is a function of exposure frequency (how often a language user encounters the word in that use), modulated by prototypicality.⁵ The more salient a use is for a speaker/hearer, the more rapidly and strongly the situation schema associated with that use is activated. Highly salient uses will strongly activate the associated situation schema, regardless of contextual (im)propriety. The more strongly activated a schema is, the longer its activation takes to decay (Farah and McClelland 1991; Loftus 1973) and the more effortful it is to suppress (De Neys et al. 2003; Levy and Anderson 2002; Giora 1997).

When the word is used in a different, less salient sense, context-sensitive processes may lead to suppression of the initially activated, but contextually inappropriate, dominant schema that is associated with the most salient sense. This happens, specifically, where the less salient sense is associated with an entirely different schema whose activation is enhanced by explicit marking of the less salient sense (Givoni et al. 2013). But the less salient sense need not be associated with an entirely different schema. Rather, according to the *Retention/Suppression Hypothesis* (Giora 2003; Giora et al. 2014), its interpretation (e.g., ‘I see your point’) may involve retaining the dominant schema and suppressing its contextually inappropriate features (*agent S uses her eyes; patient X is located in front of S*, etc.) while deploying the contextually relevant ones (*S knows what X is*) (Fischer and Engelhardt 2019). This ‘retention strategy’ (for short)

⁵ Since exposure frequency cannot be directly measured, it is inferred from occurrence frequencies in corpora or from familiarity and conventionality ratings, leading to the early, but not strictly accurate, explanation of (non-contextual) salience as a function of ‘frequency, familiarity, conventionality, and prototypicality’ (e.g., Giora 2003, pp. 15–22). Giora now concurs with the above explication (personal communication).

has been shown to be used in interpreting irony (Giora et al. 2007b), sarcasm (Fein et al. 2015), and metaphor (Giora et al. 2007a; Giora and Fein 1999).

Under some conditions, suppression of irrelevant features will remain partial: Where a word

1. is frequently used and
2. has a dominant sense that is far more salient than the others,

the semantic and stereotypical features that make up its associated situation schema may go together so often that initial stimulus-driven activation (due to salience) will be complemented by lateral activation between frequently co-occurring features, as elements of a situation schema activate others (Hare et al. 2009; McRae et al. 2005). It will then be difficult to suppress only some, but not all of them, when

3. the retention strategy is used to interpret less salient uses.

Where suppression of contextually irrelevant features remains partial, these irrelevant schema components will support contextually inappropriate stereotypical inferences which are presupposed in further reasoning. We have thus built up to a set of jointly vitiating conditions—(1) to (3)—in which stereotypical enrichment leads to contextually inappropriate inferences, despite the general context-sensitivity of comprehension inferences noted already by Austin (1962, pp. 40–41). These conditions are articulated by the *Salience Bias Hypothesis (SBH)* (Fischer and Engelhardt 2019, [under review](#)):

SBH When frequently used polysemous words have a clearly dominant sense or use whose associated schema is deployed in interpreting less salient uses, the latter uses will prompt inferences licensed (only) by the dominant sense, also in inappropriate contexts.

Conclusions of these inappropriate inferences are particularly likely to go through and influence further reasoning

- (a) where an uninformative context fails to trigger further comprehension inferences that have a bearing on the truth of those conclusions, and
- (b) where any incompatible comprehension inferences from previous text are supported by considerably weaker stereotypical associations than those that need suppressing—which then duly sideline the weak competition (Foss and Speer 1991; Morris 1994).

To derive from the general SBH some word-specific hypotheses which are experimentally testable, we first need to identify words which fit the bill: high-frequency words with a dominant sense deployed in interpreting less salient uses. Fischer and Engelhardt (2017, 2019, [under review](#)) combined a comprehension task with pupilometry and reading-time measurements, respectively, to examine inferences from perception verbs, and provided evidence of inappropriate stereotypical inferences from less salient uses of ‘see’. We now turn to a new, if related, word class, the appearance verbs ‘look’, ‘appear’, and ‘seem’, and extend the investigation in two critical ways that will allow us to gauge the potential relevance of the inappropriate inferences identified by the SBH: We will investigate how robust stereotypical inferences from verbs are in the light of competing pragmatic inferences that may defeat them, and

will examine this not only for English but also for languages with verb-final sentence structure which accord stereotypical inferences from verbs less influence on utterance interpretation.

3.2 Appearance-verbs

The philosophically most relevant use of appearance verbs is with adjectival complement or infinitive. Only one dictionary-attested sense of ‘appear’ (*MEDAL* 1, *OD* 2), ‘seem’ (*MEDAL* 1, *OD* 1), and ‘look’ (*OD* 3, though counted as two senses in *MEDAL*, 3 and 5) allows these syntactic constructions.⁶ For all three verbs, this sense is characterised identically (*WordNet*) or almost identically (*MEDAL*, *OD*), suggesting that, in conjunction with the relevant syntactic cues (Goldberg 2003), all three verbs rapidly (*ibid.*) activate the same associated situation schema (‘appearance schema’).⁷ This schema combines doxastic and experiential elements: ‘give a certain impression or have a certain outward aspect’ (see also *WordNet*). Brogaard (2013, 2014) argues that, in their intransitive sense (‘Joe looks dirty’), appearance verbs function as subject-raising verbs (Postal 1973) which are semantically unrelated to their grammatical subjects (‘Joe’) and serve not so much to predicate any property from their complement (*dirtyness*) of those subjects’ referents (Joe) as to attribute to the often implicit patient an experiential, epistemic, or doxastic attitude towards a content (*Joe is dirty*).

To examine to what extent appearance verbs are stereotypically associated with these different patient properties, Fischer et al. (2015) conducted a distributional semantic analysis (Erk 2012) of the intransitive use of the verbs in a large corpus (a parsed Wikipedia snapshot, which disambiguates polysemous words on the basis of argument structure) (Flickinger et al. 2010). That analysis identified the ‘nearest neighbours’ of each verb, i.e., those verbs that are distributionally (significantly) more similar to them than others. Two predicates, e.g., ‘seem (x, Fx, y)’ and ‘think (y, Fx)’, have a similar distribution to the extent that they co-occur in the corpus with the same other words as arguments, and in the same proportion (!). The five nearest neighbours of ‘appear’ and ‘seem’ included the three doxastic verbs ‘believe’, ‘think’, and ‘find(mental)’, while those of ‘look’ included the latter verb (and ‘think’ and ‘believe’ among the top twenty). For all three verbs, epistemic verbs (like ‘know’ or ‘realise’) were distributionally less similar. Remarkably, the nearest neighbours identified did not include any clearly experiential terms.

Distributionally similar expressions are used interchangeably in a variety of prototypical contexts (after argument swapping). Across such contexts, ‘X seems/appears/looks F to S’ are used interchangeably most often with ‘S thinks/believes that X is F’, less often with ‘S knows that X is F’, and yet less often with experiential terms. We can infer, first, that the intransitive use of appearance verbs

⁶ See relevant entries in: <http://www.macmillandictionary.com/>; <https://en.oxforddictionaries.com/>; cf. <https://wordnet.princeton.edu/wordnet/>. Last accessed 20/05/2018.

⁷ It is an open question whether this generalised situation schema has subordinate schemas that are verb-specific and include components suggested by Austin (1962, pp. 36–37): special circumstances, agent’s possession of inconclusive evidence, etc.

is more frequently employed to attribute doxastic than epistemic or experiential attitudes; and, second, that its stereotypical association with different patient properties varies: The highly salient use at issue is strongly associated with doxastic patient-properties, less strongly with epistemic patient properties, and even more weakly with experiential properties lexicalised by other expressions. These properties are decreasingly strongly integrated in the ‘appearance schema’. Pace Austin (1962, pp. 36–37), our ‘nearest neighbours’ analysis thus suggests that, in their dominant sense, *all* three appearance verbs are stereotypically associated with doxastic patient-properties, and more strongly with these than with others. Behavioural experiments using a forced-choice plausibility-ranking task (*cf.* below, Sect. 4.1) further suggest that doxastic patient-properties are as strongly associated with ‘look’ as with ‘appear’ and more strongly with ‘seem’ (Fischer and Engelhardt 2016).

A less salient special sense is provided by philosophers of perception who have explained a ‘phenomenal’ use of appearance- and perception-verbs, in which these verbs merely serve to describe subjects’ experience, without factive, epistemic, or doxastic implications (*cf.* ‘Hit on the head, he saw stars’) (Ayer 1956, p. 90; Jackson 1977, pp. 33–49; Maund 1986; *cf.* Chisholm 1957, pp. 44–48). For example, in the argument from illusion (see below, Sects. 7.2–7.3), this phenomenal sense, and only this sense, allows them to describe familiar cases of non-veridical perception (where nobody is taken in) by saying, e.g., that a round coin appears elliptical, when viewed sideways. The intended phenomenal interpretation is metaphorical and can be obtained with the common feature-transfer strategy of metaphor interpretation (Bortfeld and McGlone 2001; Ortony 1993; Searle 1993): Subject to contextual constraints, one or more stereotypical implications of a word are chosen as its intended interpretation, and all contextually irrelevant others suppressed (‘Achilles is a lion’—Achilles is [as] noble and courageous [as a lion]). For verbs, this means that one or more components of the stereotypically associated situation schema are retained for interpretation. To interpret the phenomenal use of appearance-verbs, this strategy retains the experiential component of the appearance schema (*S looks at X, X visually looks F to S*) and interprets the utterance as stating that the agent’s experience is similar to that model (*cf.* Ayer 1956, p. 96).

The three conditions identified by our Salience Bias Hypothesis hence apply when appearance verbs are used in the phenomenal sense: (1) These verbs are high-frequency (see *MEDAL*), our nearest neighbours analysis suggests (2) their doxastic use is far more salient than all others, and we have just argued that (3) the retention strategy is used to interpret less salient phenomenal uses. According to our hypothesis, these uses should therefore prompt doxastic inferences (e.g., from ‘X appears F to S’ to *S thinks that X is F*), even in inappropriate contexts (where the viewer does not think that X is F).

3.3 Hypotheses

Our general Salience Bias Hypothesis thus motivates the experimentally testable word-specific hypothesis

H₁ Even in inappropriate contexts which invite phenomenal interpretation of the verb, ‘X looks F [to S]’, ‘X appears F [to S]’, and ‘X seems F [to S]’ all [i] trigger stereotypical inferences to *S thinks that X is F*, and [ii] their conclusions influence further cognition (judgment and reasoning).

Inferences with the I-heuristic are at the bottom of the pragmatic pecking order and can be defeated by conflicting inferences with other heuristics or maxims (Levinson 2000, pp. 157–158): Even when triggered, their conclusions may be swiftly suppressed in face of conflicting inferences and *fail* to influence further cognition. The inferences suggested by H₁ are particularly vulnerable to defeat by inferences from the Maxim of Manner (Grice 1989): In many contexts, speakers might have used ‘is’ instead of appearance verbs, and hearers will infer from preference for these verbs over the simpler and more frequent copula that it is in doubt or contention whether X is F (Grice 1961). This would render it less likely that patients are inclined to think that X is F. We therefore further examine the robustness of the hypothesised stereotypical inferences in the face of competing Manner-inferences.

To develop a testable hypothesis, consider the more precise formulation of the neo-Gricean *M-heuristic* (Levinson 2000, pp. 136–137):

(M-hearer) Where S uses a marked expression in saying ‘p’, and there is an unmarked alternate expression which the speaker might have employed in the same sentence frame, instead, infer that the situation talked about does not conform to the stereotype associated with the unmarked alternate expression.

Where will such inferences be triggered by appearance verbs? These verbs are marked by comparison to ‘is’, in virtue of being less frequent and less neutral in register (Eckman 1977; Shattuck-Hufnagel and Klatt 1979). The M-heuristic will trigger inferences from preference of appearance verbs over ‘is’, when those verbs are used in the same sentence-frames (‘X appears/is F’), i.e., leave the patient implicit, and when ‘is’ is regarded as ‘alternate expression’ to appearance verbs. This will happen only where these verbs, like ‘is’, are interpreted as predicating properties of agents. Since they are typically used, instead, to attribute doxastic and other attitudes to patients (Brogaard 2013, 2014), this only happens where, as in Grice’s (1961) examples, the implicit patient is the speaker herself: Exploiting the stereotypical doxastic implication (‘I, the speaker, am inclined to think that. . .’) then allows speakers to express judgments about the appearance-verbs’ agent X, which they might have expressed more simply by saying ‘X is F’. M-inferences from preference of appearance verbs over ‘is’ will hence only occur in case the patient-role is re-assigned from the verb’s (implicit) patient, to the speaker. Such a re-assignment occurs when it avoids attributing contradictions to the speaker.

Where appearance-verbs take non-perceptual objects as agents (‘The strategy appeared perfect’), contextually inappropriate experiential components of the appearance-schema will be suppressed, and ‘X appears F’ reduces to the attribution of epistemic or doxastic attitudes to the implicit patient. Where sequels are inconsistent with these attributions, recipients will perceive a strong clash, prompting remedial interpretative action in the shape of patient-reassignment, which facilitates M-inferences. By contrast, where appearance-verbs go with visual objects (‘The dress

appeared orange'), the verb-noun combination will reinforce the experiential implications that the patient is looking at the object, and this offers a certain aspect to her (*cf.* Bicknell et al. 2010). Sequels inconsistent with inferred doxastic attributions will therefore be perceived as clashing less strongly, and less likely prompt patient-reassignment. This reasoning motivates hypothesis

- H₂ Inferences with the M-heuristic defeat the posited inferences with the I-heuristic only when appearance-verbs go with non-perceptual objects as agent-role fillers, but not when the verbs go with visual objects (as, e.g., in arguments from illusion).

By examining psycholinguistic hypotheses H₁ and H₂ we seek empirical support for the key idea the Austinian strand of critical ordinary language philosophy relies on: the idea that even competent speakers (like philosophers) make contextually inappropriate stereotypical inferences that influence further judgment and reasoning. To begin examining Austin's further metaphilosophical hypothesis that such inferences are at the root of some philosophical paradoxes, we will deploy psycholinguistic findings to assess a diagnostic hypothesis about a particular example:

- H₃ Arguments from illusion crucially involve contextually inappropriate stereotypical inferences from phenomenal uses of appearance verbs to doxastic conclusions.

That such automatic inferences are crucially involved in these arguments means that they afford the best available explanation of a crucial step made in the arguments. This hypothesis can be followed up in two complementary ways: Analyses of different versions of the argument (below, Sect. 7) can examine to what extent the argument involves the step supposedly best explained by the inferences at issue—i.e., whether this step is indeed crucial. But further experimental work can help us assess whether these inferences indeed best explain the step: Perhaps surprisingly, we can bring findings from cross-linguistic investigations to bear on this issue.

Cross-linguistic investigation can help because the arguments of interest were advanced also in languages with increasingly rigid verb-final sentence structure, including German (reviews: Staudacher 2011; Wiesing 2002) and Japanese (Genka 2017; Nobuhara 1999). Verb-associated stereotypes affect utterance interpretation less strongly in verb-final languages than in verb-medial languages like English: Grammatical information and processing preferences are conventionalised differently in typologically distinct languages (Hawkins 1994)⁸: If the verb comes at the end (German, Japanese), then identification of arguments for the verb employs prior information. By contrast, verb-medial languages (English) allow more verb processing time, but initially leave the object undetermined (Melinger and Maurer 1999; Tanenhaus and Carlson 1989). Therefore, information associated with the verb, including stereotypes, has more time to affect interpretations, and may affect comprehension more strongly, in verb-medial than verb-final languages (Masuoka 2002; Matsui 2000). It is hence a live possibility that stereotypical inferences from appearance verbs do not affect utterance interpretation in German and Japanese strongly enough to drive paradoxical arguments. Since arguments from illusion still had their following in these

⁸ Pre-verb interpretation of nouns is facilitated by morphological case-markers (Bornkessel et al. 2004), which are highly correlated with verb-final languages (Hawkins 2004).

language communities, the inferences of interest would then arguably not offer the best available explanation for any step in the argument—even in English (where they would at best add to the persuasiveness of a step made independently). To examine H_3 , we therefore followed up a study of English with studies of German and Japanese. The stereotypical inferences we identified play a crucial role in the target arguments only if positive findings for English are replicated in these languages.

4 Experiment 1: English

4.1 Approach and predictions

We used a forced-choice plausibility-ranking task. Participants were presented with short (2-sentence) texts⁹ which differ in one critical word ('minimal pairs'):

- 6a. The hill seemed quite steep. The rambler thought it was gentle.
 6b. The hill was quite steep. The rambler thought it was gentle.

Participants are asked to indicate which of the two strikes them as more plausible, even in the absence of a clear-cut preference. Critical items paired sentences using 'look', 'appear', or 'seem' with otherwise identical sentences which employed 'is' (a verb without doxastic implications). The second sentence ('The rambler thought it was gentle') was inconsistent with the posited doxastic inferences from the appearance verb. This conflict invites a phenomenal (re)interpretation of the verb, which lacks doxastic implications. If, even so (by H_1), participants make doxastic inferences from that verb and retain (rather than suppress) the conclusion (here: 'The rambler thought the hill was quite steep'), then the conclusion's persistent clash with the sequel should lower plausibility of 'seem'- etc. sentences (6a), while no such inferences will affect plausibility of 'is'-sentences (6b). Other things being equal (or given exclusion of other relevant factors, see Sect. 4.2), these 'is'-sentences will then strike participants as more plausible than counterparts with appearance verbs. By contrast, if the posited inferences are not triggered, or are swiftly suppressed, preferences should be random (i.e., no difference in plausibility). In view of conflicts with background beliefs or simultaneous inferences, initial inferences can be swiftly suppressed, and fail to influence plausibility judgments in as little as 1 s after the sentence (Fischer and Engelhardt 2017). While it cannot assess whether inferences are first triggered and then suppressed, the forced-choice paradigm is well suited to examining whether inferences are triggered *and influence further cognition*.

Since the task requires participants to compare sentences differing in one word, this paradigm lends itself particularly well to studying the robustness of stereotypical inferences (with the I-heuristic) in the light of competing pragmatic inferences with the M-heuristic. The reasoning motivating H_2 has us complement items with visual objects (like 6) with items containing non-perceptual objects:

⁹ In view of criticism against experimental philosophy from 'radically contextualist' OLP (Baz 2017), note the context-free presentation of brief items is well suited precisely to study automatic inferences from brief premises of philosophical arguments, presented out of any (non-theoretical) context.

19a. The plan looked good. Cole believed it was terrible.

19b. The plan was good. Cole believed it was terrible.

If participants re-assign the implicit patient role of the first sentence from Cole to the statement's author, they will infer with the M-heuristic from (19a) that the quality of the plan was in doubt. This makes it more plausible that Cole believed it was terrible, and attenuates preferences for 'is'-sentences (19b). H_2 predicts such attenuation for items involving non-perceptual, but not visual objects.

We thus infer from our first two hypotheses that

[Prediction 1] participants will prefer (find more plausible) 'is'-sentences over all alternatives, in pairs with visual objects, and

[Prediction 2] this (plausibility) preference will be attenuated, possibly to the point of randomness, in pairs with non-perceptual objects.

H_3 predicts that positive findings for English will be replicated in German and Japanese.

By inviting comparisons and providing salient triggers for the M-heuristic, our task seems to provide the strongest possible invitation to make defeating M-inferences. Any such inferences picked up (through decreased preference for 'is'-sentences) may therefore be task artefacts. But, by the same token, significant preferences for 'is'-sentences will provide strong evidence that the stereotypical (I-heuristic) inferences of interest are not defeated by competition (M-heuristic inferences), in the contexts of interest (arguments from illusion).

4.2 Methods

4.2.1 Participants

Seventy-three undergraduate psychology students from the University of East Anglia participated for course credit. All were native speakers of English. Four were bilingual.

4.2.2 Materials

Participants received a pen-and-paper questionnaire with 120 items, and were asked to indicate which of each pair 'strikes you as more plausible'. The questionnaire contained 36 critical items, 12 for each pairing 'look'/'is', 'appear'/'is', and 'seem'/'is'. For each pairing, half of the items had visual objects, and half had non-perceptual objects. Items with visual objects attributed to them colour, shape, and size (one each per verb) or more complex visually ascertainable properties, e.g., 'Tom's shoes appeared/were dirty. Tom believed they were clean.' Items were controlled for verb-order: In items with visual and non-perceptual objects, respectively, each critical verb appeared in sentence (a) precisely half the time. Cancellation sentences employed 'believed' or 'thought', in equal number.

When participants cannot base plausibility assessments on factual knowledge (as in uninformative items, e.g., about fictitious people and their shoes), participants base them on metacognitive cues, in particular on the level of fluency or subjective ease they

experience in processing the sentence(s) (review: Alter and Oppenheimer 2009). We predicted differences in plausibility due to the presence versus absence of stereotypical inferences to conclusions inconsistent with the sequel: Where incongruence engenders difficulty or ‘dysfluency’, it leads to lower (subjective) plausibility. While most factors that influence fluency, including familiarity and pronounceability of individual words (Oppenheimer 2006), syntactic complexity of the sentence (Lowrey 1989), and priming by earlier words, are controlled by using minimal pairs or exclusion norming (below), we cannot eliminate differences in frequency and hence familiarity between appearance verbs and ‘is’ (Supplementary Appendix, Section A).

To examine the extent in which participants’ plausibility rankings are influenced by word frequency, we constructed 30 minimal pairs whose critical verbs differed in frequency. Each of the 15 verbs occurred once in a ‘*frequency-congruent*’ item where the text employing the more frequent verb was also more consistent with its associated stereotype, and once in a ‘*frequency-reversed*’ item, where word-frequency and stereotype-consistency work in opposite directions. E.g., the more frequent verb ‘obey’ stereotypically implies submission to formal authority:

[*frequency-congruent*]

80a/b. The colonel told the captain not to change his company’s position until further notice. The captain thought this reckless but obeyed/complied.

[*frequency-reversed*]

18a/b. Jane asked the campers on her land to move somewhere else by tomorrow afternoon. They weren’t happy but complied/obeyed.

If participants make judgments predominantly in line with stereotype-consistency, and do not make fewer such judgments about frequency-reversed than frequency-congruent items, their plausibility judgments are unlikely to be influenced by frequency.

4.2.3 Procedure and design

In constructing frequency-congruent and -reversed fillers, we used a written British English corpus (Leech et al. 2001), appropriate for our participants.

In a prior norming study, four philosophy graduate students responded to a draft questionnaire and then explained their responses with the first reasons coming to mind, first independently on paper, then in group discussion. To exclude the influence of associative priming, which makes sentences ‘sound better’ or ‘more idiomatic’, we discharged or modified critical items where even one participant mentioned that one formulation ‘sounded better’ or ‘more idiomatic’ than the other. To exclude extraneous content-based considerations, we also changed items where even one participant offered such considerations. As frequency-congruent and -reversed filler items, we only kept items where at least two participants independently invoked the same stereotypical association of the critical verb, all agreed in subsequent discussion that the verb had these ‘connotations’ or ‘associations’, and no other association was mentioned by more than one student.

In the actual study, participants were instructed to ‘read each text carefully and then respond as quickly as you can’, to ensure responses in under 5 s, before controlled processes could modify automatic cognition (De Neys 2006). Brief discussion of three practice items ensured understanding of the procedure. To prevent fatigue, participants were instructed to take a 5 min break half way (remaining seated, no phones).

We manipulated verb (look/appear/seem) and object (visual/non-perceptual) within subject, in a 3×2 design, and measured consistency of preference, which admits parametric tests. Predicted responses were coded as ‘1’, ‘incorrect’ responses as ‘0’. Data were screened for outliers prior to analysis, but there were no datapoints greater than 3 SDs from the mean in any condition.

4.3 Results

For the reader’s convenience, results for all languages (Experiments 1–3) are presented together, in Tables 1 and 2.¹⁰ To preview results, both predictions were confirmed, for all languages: As per Prediction 1, participants had a significant preference for ‘is’-sentences with visual objects over counterparts using any appearance verb. As per Prediction 2, preferences were attenuated for ‘is’-sentences with non-perceptual objects, in many conditions to the point of randomness. Given the binary choice, preferences are random when ‘is’-sentences are deemed more plausible than counterparts half the time. Whether preferences were significant or random was determined with one-sample *t*-tests with a test-value of .5.

For English, a 3×2 (verb \times object) repeated measures ANOVA revealed no significant interaction ($p > .50$), but showed a significant main effect of verb $F(2,144) = 14.05, p < .001, \eta^2 = .16$ and object $F(1,72) = 8.40, p = .005, \eta^2 = .10$. That is: Which appearance-verb was used made a difference for participants’ plausibility preferences, and whether sentences had a visual or non-perceptual object also affected preferences. Preferences concerning ‘look’-items (mean across objects .59) and ‘appear’-items (mean across objects .59) were not significantly different from each other ($p > .69$), but ‘is’ preferences were significantly stronger for them than for ‘seem’-items (mean across objects .51) (‘look’ vs ‘seem’ $t(72) = 4.43, p < .001$; ‘appear’ vs ‘seem’ $t(72) = 4.94, p < .001$). The significant effect of object was based on higher preferences for ‘is’-sentences with visual objects than with non-perceptual objects (means across verbs .64 vs .48). In the absence of a significant interaction, there is no statistical support for more detailed comparisons. As predicted, however, ‘is’-preferences were significant for items with all three appearance-verbs and visual objects, and attenuated to randomness, or even reversed, for items with non-perceptual objects (Table 1).¹¹

¹⁰ To assess whether there were significant differences within the full dataset, we conducted an omnibus analysis which included language as a variable. This analysis helps guard against family-wise Type I error, and reveals whether some conditions differ significantly from each other. A $3 \times 3 \times 2$ (language \times verb \times object) repeated measures ANOVA showed a significant main effect of verb $F(2,414) = 10.51, p < .001, \eta^2 = .05$, object $F(1,207) = 22.36, p < .001, \eta^2 = .10$, and language $F(2,207) = 6.73, p = .001, \eta^2 = .06$. The 3-way interaction was also significant $F(4,414) = 2.41, p = .049, \eta^2 = .02$.

¹¹ In view of forceful arguments that correction for multiple comparisons is too conservative (e.g., Armstrong 2014; Cabin and Mitchell 2000; Nakagawa 2004), we used the conventional significance threshold

Table 1 Means (SDs) for verb and object broken down by language

	Visual			Non-visual		
	Look	Appear	Seem	Look	Appear	Seem
English	.68 (.28)***	.66 (.30)***	.59 (.27)**	.49 (.28)	.52 (.30)	.43 (.29)*
German	.73 (.28)***	.78 (.25)***	.69 (.28)***	.65 (.31)***	.67 (.29)***	.57 (.30)
Japanese	.57 (.28)*	.58 (.31)*	.66 (.27)***	.56 (.30)	.56 (.27)*	.51 (.29)

Significant one-sample *t* tests (test value .5) are indicated (* $p < .05$, ** $p < .01$, *** $p < .001$)

Table 2 Means (SDs) for verb and object broken down by language, after exclusion of potentially frequency-sensitive responders

	Visual			Non-visual		
	Look	Appear	Seem	Look	Appear	Seem
English	.72 (.23)***	.69 (.27)***	.63 (.26)**	.46 (.24)	.55 (.29)	.44 (.28)
German	.74 (.27)***	.77 (.28)***	.69 (.30)**	.59 (.34)	.68 (.31)**	.57 (.32)
Japanese	.59 (.28)*	.60 (.31)*	.68 (.26)***	.59 (.29)*	.58 (.27)*	.55 (.27)

English ($N = 44$), German ($N = 35$), Japanese ($N = 64$). Significant one-sample *t*-tests (test value of .5) are indicated (* $p < .05$, ** $p < .01$, *** $p < .001$)

To examine the potential influence of word frequency, we used responses to filler items to identify participants whose responses might be influenced by frequency: those who give ‘correct’ (stereotype-consistent) responses more frequently for frequency-congruent than for frequency-reversed fillers (where frequency and stereotypes work in opposite directions). We derived relevant criteria empirically, by considering meaningful gaps in the distribution of responses (Supplementary Appendix, Section B). In this experiment, we defined ‘potentially frequency-sensitive responders’ as those who responded ‘correctly’ to over 70% of frequency-congruent fillers but to under 70% of frequency-reversed items. 29 participants met this criterion and were excluded from further analysis. Results for the 44 remaining ‘frequency-insensitive’ participants showed the same pattern as those for the overall sample: We observed no significant interaction ($F(2,86) = 2.59, p = .08, \eta^2 = .06$), but main effects of object ($F(1,43) = 9.15, p = .004, \eta^2 = .18$) and verb ($F(2,86) = 8.47, p < .001, \eta^2 = .17$), and similar paired comparisons (look = appear; look > seem; appear > seem) (appear vs. look: $t(43) = 1.10, p = .28$; appear vs. seem: $t(43) = 3.91, p < .001$; look vs. seem: $t(43) = 3.03, p = .004$). ‘Is’-preferences remained significant for items with all appearance-verbs and visual objects (Table 2). This excludes word-frequency as a confound, and allows us to interpret results as indicative of the hypothesised stereotypical inferences from appearance-verbs.

Footnote 11 continued

of .05. However, all *p*-values reported in Tables 1 and 2 as < .01 would retain significance upon Bonferroni correction.

5 Experiment 2: German

To identify the relevant German verbs, we consulted translations of English philosophical texts discussing the argument from illusion, and two major German-English dictionaries. Austin (1975, transl. of 1962) translates ‘look’ as ‘aussehen’, ‘appear’ as ‘erscheinen’, and ‘seem’ as ‘(zu sein) scheinen’, as does Staudacher (2011, p. 74, Fn.85). In translating from Ayer (1940), Wiesing (2002, p. 3) renders ‘look’ as either ‘aussehen’ or ‘erscheinen’, ‘appear’ as either ‘erscheinen’ or ‘scheinen’, and ‘seem’ as ‘scheinen’ or ‘zu sein scheinen’, while Russell (1967, transl. of 1912) uses both ‘aussehen’ and ‘erscheinen’ for ‘look’, and ‘erscheinen’ for both ‘appear’ and ‘seem’. The *Duden-Oxford* (Scholze-Stubenrecht and Sykes 1999) translates relevant senses of ‘look’ (sense indicator: *appear*) as ‘aussehen’, and ‘appear’ (sense indicator: *seem*) and ‘seem’ (sense indicator: *appear*) as ‘(zu sein) scheinen’. The *Great Muret-Sanders* (Springer 2000) concurs and adds the more explicitly doxastic ‘den Anschein haben’ and ‘den Eindruck erwecken’ for ‘appear’ (sense 3), as well as ‘anscheinend sein’ and ‘erscheinen’ for ‘seem’ (sense 1). We therefore examined ‘aussehen’, ‘erscheinen’, and ‘(zu sein) scheinen’, using the same approach as for English.

5.1 Methods

5.1.1 Participants

48 undergraduate philosophy students from the University of Cologne participated without remuneration. All were native speakers of German.

5.1.2 Materials

Participants received a pen-and-paper questionnaire with 66 items, and the same instructions as in English. The questionnaire contained translations of the 36 critical items, translating ‘look’ with ‘aussehen’, ‘appear’ with ‘erscheinen’, and ‘seem’ with ‘scheinen (zu sein)’. Half of the 30 fillers were ‘frequency-congruent’, the other half were ‘frequency-reversed’ items with the same verb (see Sect. 4.2.2). Critical items were slightly modified in translation, where this was necessary to retain idiomaticity. This included using an infinitive (‘scheint zu sein’) for half of visual and non-perceptual ‘seems’ items. All other critical items used an adjectival complement. Cancellation phrases employed ‘dachte’ (‘thought’) or ‘glaubte’ (‘believed’). Items were controlled for verb-order.

5.1.3 Procedure and design

In constructing frequency-congruent and -reversed filler items, we used word-frequency information from the Leipzig University *Wortschatz*.¹² A prior norming

¹² <http://corpora.informatik.uni-leipzig.de/>, using a German newspaper corpus (deu_newscrawl_2011).

study with 4 participants followed the same protocol as Experiment 1, as did the main study (except that no break was required). Design and coding were the same as before.

5.2 Results

A 3×2 (verb \times object) repeated measures ANOVA revealed no significant interaction ($p > .70$) but showed significant main effects of verb ($F(2,94) = 8.53, p < .001, \eta^2 = .15$) and object ($F(1,47) = 14.46, p < .001, \eta^2 = .24$). The main effect of verb was due to 'is'-preferences concerning 'look'-items (mean across objects .69) and 'appear'-items (mean .73) being higher than for 'seem'-items (mean .63). Preferences concerning 'look' and 'appear' did not differ significantly ($t(47) = 1.74, p = .088$), but were significantly stronger for them than for 'seem' ('look' vs. 'seem' $t(47) = 2.29, p = .026$; 'appear' vs. 'seem' $t(47) = 4.28, p < .001$). The main effect of object was due to significantly higher 'is'-preferences for items with visual objects (mean across verbs .74) than with non-perceptual objects (mean .63) ($t(47) = 3.80, p < .001$). As predicted, 'is'-preferences were significant for items with all appearance-verbs and visual objects, and attenuated with non-perceptual objects, to the point of randomness for 'seem' (Table 1).

We again used gaps in the distribution of data to derive a criterion to identify potentially frequency-sensitive responders. In this study, participants qualified when responding 'correctly' to over 90% of frequency-congruent fillers but to under 90% of frequency-reversed items. 13 participants met this criterion and were excluded from further analysis. Results for the 35 remaining participants showed the same pattern as the overall sample: We observed no interaction ($F(2,68) = .63, p = .537, \eta^2 = .02$) but main effects of object ($F(1,34) = 12.15, p = .001, \eta^2 = .26$) and verb ($F(2,68) = 5.47, p = .006, \eta^2 = .14$). However, 'is'-preferences were now marginally lower concerning 'look'-items (mean across objects .67) than 'appear'-items (mean .72) ($t(34) = 1.90, p = .066$), and no longer significantly higher than for 'seem'-items (mean .63) ($t(34) = 1.37, p = .179$), while 'is'-preferences were still significantly higher concerning 'appear'-items than 'seem'-items ($t(34) = 3.31, p = .002$); 'is'-preferences remained significant for all items with visual objects (Table 2). Again, this excluded frequency as a confound.

Half the German 'seem'-items (with visual and non-perceptual objects), employed an infinitival construction, which should strengthen the doxastic implications. We therefore predicted stronger 'is'-preferences for infinitival than adjectival items with visual objects, and expected stronger interference of the M-heuristic in items with non-perceptual objects, resulting in a more pronounced difference between 'is'-preferences concerning visual and non-perceptual objects for infinitival than for adjectival 'seem'-items. A 2 (infinitive vs. adjective) $\times 2$ (visual vs. non-perceptual object) repeated measures ANOVA revealed a (marginal) interaction ($F(1,47) = 3.97, p = .052, \eta^2 = .08$) and a main effect of object ($F(1,47) = 8.00, p = .007, \eta^2 = .15$), allowing us to make the relevant comparisons. We indeed observed that 'is'-preferences for adjectival items with visual objects were marginally lower than for infinitival items with such objects ($t(47) = -1.80, p = .078$), and that 'is'-preferences for infinitival items with non-perceptual objects were significantly lower than for infinitival

items with visual objects ($t(47) = -3.47, p = .001$), while there was no significant difference for the corresponding adjectival items ($t(47) = -.82, p = .419$). Crucially, while ‘is’-preferences concerning infinitival items with visual objects (mean .75) were very pronounced ($t(47) = 5.06, p = .000$), preferences concerning their adjectival counterparts (mean .66) remained significant ($t(47) = 3.59, p = .001$). The posited inferences are therefore not due to the infinitival construction, which merely reinforces them.

6 Experiment 3: Japanese

To identify the relevant verbs in Japanese, we consulted Japanese translations of philosophical texts discussing arguments from illusion (Austin 1984, transl. of 1962; Ayer 1991, transl. of 1940; Ayer 1981, transl. of 1956; Russell 1964, 1965, 2005, three transl. of 1912). We adopted the only consistent translation schema extant, provided by Austin (1984), which renders ‘seem’ as ‘omowareru’, and ‘look’ and ‘appear’ as ‘mieru’, using the Kanji and the hiragana character ‘mi’, respectively.¹³ ‘Omowareru’ is a derivative of the polysemous but clearly doxastic verb ‘omou’.¹⁴ Standard dictionaries explain ‘omowareru’ in the relevant syntactic constructions as ‘something gives some impression’ (Minamide and Nakamura 2011, p. 268) and suggest ‘look’, ‘appear’, and ‘seem’ as translations. ‘Mieru’ with kanji and hiragana ‘mi’ are largely interchangeable derivatives of the visual verb ‘miru’.¹⁵ Standard dictionaries explain their meaning with the relevant syntactic constructions as ‘some impression is given’ or ‘something is interpreted/supposed as such and such’, and again suggest ‘look’, ‘appear’, and ‘seem’ as translations (Minamide and Nakamura 2011, p. 1720; Watanabe et al. 2003, p. 2486). The two versions of ‘mieru’ differ in that Kanji symbols convey meanings and sounds, whereas hiragana only represent sounds. Given processing differences between these scripts (Goryo 1987; Sasanuma 1980), the Kanji character ‘mi’, connoting (literal or metaphorical) vision, may suggest more strongly that visual perception is involved.

Whereas the German verbs of interest preceded their complements, the Japanese verbs appeared in sentence-final position, behind their complement, as in the translation of ‘The coin looks (seems) elliptical’:

Coin/wa/daen/ni/mieru (omowareru).

The coin/[topic-marker]/elliptical/[case forming particle]/looks (seems).

¹³ Ayer (1991) employs the same translations, but uses both spellings of ‘mieru’, in translating ‘appear’. The different translations of Russell (1912) render both ‘look’ and ‘seem’ as ‘mieru’ with Kanji ‘mi’, but ‘appear’ variously as ‘mieru’ with Kanji or hiragana, ‘omowareru’, ‘arawareru’, and ‘arawareteiru’. The last two are idiomatic only where ‘appear’ applies in the sense ‘become visible’, despite sharing kanji with the noun for ‘appearance’ (possibly motivating this problematic translation), and we disregarded them. Ayer (1981) also translates ‘look’ as ‘yōsu wo shiteiru’, but this is mainly used to express hedged judgments and has strong doxastic implications.

¹⁴ ‘think’, ‘intend’, ‘judge’, ‘evaluate’, ‘imagine’, ‘assume’, ‘expect’, ‘pray’, ‘desire’, ‘believe’, ‘suspect’ and ‘recall’ (Watanabe et al. 2003, pp. 426–428). See also <http://jisho.org/>.

¹⁵ ‘see’, ‘look at’, ‘witness’, ‘stare at’, ‘watch’, and ‘attend’ (Watanabe et al. 2003, p. 2513).

6.1 Methods

6.1.1 Participants

89 undergraduate humanities and social science students from Musashino University, Tokyo, participated without remuneration. All were native speakers of Japanese.

6.1.2 Materials

Participants received a pen-and-paper questionnaire with 66 items, and the same instructions as in the previous study. 12 ('look') items employed the verb 'mieru' with Kanji 'mi'; 12 ('appear') items used 'mieru' with hiragana 'mi'; and 12 ('seem') items employed 'omowareru'. 19 critical items employed the construction with 'ni'; 9 used the auxiliary verb 'yôni'. Cancellation sentences used the verbs 'kangaeta' (= thought) or 'omowareta' (= believed) in equal number. The grammatical subject of the cancellation sentences was accompanied by the topic-marker 'wa', which does not determine by itself whether the agent of the second sentence is identical with the implicit patient from the first. 30 fillers were constructed as before. As in Exp. 1–2, items were controlled for word-order.

6.1.3 Procedure and design

In constructing frequency-congruent and -reversed filler items, we used word-frequency information from *NINJAL-LWP for BCCWJNational*.¹⁶ A prior norming study with 6 participants followed the same protocol as English and German studies. The protocol, design and coding were as before.

6.2 Results

A 3×2 (verb \times object) repeated measures ANOVA showed a significant interaction ($F(2,176) = 8.62, p < .001, \eta^2 = .09$) and a significant main effect of object ($F(1,88) = 9.41, p = .003, \eta^2 = .10$), though not of verb ($p > .40$). The significant interaction permits more fine-grained comparisons. We first considered 'is'-preferences for items with visual vs. non-perceptual objects, for each type of verb. 'Is'-preferences were significantly higher concerning items with visual objects for 'seem'-items (means .66 vs. .51; $t(88) = -5.54, p < .001$), but not for 'look'- and 'appear'-items (p 's $> .60$). We then compared preferences for items with visual objects and different verbs. 'Is'-preferences were similar for 'look'- and 'appear'-items ($t(88) = -.39, p = .701$), and significantly higher for 'seems' (look vs. seem $t(88) = -3.24, p = .002$; appear vs. seem $t(88) = -3.24, p = .002$). For items with non-perceptual objects, 'is'-preferences were similar for 'look'- and 'appear'-items (both means .56), but numerically lower for 'seem'-items, though only the difference between 'appear' and 'seem' was statistically (marginally) significant ($t(88) = 1.97, p = .052$). As predicted, 'is'-preferences were

¹⁶ <http://nlb.ninjal.ac.jp/>, using *The Balanced Corpus of Contemporary Written Japanese*.

significant for items with all appearance-verbs and visual objects, while attenuation with non-perceptual objects was only observed with ‘seem’ (Table 1).

We identified potentially frequency-sensitive responders in the same way as in the German study (Exp.2). Results for the remaining 64 clearly frequency-insensitive responders (Table 2) showed exactly the same pattern as those for the overall sample: We observed a significant interaction ($F(2,126) = 4.67, p = .02, \eta^2 = .07$) and main effect of object ($F(1,63) = 4.38, p = .04, \eta^2 = .07$), but not of verb ($p = .46$). The paired comparisons were similar (seem-visual vs. seem-nonvisual $t(63) = -3.78, p < .001$; look-visual vs. seem-visual $t(63) = -2.53, p = .014$; appear-visual vs. seem-visual: $t(63) = -2.46, p = .017$). Again, we can exclude word frequency as a confound.

7 General discussion

7.1 Main findings

Our three cross-linguistic experiments (English, German, Japanese) followed up the Saliency Bias Hypothesis (SBH): When frequently used polysemous words have a clearly dominant sense or use whose associated schema is deployed in interpreting less salient uses, the latter uses will prompt inferences licensed (only) by the dominant sense, also in inappropriate contexts. Our experiments examined, in the first instance, two hypotheses about such inferences from appearance verbs, and their robustness: (H₁) ‘X looks F’, ‘X appears F’, and ‘X seems F’ [to S] as well as German and Japanese equivalents all prompt stereotypical inferences (with the I-heuristic) to *S thinks that X is F*, whose conclusions frequently influence further cognition (judgment and reasoning), even in inappropriate contexts which invite a phenomenal interpretation of the verb. (H₂) Even where sequels are inconsistent with them, such doxastic inferences from appearance verbs are defeated by competing pragmatic inferences (with the M-heuristic) only when these verbs are paired with non-perceptual objects, but not when they go with visual objects. In the plausibility-ranking paradigm we used, these hypotheses translated into predictions about (plausibility) preferences for ‘is’-sentences over counterparts using appearance-verbs. These predictions were borne out, for all three languages.

As per Prediction 1, we observed significant preferences for ‘is’-sentences in items with visual objects, with all appearance verbs. This suggests that hypothesised doxastic inferences are made and influence further cognitive processing, when sentences combine appearance-verbs with visual objects as agent-role fillers. As per Prediction 2, we observed that preferences for ‘is’-sentences were lower with non-perceptual objects than with visual objects: We measured a significant main effect of object, across verbs, in all three languages.¹⁷ In English and German, ‘is’-preferences were numerically lower with non-perceptual objects for all three verbs. In English, they

¹⁷ We measured medium-size effects in German. In English and Japanese, effects were comparatively small, as often in social science experiments (Rosnow and Rosenthal 2003); but they were well above the anchor point of $\eta^2 = .04$ for practical significance (Ferguson 2009). Further relevant sources of variance (individual differences in verbal IQ, inhibitory abilities, etc.) are unfortunately ill understood at this point.

became random, for all three verbs, upon exclusion of potentially frequency-sensitive responders; in German, they remained significant only for ‘appear’. In Japanese, ‘is’-preferences were significantly lower (attenuated to randomness) for ‘seems’-items with non-perceptual than with visual objects, though the object-manipulation made no difference for preferences concerning ‘mieru’-items. The observed attenuation of ‘is’-preferences provides evidence that doxastic inferences with the I-heuristic were frequently countermanded by inferences with the M-heuristic, which were probed by our task.

These findings provide additional experimental support for the Salience Bias Hypothesis: They complement prior findings about perception verbs (Fischer and Engelhardt 2017, 2019, under review) with fresh findings about appearance verbs. Second, these findings reveal the robustness of the inappropriate stereotypical inferences posited by the SBH: Our experimental task and design created the most favourable conditions for defeating stereotypical inferences with the M-heuristic. Even so, such defeat occurred only where appearance verbs took non-perceptual objects as agent-role fillers, while doxastic inferences with the I-heuristic went through undefeated where the verbs took visual objects (*pace* Grice 1961). Third, cross-linguistic replication for increasingly rigid verb-final languages (German, Japanese), where verb-associated stereotypes influence utterance interpretation less strongly than in verb-medial English, suggests that inappropriate but influential stereotypical inferences are not only triggered by words which play a pivotal role in utterance interpretation, but more generally.

Finally, we found that ‘look’ and ‘appear’ behaved the same in all languages while ‘seem’ behaved differently. This is least surprising in Japanese, where we were merely dealing with different spellings of the same verb ‘mieru’. But it seems inconsistent with Austin’s (1962, pp. 36–37) hypothesis that ‘appear’ and ‘seem’ are more closely tied to judgment and belief than ‘looks’ (which supposedly is typically used to comment on the mere look of things): Not ‘look’, but ‘seem’, is the odd one out in terms of doxastic inferences.¹⁸ That even the ordinary language philosopher credited with the most extraordinary ‘genius for spotting linguistic differences and distinctions’ (Searle 2001, p. 226; *cf.* Cavell 1994, p. 21; Williams 2014, p. 44) misperceived the relevant linguistic pattern, forcefully illustrates how useful it is to complement informal with formal experiments.

By supporting the SBH, our findings support the psycholinguistic assumption underwriting Austin’s brand of critical ordinary language philosophy: Under certain conditions, even competent speakers make contextually inappropriate stereotypical inferences that go through undefeated and influence further judgment and reasoning. The SBH identifies a first relevant set of conditions. We now turn to the metaphilosophical hypothesis the Austinian approach seeks to make good: Such inappropriate inferences are at the root of some philosophical paradoxes and problems. Austin (1962) examines one such paradox, the ‘argument from illusion’. We will now follow up the

¹⁸ Since stereotypical association with doxastic patient-properties is stronger for ‘seem’ than ‘look’ or ‘appear’ (Fischer and Engelhardt 2016), we submit our ‘seem’-items are perceived as more contradictory. In our deliberately artificial forced-choice setting (Sect. 4.1), this prompts occasional reference reassignment and M-heuristic inferences (Sect. 3.3), even with visual objects, and thus leads to the lower ‘is’-preferences observed in this study (Exp.1, replicated for German in Exp.2).

metaphilosophical hypothesis by exploring to what extent our word-specific hypotheses (H_1) and (H_2) help develop the specific diagnostic hypothesis that has already received some preliminary support from the replication of English results in verb-final German and Japanese (see end Sect. 3.3):

H_3 The ‘argument from illusion’ relies on contextually inappropriate stereotypical inferences from phenomenal uses of appearance verbs (in the argument’s initial premises) to doxastic conclusions.¹⁹

The development of H_3 will illustrate, more generally, how experimental investigations into automatic cognitive processes can support the analysis of philosophical arguments and, specifically, ‘diagnostic’ analyses that seek to expose fallacies. The more blatant the identified fallacies are, once spelled out, the more strongly the proposed analysis violates hermeneutic principles of charity (Adler 1994; Lewinski 2012), and the more strongly it is in need of empirical validation (Thagard and Nisbett 1983). Such validation is provided by psychological explanations of when and why competent thinkers should commit the fallacies at issue (see above, Sect. 3). In line with general trends in cognitive science (reviews: Kahneman 2011; Wilson 2002), experimental philosophy has begun to examine how philosophical thought is shaped by largely unconscious automatic processes into which only experiments give us insight (see above, Sect. 1). Where an automatic inference process leads from an explicit premise to a conclusion that remains implicit but provides input for another cognitive process that generates an explicit conclusion, a thinker may leap from explicit premise to explicit conclusion, without awareness of the initial inference and the intermediate conclusion. If that initial automatic inference is inappropriate, the fallacy will be committed below the radar of the thinker’s conscious awareness.

We will now develop (H_3) in conjunction with the suggestion that this happens in the argument from illusion: that the automatic comprehension inferences explained by the SBH and documented by our experiments provide input for a well-researched judgment heuristic that delivers an explicit verdict, without thinkers being aware of the initial inappropriate inference or its implicit conclusion (Sect. 7.2). Then we will explore to what extent (H_3) can contribute to resolving the paradox (Sect. 7.3), and clarify the sense in which it might help ‘dissolve’ a philosophical problem (Sect. 7.4). We thus contribute towards first proof of concept for an experimental implementation of Austin’s critical project that may provide fresh inspiration for restrictionist experimental philosophy (*cf.* Sect. 1).

7.2 Reanalysing the argument from illusion

The commonsense conception of sense-perception as experiential awareness mainly of material objects has been challenged by arguments that proceed from two kinds of cases: (1) from mostly familiar cases of non-veridical perception (‘illusions’) where physical objects look or otherwise appear to have a shape, size, colour, or other property they do not actually possess, and (2) from often fictitious cases of ‘hallucination’,

¹⁹ Austin (1962) discusses appearance verbs (see above, Sect. 2), but does not deploy conclusions to analyse the argument from illusion. H_3 was first mooted by Fischer (2014a).

where someone has an experience as of perceiving a physical object though no suitable object is actually around. These arguments lead to the same conclusion and the label ‘argument from illusion’ has sometimes been loosely applied to both. Despite occasional assimilation (e.g., Ayer 1940, p. 3; Fish 2010, pp. 12–13), however, these arguments are now generally treated as distinct (Crane and French 2015, Smith 2002). We here consider ‘arguments from illusion’ in the now common stricter sense, which proceed from cases of the first kind.²⁰ E.g.:

1. When a subject looks at a round coin sideways, the coin appears elliptical to her.

Seminal statements of the argument (e.g., Hume [1748] 1739, p. 152) infer directly that, in these cases, an ‘image’ (aka ‘sense-datum’) rather than a physical object must be ‘present to the mind’. Subsequent proponents of the argument then sought to make explicit, and rationalize, the implicit reasoning driving this (to them) intuitively compelling leap of thought. Austin addresses early 20th century statements (e.g., Ayer 1940, p. 4; Broad 1923, p. 240; Price 1932, pp. 27–30; Russell 1912, pp. 1–3), which break this decisive ‘sense-datum inference’ (Smith 2002, p. 25) up into two parts. From (1) they infer the ‘*negative conclusion*’:

2. When a subject looks at a round coin sideways, she is not (directly) aware of the round coin.

The positive conclusion that subjects are, instead, aware of an ‘elliptical sense-datum’ is then obtained from an uncontroversial response to (2):

3. When a subject look at a round coin sideways, she is (directly) aware of *something*.
4. By (2) & (3), the subject is then (directly) aware of something other than the round coin (namely, a ‘sense-datum’).

The sense-datum is then credited with the shape, size, and colour that the coin merely looks (there and then). This yields so-called ‘phenomenal judgments’, such as ‘The object that viewers are then (directly) aware of is elliptical.’ The remainder of the argument then generalises from (4) to argue that in all cases of visual perception, we are (directly) aware of sense-data.

This early 20th century version has been superseded by other versions of the argument, in current debates (Robinson 2001, pp. 57–58; Smith 2002, pp. 25–27; cf. Crane and French 2015; Fish 2010, pp. 12–13). However, analysis of the currently most prominent version (in Sect. 7.3) will benefit from prior analysis of its predecessor. We therefore now reconstruct how contextually inappropriate inferences may drive the crucial inference from (1) to (2) above. To do so, we turn from explicit reasoning to automatic inferences that remain implicit. Proponents of the argument divided their ‘sense-datum inference’ into a negative step (from 1 to 2) and a positive step (via 3 to 4). We divide the negative step, in turn, into two automatic inferences governed by well-researched heuristics: We suggest that comprehension inferences with the I-heuristic provide input for judgments with the representativeness heuristic.

Statements of arguments from illusion occasionally use the appearance verb ‘look’ (e.g., Ayer 1940, p. 4) but mainly ‘appear’ (e.g., Ayer 1940, p. 3; Fish 2010, pp. 12–13;

²⁰ For a parallel examination of ‘arguments from hallucination’, see Fischer and Engelhardt (2017, 2019, under review).

Robinson 2001, p. 57; Russell 1912, p. 2; Smith 2002, p. 25) and ‘seem’ (e.g., Ayer 1940, p. 3; Broad 1923, pp. 239–240; Crane and French 2015, p. 3; Moore 1918/19, pp. 21–23; Russell 1912, p. 2).²¹ These verbs are used here in a phenomenal sense devoid of factive, epistemic, or even doxastic implications (see above, Sect. 3.2). The Salience Bias Hypothesis explains why such uses trigger contextually inappropriate doxastic inferences and our experiments show such inferences are made and go through in uninformative contexts. Such inferences with the I-heuristic lead from the initial premise (1) to the implicit conclusion:

(C) The viewer thinks that the object viewed is elliptical.

Since the appearance verb goes with perceptual objects in all relevant premises, this inference will not be defeated by inferences with the competing M-heuristic from the same verb (Sect. 7.1).

But will the inference get defeated by conflicting inferences from other words figuring in statements of arguments from illusion? We identified two conditions under which the inferences of interest are particularly likely to go through undefeated (see end Sect. 3.1). Both are satisfied by statements of arguments from illusion: (a) Some statements provide poor context, and the case description triggers no further comprehension inferences with a bearing on the truth of doxastic conclusions, as in ‘When partially immersed in water, the straight stick looks bent’ (Ayer 1940, p. 3). Hence the doxastic conclusions are not suppressed. (b) Other statements do trigger incompatible inferences, but these are supported by weaker stereotypical associations. E.g., ‘When subjects *look at* a round coin sideways...’, will trigger stereotypical inferences to ‘The viewer knows there is a round coin’. But ‘S looks at an F’ is less strongly associated with epistemic or doxastic agent properties than appearance verbs are associated with doxastic patient-properties.²² Hence the weak competition gets sidelined by the stronger inferences (Foss and Speer 1991; Morris 1994), whose conclusion (C) goes through. Either way, doxastic inferences provide input for further cognitive processing.

Arguments from illusion are commonly presented as addressing the guiding question whether perceivers are aware of physical objects. The present input is hence processed in addressing the task of judging whether or not the viewer is aware of the physical object (the round coin). Such categorization tasks are addressed with a version of the representativeness heuristic (Kahneman and Frederick 2002; Tversky and Kahneman 1982; *cf.* Morewedge and Kahneman 2010). This heuristic has us base (probabilised) categorisation judgments (how probable is it that the ordered pair of

²¹ An anonymous reviewer questioned whether the use of appearance verbs is crucial for the argument. Brief ‘roadmaps’ of the argument (e.g., the first outline in Crane and French 2015, Sect. 2.1) indeed do without them, but start the argument with the controversial negative claim (2 above). Appearance verbs are required, however, for stating the uncontroversial case descriptions (like 1 above) that fuller statements of the argument treat as initial premise.

²² A nearest neighbours analysis provided supporting evidence. Aurélie Herbelot complemented the data from Fischer et al. (2015) (see above, Sect. 3.2) with similar analysis for ‘look at’. While its five nearest neighbours included epistemic terms ‘notice’ and ‘find’, distributional similarity was strikingly low, with a cosine of 0.14 for the nearest neighbour and no words clearly standing out in terms of distributional similarity. By contrast, appearance verbs had distributionally highly similar nearest neighbours (cosine 0.45 for 5th-nearest neighbour of ‘seem’ and ‘appear’), which clearly stood out.

viewer and round coin falls under the category ‘S is aware of X’?) on the degree of conformity with the relevant stereotype. To gauge what judgment this heuristic would deliver in the present situation, we need to determine the components of this stereotype and their relative weights.

An eye-tracking study revealed extensive similarities in intricate processing patterns for ‘aware’- and ‘see’-sentences which strongly suggest similar schemas are deployed in interpreting them (Fischer and Engelhardt 2019) with the retention strategy (see above, Sect. 3.1). We infer that ‘S is aware of X’ is associated with a variant of the *seeing* stereotype. As components, this situation schema includes epistemic agent features (*S knows what X is, S knows X is there, etc.*) in addition to non-epistemic agent and patient features (*S looks at X, X is before S, X is near S, etc.*).

Further data suggests that, while the stereotypes associated with ‘aware of’ and ‘see’ are similar in terms of the features they include, these features differ in their ‘weight’ or strength of association with the different verbs: ‘S is aware of X’ is mostly applied where S does not see, hear, or feel (etc.) X: In a random sample of 1000 ‘aware of’ sentences from the *British National Corpus*, 77% of occurrences fell into this category (Fischer et al., in prep). In these cases, knowledge is the only agent feature attributed to S, and the other features of the *seeing*-stereotype are contextually irrelevant. We infer that epistemic agent features are more strongly associated with ‘aware’ than ‘see’, and the other features less strongly. A forced-choice plausibility ranking experiment (reported in Supplementary Appendix, Section C, to the present paper) confirmed that epistemic agent features are yet more strongly associated with ‘aware’ than ‘see’—where the association is strong enough to support a prominent epistemic use (‘I see your point’), arguably interpreted with the common metaphor-interpretation strategy of stereotype-feature transfer (Bortfeld and McGlone 2001; Searle 1993). Plausibility ratings elicited in a comprehension study with eye tracking confirmed that, where contextually irrelevant and unsupported, spatial patient features (*X is before S*) get completely suppressed in interpreting ‘aware’ (though not ‘see’) sentences (Fischer and Engelhardt 2019); this suggests they are weakly associated with ‘aware’. We tentatively conclude that, in the ‘aware’ stereotype, epistemic agent features are very strongly associated with the verb, whereas the other features are weakly associated.

Application of the representativeness heuristic to the present input therefore would deliver a negative judgment: Premise (1) tells us that the object viewed is round. Integration of (C) (‘The viewer thinks that the object viewed is elliptical’) with this contextual information leads to the conclusion that the viewer has a wrong belief about the coin, and does not know that it is round, or that there is a round coin. This input suggests that conformity with the ‘aware’ stereotype is low: The agent lacks the most highly weighted component feature of the stereotype, and the other component features have such a low weighting that even conformity with all remaining features could not compensate the lack. Application of the representativeness heuristic therefore delivers the judgment that (more likely than not) the viewer is not aware of the round coin (= 2 above).

This conclusion is strengthened, rather than defeated, by the common qualifier ‘directly’: The philosophical notion of ‘direct awareness’ does not cancel epistemic implications but rather imposes the stricter requirement that the relevant knowledge

be acquired without conscious inference or other intellectual process (Price 1932, p. 3; Russell 1912, p. 4; cf. Fischer 2011, pp. 114–116).²³ Hence the ignorant viewer is not ‘directly aware’ of the round coin, either.

Further empirical evidence is required to support (H₃) and this explanation of the key inference in the argument from illusion, from (1) to (2) (see Sect. 7.5). If confirmed, however, our explanation resolves at any rate the early 20th century version of the paradox by exposing in its very first step an automatic stereotypical inference, from (1) to (C), which is contextually inappropriate and leads to a conclusion proponents of the argument explicitly reject but presuppose in further reasoning: They typically intend to use appearance- and perception-verbs in a ‘phenomenal’ sense devoid of their usual factive, epistemic, or even doxastic implications (Ayer 1956, p. 90; Jackson 1977, pp. 33–49; cf. Chisholm 1957, pp. 44–48; Maund 1986), so that the inference is not licensed by the intended sense of ‘appear’ and its cognates. Second, proponents explicitly acknowledge that, in the familiar cases at issue, viewers confidently judge that things actually have some shape, size, or colour distinct from the one they look under the circumstances (e.g., Ayer 1956, p. 88; Broad 1923, pp. 236–237, 241; cf. Price 1932, p. 27). Finally, also the inference from (C) to (2) is defective: Since ‘is aware of’ is one of the perception verbs proponents of the argument want to use in a phenomenal sense, the verb’s epistemic implications should be completely suppressed. This impugns those rare versions of the argument that proceed from unfamiliar cases of illusion, where viewers are taken in.

7.3 Paradox resolution

To explore whether our diagnostic hypothesis can meaningfully contribute to resolving the paradox, we now examine whether it can be extended to the currently most prominent version of the argument from illusion (Robinson 2001, pp. 57–58; Smith 2002, pp. 25–27; cf. Crane and French 2015; Fish 2010, pp. 12–13). Early 20th century authors leaped from case-descriptions (1 above) to negative conclusions (2 above), and based ‘phenomenal’ judgments on these (above). By contrast, more recent authors base negative conclusions (like 5 below) on ‘phenomenal’ judgments (3 below), inferred from case-descriptions with the ‘Phenomenal Principle’ (2 below). E.g.:

1. When subjects view a round coin sideways, the coin appears elliptical to them.
2. Whenever something appears a shape, size, or colour F to observers, they are (directly) aware of something that actually is F. Hence:
3. When subjects view a round coin sideways, they are (directly) aware of something that actually is elliptical (an elliptical patch).
4. If b has a property a lacks, $a \neq b$. (Leibniz’ Law)
5. When subjects view a coin sideways, they are (directly) aware of something other than the round coin (an elliptical ‘sense-datum’).

Again, the remainder of the argument generalises to all cases of visual perception. We now outline for further follow-up an empirically informed analysis which will suggest

²³ Some authors exclude inferences by admitting as objects of ‘direct awareness’ only things to which the appearance/reality-distinction does not apply (e.g. Ayer 1940, pp. 59, 61, 69), so that no inference is required to find out whether they merely appear or actually are F (cf. Broad 1923, pp. 239–240, 248).

– against first appearances to the contrary – that (H₃) applies also to this version of the argument.

The Phenomenal Principle has been either advanced to explain the phenomenal character of our perceptual experience (Broad 1923, pp. 240–241; *cf.* Smith 2002, pp. 36–37; Fish 2010, p. 6) or treated as obvious or intuitive (e.g., Price 1932, p. 63; Robinson 2001, p. 54). We now argue that thinkers only regard the principle as intuitive when they presuppose the negative conclusion (5 above) it is meant to support, so that at least where the principle is treated as obvious or intuitive, the present version of the argument continues to rely at its earliest stage on the negative conclusion (H₃) can explain.²⁴

Intuitive plausibility results from high fluency of the underlying processes (Simmons and Nelson 2006), and promotes swift acceptance of judgments (Thompson et al. 2011) and ex-post rationalisations of initial responses (Shynkaruk and Thompson 2006). Syntactic complexity reduces fluency (Lowrey 1989), and abstract or general wording reduces the effect of fluency on judgments (Tsai and Thomas 2011). This suggests that what strikes proponents of the argument as intuitive is not the general principle, in its abstract formulation (which many students find outright incomprehensible, until given concrete examples), but rather particular phenomenal judgments, phrased in syntactically simple and concrete terms. The relevant phrasing in statements of the argument is that ‘viewers are aware of an elliptical patch’ (*cf.* Price 1932, p. 3) or ‘speck’ (e.g., Ayer 1940, pp. 22–23). These statements express the intuitions to which the argument has been held to appeal (e.g., Robinson 2001, p. 54). The general principle is then formulated only in efforts to transform intuitive reasoning into a deductive argument, namely, to turn the inference of phenomenal judgments from initial case descriptions into a deductive inference, and it is accepted as ‘intuitive’ due to the intuitive plausibility of the particular phenomenal judgments it appears to justify.

The intuitive judgments that thus do all the work have the form ‘S is aware of an F patch’. ‘F patch’ has a literal use, which attributes F-ness to a patch of some sort, and a metaphorical use, which refers to something by saying it looks like an F patch (perhaps from here, now).²⁵ We employ the latter, e.g., when we cannot tell what it is we are looking at (‘Do you see the small red patch in the valley? Might that be our

²⁴ This conclusion will arguably also apply where the Phenomenal Principle (PP) is invoked on explanatory grounds. E.g., C.D. Broad (1923) invokes the PP to explain ‘why the penny should seem elliptical rather than of any other shape’. But, as Broad grants, familiar ‘laws of perspective’ explain this (p. 235); what these laws supposedly cannot explain is ‘the compatibility of these changing elliptical appearances ... with the ... constancy and roundness of the physical object’ (p. 236). This compatibility problem arises from an apparent tension between, e.g., the object’s elliptical appearance and the fact that it is round. Since people ordinarily expect round objects to look elliptical from various perspectives (Austin 1962, p. 26), the felt tension is only generated by the expectation that when something appears F there should be something that *is* F. Without such prior commitment to the PP, this specific explanatory challenge does not arise. Alternatively, authors insist that only instantiations of F can ‘adequately explain’ why our experience of an F-looking thing is as it is (Fish 2010, p. 6)—without considering scientific explanations, which take a different line (review: Clark 1996). Either way, thinkers seem committed to the PP from the start, instead of basing their acceptance of it on an inference to the best explanation, and our account below may apply.

²⁵ The common feature transfer strategy (Bortfeld and McGlone 2001; Ortony 1993; Searle 1993) has language users select one or more stereotypical implications of the dominant (literal) sense of an expression, as metaphorical interpretation. Here, we select the stereotypical looks of the literal referent.

car?’), or when we wish to avoid the stereotypical implication that the agent knows what it is she is seeing (‘She watched the small specks climbing towards her, and would have fled, had she recognised them as her pursuers.’), in line with the speaker’s maxim of the M-heuristic: ‘Use unusual (marked) expressions for stereotype-deviant situations’, where marked expressions ‘contrast with those you would use to describe the corresponding stereotypical situation’ (Levinson 2000, p. 136). There is then no suggestion that ‘F patch’ refers to something that actually is F (the small red patch may turn out to *be* a big SUV).

On a literal interpretation of the phrase, phenomenal intuitions like ‘The viewer is aware of an elliptical patch’ are controversial, and take for granted too much of what the argument needs to show. We therefore submit that any pre-theoretical acceptance of them as obvious is due to metaphorical interpretation: ‘Elliptical patch’ then refers to the round coin just mentioned before. *Thus* understood, the judgment is pre-philosophically uncontroversial. But this metaphorical interpretation does not support the key moves required by the argument from illusion: First, it does not support generalization from intuitive inferences (which lead from specific case descriptions to particular phenomenal judgments), to the Phenomenal Principle that whenever something appears F, observers are aware of something that *actually is F*. Alternatively, proponents of the argument could base inferences with Leibniz’ Law directly on intuitive phenomenal judgments. But, second, their metaphorical interpretation does not permit such inferences, either (*cf.* ‘That small red patch cannot be our car—our SUV is big’). The argument thus requires the switch to a literal interpretation.

This switch can be explained by the partial match heuristic for determining reference, which has been invoked to explain semantic illusions (Barton and Sanford 1993; Kamas et al. 1996; Park and Reder 2004): ‘Pick the domain element semantically most similar to the stimulus concept, if the similarity exceeds a threshold; otherwise, assume the expression has a referent satisfying the concept, outside the domain of discourse.’ This heuristic has us initially interpret ‘elliptical patch’ as referring to the reasonably similar sole object mentioned—the coin which then looks similar to an elliptical patch. But any further negative conclusion to the effect that the viewer is unaware of the round coin will remove this object from the relevant domain of discourse (objects of awareness). The partial match heuristic then has people posit a new object, not introduced by the premises, which satisfies the description on the default literal interpretation.

We therefore submit that the current textbook version of the argument from illusion relies on the same inference from initial case descriptions to negative conclusions that earlier versions explicitly endorsed. Only these negative conclusions effect the switch in the interpretation of phenomenal judgments that allows proponents of the argument to first regard them as intuitive or pre-philosophically uncontroversial and then rationalize intuitive inferences to phenomenal judgments with the Phenomenal Principle that supposedly licences them.

This means that implicit reliance on intuitive phenomenal judgments and Leibniz’ Law cannot explain how negative conclusions are obtained, in the first place. By contrast, the ‘textbook reasoning’ can be explained by the hypothesis (H₃) that initial case descriptions trigger contextually inappropriate stereotypical inferences to attributions of doxastic attitudes (‘The viewer thinks the coin is elliptical’) and ignorance (since

the coin is round). The moment we ask, ‘What is the viewer aware of?’, the speaker’s maxim of the M-heuristic has us respond to inferred ignorance by opting for the marked expression ‘an elliptical patch’, which signals deviation from relevant stereotypes and is often used to avoid the stereotypical implication that the viewer knows what she is viewing (above). The moment we ask, ‘Is the viewer aware of the round coin?’, the same ignorance attribution has the representativeness heuristic deliver the negative judgment that the viewer is not aware of the coin (Sect. 7.2). The automatic doxastic inference thus facilitates both phenomenal judgments and negative conclusions. And only the interpretation of those judgments in the light of these conclusions supports the Phenomenal Principle.

Contextually inappropriate doxastic inferences from appearance verbs thus seem to provide the best available explanation for the spontaneous inferences from initial case descriptions to negative conclusions (‘The viewer is not aware of the coin’) which we submit are crucially involved in both early analytic and current versions of the argument from illusion. If so, the explanation warrants the evaluative conclusion that both versions ultimately rely on contextually inappropriate stereotypical inferences. In addition, it identifies at the root of the more recent version a fallacy of equivocation: Phenomenal judgments (‘... elliptical patch’) receive a metaphorical interpretation when accepted as intuitive or obvious, but a literal interpretation in acceptance of the Phenomenal Principle that supposedly licenses them. We tentatively conclude that (H₃) can meaningfully contribute towards a resolution of this classic paradox about perception.

7.4 Problem-‘dissolution’

Arguments from illusion lead from uncontroversial premises to the conclusion that when we use our five senses, we are never (directly) aware of physical objects, but only of sense-data. Together with ‘arguments from hallucination’ for the same conclusion, they generate the ‘problem of perception’ (Crane and French 2015; Fish 2009; Smith 2002). This is the problem of reconciling the conclusion of these paradoxes, or as much of these arguments as one still accepts, with the common-sense convictions with which it appears to conflict. It thus exemplifies a recurrent structure: It is a ‘*paradox-generated reconciliation problem*’ (Fischer 2011). *Theoretical responses* try to solve such problems by showing that, properly understood, the parties to the apparent conflict are mutually consistent (Dancy 1985). *Diagnostic responses* try to resolve such problems by identifying mistakes in the underlying paradoxes. Relevant ‘mistakes’ can range from substantive theoretical presuppositions and implicit general principles that are wrong (Papineau 2009; Williams 1996) to contextually inappropriate default inferences (Austin 1962). Diagnostic responses can involve either more or less empirical argument and theoretical reflection about the topic under investigation (say, sense-perception): the assessment of implicit theories and principles will typically involve more, the examination of contextually inappropriate default inferences perhaps less.

We propose to give more precise content to the distinction between ‘solving’ and ‘dissolving’ such problems by considering to what extent responses require the acqui-

sition of new theoretical or empirical knowledge about the topic under investigation. The less such knowledge they require, the more ‘solutions’ turn into ‘dissolutions’.²⁶ While they do not require the acquisition of knowledge about the topic under investigation (say, sense-perception), they may involve the acquisition of semantic knowledge about words used in discussion about that topic (Austin 1962, p. 5) (*semantic dissolution*) or of psychological knowledge about the cognitive processes that drive reasoning about that topic (and the formulation of the paradox, in particular), as well as about the cognitive structures that support these processes (Fischer 2011, pp. 218–223; Weinberg 2017, p. 179) (*psychological dissolution*).²⁷

Both approaches can target different defects for exposure. As traditionally conceived (e.g., Hanfling 2000; cf. Hansen 2014), OLP seeks to expose semantic defects, namely, lack of meaning or truth, in philosophical questions or the assumptions or conclusions that motivate them. Alternatively, however, diagnostic responses may seek to expose epistemic defects, namely, show proponents lack justification for some of the assumptions or conclusions that engender the problem (Fischer 2011, pp. 61–72). Paradoxical arguments provide *prima facie* justification for conclusions that engender reconciliation problems. Exposing fallacies in such arguments then provides an undercutting defeater (Pollock 1986, p. 39) that undermines that *prima facie* justification. By identifying fallacies in the very first step of arguments from illusion, this paper provides proponents of these arguments with an undercutting defeater for their reasons to accept already the arguments’ initial conclusions. This contributes to showing that the paradox-generated reconciliation problem is ill-motivated. Together with a parallel diagnostic response to the argument from hallucination (Fischer and Engelhardt 2017, 2019, under review), the proposed diagnostic account may ‘dissolve’ the problem of perception: If further vindicated, it will show this problem ill-motivated, and will show this by depending on facts about verbal cognition, rather than about sense-perception.

The more glaring the fallacies are that a diagnostic response attributes to a philosophical paradox, the more urgent it is to support the diagnosis through empirical accounts that explain when and why competent thinkers commit those fallacies (Thagard and Nisbett 1983). Diagnostic responses to paradox-generated problems can receive such support from second-generation contributions to restrictionist experimental philosophy (see Sect. 1). These seek to develop epistemological profiles of automatic cognitive processes that tell us under which conditions we may (not) rely on their outputs (Weinberg 2015, 2016). By showing that the paradox is formulated under vitiating conditions (like those identified by the Saliency Bias Hypothesis) where automatic language processes that are generally reliable lead to inappropriate inferences, we can vindicate attribution of the resulting fallacies to competent thinkers—and develop psychological dissolutions of paradox-generated problems.

²⁶ This criterion is inspired by Wittgenstein ([1933] 2005), who wished to ‘completely dissolve’ some philosophical problems (p. 421) and suggested that ‘taking care of a philosophical problem is not a matter of pronouncing new truths about the subject of the investigation’ (p. 416).

²⁷ For discussion of how this approach is in line with a new ‘metaphilosophical naturalism’, while deviating from traditional ‘first-order methodological naturalism’, see Fischer and Collins (2015) and Fischer (2018a).

7.5 Limitations and future research

Our study used a plausibility-ranking task to examine (H_1) and (H_2), and provide indirect evidence of contextually inappropriate stereotypical inferences from appearance verbs. More direct evidence can be provided by online measures including reading time measurements with eye tracking (Patson and Warren 2010; Rayner 1998), and comprehension experiments with pupillometry (Kahneman 1973; Laeng et al. 2012). We used both techniques to document contextually inappropriate stereotypical inferences from perception-verbs (Fischer and Engelhardt 2017, 2019) which may drive arguments from hallucination. We plan to use these techniques to follow up the present investigation of appearance verbs, to provide direct evidence of contextually inappropriate doxastic inferences.

To provide initial empirical support for the metaphilosophical hypothesis (H_3), this paper presented evidence that the ‘aware’-stereotype has a structure which ensures that input from the documented doxastic inferences *would* lead the representativeness heuristic to yield negative judgments (like ‘The viewer is not aware of the coin’) (Sect. 7.2). Follow-up experiments will examine whether this heuristic is actually used in moving from premises of arguments from illusion to such negative conclusions. Relevant experiments include, e.g., plausibility assessments where participants assess answers to questions about cases described by the arguments’ premises (e.g., ‘The round coin appears elliptical to Joe’). Questions employ either ‘see’ or ‘aware’ (‘Does Joe see/Is Joe aware of/the round coin?’). ‘see’ is less strongly associated with epistemic and doxastic agent-properties than ‘is aware of’, and more strongly with the other components of its associated stereotype. If participants employ the representativeness heuristic to answer the question, doxastic inferences from appearance verbs should affect answers to ‘aware’-questions more strongly, and negative answers should be deemed more plausible in response to ‘aware’-questions than ‘see’-counterparts.

This paper has examined one source of the fallacies we identified in the classical paradoxes we considered: Automatic inferences with the I-heuristic and the representativeness heuristic lead to conclusions (e.g., ‘the viewer is unaware of the coin’) which appear to clash with background beliefs and contextual inferences (e.g., ‘the viewer is aware of something’). But perceived conflicts lead to lower subjective confidence and plausibility (De Neys et al. 2011) and increased critical scrutiny (Thompson et al. 2011). Arguably, the inferences at the root of arguments from illusion (and hallucination) only strike their proponents as so intuitively plausible because they believe from the outset in the existence of a complementary perceptual space, ‘the mind’, in which objects of awareness can be placed when evicted from the viewer’s physical environment. Accordingly, one of us has developed a debunking explanation of introspective conceptions of the mind that have traditionally struck proponents of these arguments as intuitively plausible (Fischer 2014b, 2018b). How this conception and contextually inappropriate stereotypical inferences interact to generate these paradoxes and the ‘problem of perception’ remains to be examined.

Further profitable applications may include examination of inferences from the verb ‘to know’: Experimental philosophers have started to collect data relevant for assessing the salience of its different senses or uses (Hansen et al. [under review](#)).

Experimental and ordinary language philosophers have clarified philosophically relevant uses, including infallibilist uses (Nichols and Pinillos 2018) and uses according varying relevance to relevant alternatives (Baz 2017), suggesting interesting hypotheses about how skeptical paradoxes and other epistemological problems may arise from contextually inappropriate inferences from the verb. We would welcome application of the approach presented, to these and further problems.

8 Conclusion

This paper provides critical ordinary language philosophy with fresh, empirical foundations. Critical OLP examined default inferences from words, which have subtle contextual defeaters. It sought to ‘dissolve’ philosophical problems by disentangling such inferences. Our Saliency Bias Hypothesis identifies a first set of conditions under which even competent speakers make contextually inappropriate stereotypical inferences: Such inferences occur when speakers give a word with a clearly dominant sense rarefied uses for whose interpretation the dominant sense is functional. This is liable to happen when philosophers give special uses to words that already have well-established uses in ordinary discourse. Our psycholinguistic hypothesis thus lends empirical substance to Austin’s observation that ‘tampering with words ... is always liable to have unforeseen repercussions... we must always be particularly wary of the philosophical habit of dismissing some (if not all) the ordinary uses of a word as “unimportant”’ (Austin 1962, p. 63). We must be wary because ordinary uses may be most salient and continue to shape automatic inferences; and ‘unforeseen repercussions’ include inappropriate inferences from rarefied uses (say, technical uses resulting from well-motivated philosophical ‘tampering’), which go through especially in uninformative contexts (typical of philosophical arguments). The Saliency Bias Hypothesis thus provides an empirical rationale for critical OLP.

Three cross-linguistic experiments supported the psycholinguistic hypothesis by providing evidence for contextually inappropriate doxastic inferences from phenomenal uses of appearance verbs, and their robustness in the face of competing pragmatic inferences. We empirically developed the metaphilosophical hypothesis that the documented inferences are at the root of classical paradoxes about perception (‘arguments from illusion’). Philosophical problems arising from such paradoxes can be resolved by identifying the inappropriate inferences involved. Where inferences remain tacit, or their attribution would otherwise violate principles of charity, experimental evidence is required. Psycholinguistic experiments can provide such evidence. Psycholinguistic methods and findings thus motivate and support a more widely applicable ‘critical’ approach in experimental ordinary language philosophy that seeks to ‘dissolve’ paradox-generated problems by disentangling context-sensitive inferences language users automatically make from words.²⁸

²⁸ For helpful comments on previous drafts, the authors thank two anonymous reviewers as well as James Andow, Avner Baz, Chi-He Elder, Rachel Giora, Nat Hansen, Jennifer Nado, and conference audiences in Turku (Finland, May 2017) and Osnabrück (Germany, November 2017). Linguist Akiko Tomatsuri kindly provided advice on the development of Japanese materials. For help with gathering and entering English, German, and Japanese data, respectively, we thank Oliver Afrijdijanta, Karolin Meinert, and Junichiro

Compliance with ethical standards

Ethical standards The research conformed to the ethical standards for conducting research as outlined by the British Psychological Society.

Human and animal rights The use of human research participants was approved by the relevant Research Ethics Committee of the University of East Anglia.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adler, J. E. (1994). Fallacies and alternative interpretations. *Australasian Journal of Philosophy*, *72*, 271–282.
- Alexander, J. (2012). *Experimental philosophy: An introduction*. Cambridge: Polity.
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*, 219–235.
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Psychological Optics*, *34*, 502–508.
- Austin, J. L. (1957). A plea for excuses. *Proceedings of the Aristotelian Society*, *57*, 1–30.
- Austin, J. L. (1962). *Sense and sensibilia*. Oxford: Oxford University Press.
- Austin, J. L. (1975). *Sinn und Sinneserfahrung* (E. Cassirer, Trans.). Stuttgart: Reclam.
- Austin, J. L. (1984). *Chikaku no Gengo: Sense to Sensibilia* (N. Tanji & S. Moriya, Trans.). Tokyo: Keiso Shobo.
- Ayer, A. J. (1940). *Foundations of empirical knowledge*. London: Macmillan.
- Ayer, A. J. (1956/1990). *The problem of knowledge*. London: Penguin.
- Ayer, A. J. (1981). *Chishiki no Tetsugaku* (K. Kamino, Trans.). Tokyo: Hakusuisha.
- Ayer, A. J. (1991). *Keikenteki Chishiki no Kiso* (K. Kamino, T. Nakasai, & T. Nakatani, Trans.). Tokyo: Keiso Shobo.
- Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: mediated by priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 336–345.
- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory and Cognition*, *21*, 477–487.
- Baz, A. (2012). *When words are called for: A defense of ordinary language philosophy*. Cambridge, MA: Harvard University Press.
- Baz, A. (2016). Ordinary language philosophy. In H. Cappelen, T. Szabo Gendler, & J. Hawthorne (Eds.), *Oxford handbook of philosophical methodology* (pp. 112–129). Oxford: OUP.
- Baz, A. (2017). *The crisis of method*. 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.
- Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, *51*, 495–522.
- Bortfeld, H., & McGlone, M. S. (2001). The continuum of metaphor processing. *Metaphor and Symbol*, *16*, 75–86.
- Broad, C. D. (1923). *Scientific thought*. Repr. 2000. London: Routledge.
- Brogaard, B. (2013). It's not what it seems: a semantic account of 'seems' and seemings. *Inquiry*, *56*, 210–239.

Footnote 28 continued

Wada. Joachim Horvath's work on this paper was supported by an Emmy Noether grant of the *Deutsche Forschungsgemeinschaft* (project number 391304769).

- Brogaard, B. (2014). The phenomenal use of 'look' and perceptual representation. *Philosophy Compass*, 9(7), 455–468.
- 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.
- Cavell, S. (1994). *A pitch of philosophy: Autobiographical exercises*. Cambridge, MA: Harvard University Press.
- Chisholm, R. (1957). *Perceiving*. Ithaca: Cornell UP.
- Chow, W., Smith, C., Lau, E., & Phillips, C. (2016). A 'bag-of-arguments' mechanism for initial verb predictions. *Language, Cognition, and Neuroscience*, 31, 577–596.
- Clark, A. (1996). *Sensory qualities*. Oxford: OUP.
- Cova, F. et al. (2018). *Estimating the reproducibility of experimental philosophy*. <https://doi.org/10.17605/OSF.IO/SXD4H>.
- Crane, T., & French, C. (2015). The problem of perception. In N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Summer 2015. <http://plato.stanford.edu/entries/perception-problem/>.
- Dancy, J. (1985). *Introduction to contemporary epistemology*. Oxford: Blackwell.
- De Neys, W. (2006). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology*, 59, 1070–1100.
- De Neys, W., Schaeken, W., & D'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory and Cognition*, 31, 581–595.
- De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0015954>.
- Deutsch, M. (2015). *The myth of the intuitive*. Cambridge, MA: MIT Press.
- Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27, 315–330.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognition*, 33, 547–582.
- Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, 6, 635–653.
- Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120, 339–357.
- Faust, M. E., & Gernsbacher, M. A. (1996). Cerebral mechanisms for suppression of inappropriate information during sentence comprehension. *Brain and Language*, 53, 234–259.
- Fein, O., Yeari, M., & Giora, R. (2015). On the priority of salience-based interpretations: The case of sarcastic irony. *Intercultural Pragmatics*, 12, 1–32.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40, 532–538.
- Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 182–196.
- 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. (2011). *Philosophical delusion and its therapy. Outline of a philosophical revolution*. New York: Routledge.
- Fischer, E. (2014a). Verbal fallacies and philosophical intuitions: The continuing relevance of ordinary language analysis. In B. Garvey (Ed.), *J. L. Austin on language* (pp. 124–140). Basingstoke: Palgrave Macmillan.
- Fischer, E. (2014b). Philosophical intuitions, heuristics, and metaphors. *Synthese*, 191, 569–606.
- Fischer, E. (2018a). Wittgensteinian 'therapy', experimental philosophy, and metaphysical naturalism. In K. Cahill & T. Raleigh (Eds.), *Wittgenstein and naturalism* (pp. 260–286). New York: Routledge.
- Fischer, E. (2018b). Two analogy strategies: The cases of mind metaphors and introspection. *Connection Science*, 30, 211–243.
- Fischer, E., & Collins, J. (2015). Rationalism and naturalism in the age of experimental philosophy. In E. Fischer & J. Collins (Eds.), *Experimental philosophy, rationalism and naturalism* (pp. 3–33). London: Routledge.
- 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). London: Bloomsbury.
- Fischer, E., & Engelhardt, P. E. (under review). *Lingering stereotypes: Salience bias in philosophical arguments*.
- Fischer, E., Engelhardt, P. E., & Herbelot, A. (2015). Intuitions and illusions: From experiment and explanation to assessment. In E. Fischer & J. Collins (Eds.), *Experimental philosophy, rationalism and naturalism* (pp. 259–292). London: Routledge.
- Fischer, E., Engelhardt, P. E., & Herbelot, A. (in prep). *The expertise objection to experimental philosophy: A psycholinguistic perspective*.
- Fish, W. (2009). *Perception, hallucination, and illusion*. Oxford: Oxford University Press.
- Fish, W. (2010). *Philosophy of perception*. London: Routledge.
- Flickinger, D., Oepen, S., & Ytrestol, G. (2010). Wikiwoods: Syntacto-semantic annotation for the English Wikipedia. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner and D. Tapias (Eds.), *Proceedings of the seventh international conference on language resources and evaluation (LREC2010)* (pp. 1665–1671). Paris: European Language Resources Association.
- Fodor, J. (1983). *The modularity of mind*. Cambridge (MA): MIT Press.
- Foss, D. J., & Speer, S. R. (1991). Global and local context effect in sentence processing. In R. R. Hoffman & D. S. Palermo (Eds.), *Cognitive and the symbolic processes: Applied and ecological perspectives* (pp. 115–139). Hillsdale, NJ: Erlbaum.
- Garrod, S., & Sanford, A. J. (1981). Bridging inferences and the extended domain of reference. In J. Long & A. Baddley (Eds.), *Attention and performance IX* (pp. 331–346). Hillsdale, NJ: Erlbaum.
- Garvey, B. (Ed.). (2014). *J.L. Austin on Language*. Basingstoke: Palgrave.
- Genka, T. (2017). *Chikaku to Handan no Kyoukaisen [The distinction between perception and cognition]*. Tokyo: Keio University Press.
- Gerken, M., & Beebe, J. (2016). Knowledge in and out of contrast. *Nous*, 50, 133–164.
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 8, 183–206.
- Giora, R. (2003). *On our mind. Salience, context, and figurative language*. Oxford: OUP.
- Giora, R., & Fein, O. (1999). On understanding familiar and less-familiar figurative language. *Journal of Pragmatics*, 31, 1601–1618.
- Giora, R., Fein, O., Aschkenazi, K., & Alkabetz-Zlozover, I. (2007a). Negation in context: A functional approach to suppression. *Discourse Processes*, 43, 153–172.
- Giora, R., Fein, O., Laadan, D., Wolfson, J., Zeituny, M., Kidron, R., et al. (2007b). Expecting irony: Context versus salience-based effects. *Metaphor and Symbol*, 22, 119–146.
- Giora, R., Raphaely, M., Fein, O., & Livnat, E. (2014). Resonating with contextually inappropriate interpretations: The case of irony. *Cognitive Linguistics*, 25, 443–455.
- Givoni, S., Giora, R., & Bergerbest, D. (2013). How speakers alert addressees to multiple meanings. *Journal of Pragmatics*, 48, 29–40.
- Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7, 219–224.
- Goryo, K. (1987). *Yomu toiu koto. (on reading)*. Tokyo: University of Tokyo Press.
- Grice, H. P. (1961). The causal theory of perception. *Proceedings of the Aristotelian Society*, 35, 121–152.
- Grice, H. P. (1989). Logic and conversation. In His: *Studies in the ways of words* (pp. 22–40). Cambridge, MA: Harvard UP.
- Grindrod, J., Andow, J., & Hansen, N. (2018). Third-person knowledge ascriptions: A crucial experiment for contextualism. *Mind and Language*. <https://doi.org/10.1111/mila.12196>.
- Gustafsson, M., & Sørli, R. (2011). *The philosophy of J.L. Austin*. Oxford: OUP.
- Hampton, J. A., & Passanisi, A. (2016). When intensions do not map onto extensions: Individual differences in conceptualization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 505–523.
- Hanfling, O. (2000). *Philosophy and ordinary language*. London: Routledge.
- Hansen, N. (2014). Contemporary ordinary language philosophy. *Philosophy Compass*, 9, 556–569.
- Hansen, N. (2017). Must we measure what we mean? *Inquiry*, 60, 785–815.

- Hansen, N. (2018). "Nobody would really talk that way!": The critical project in contemporary ordinary language philosophy. *Synthese*. <https://doi.org/10.1007/s11229-018-1812-x>.
- Hansen, N., & Chemla, E. (2013). Experimenting on contextualism. *Mind and Language*, 28, 286–321.
- Hansen, N., & Chemla, E. (2015). Linguistic experiments and ordinary language philosophy. *Ratio*, 28, 422–445.
- Hansen, N., Porter, J. D., & Francis, K. (under review). *A corpus study of "know": On the verification of philosophers' frequency claims about language*.
- Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, 111, 151–167.
- Harmon-Vukić, M., Guéraud, S., Lassonde, K. A., & O'Brien, E. J. (2009). The activation and instantiation of instrumental inferences. *Discourse Processes*, 46, 467–490.
- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: CUP.
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: OUP.
- Horne, Z., & Livengood, J. (2017). Ordering effects, updating effects, and the spectre of global scepticism. *Synthese*, 194, 1189–1218.
- Horvath, J. (2010). How (not) to react to experimental philosophy. *Philosophical Psychology*, 23, 447–480.
- Hume, D. (1739/1975). A treatise of human nature (revised by P.H. Nidditch). In L. A. Selby-Bigge (Ed.), 2nd edn. Oxford: Clarendon Press.
- Institute for Japanese Language and Linguistics and Lago Institute of Language. (2012). *NINJAL-LWP for BCCWJNational* (<http://nlb.ninjal.ac.jp/>).
- Jackson, F. (1977). *Perception. A representative theory*. Cambridge: CUP.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D. (2011). *Thinking fast and slow*. London: Allen and Lane.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, et al. (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge: CUP.
- Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory and Cognition*, 24, 687–699.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and co-reference revisited. *Journal of Semantics*, 25, 1–44.
- Kim, A. E., Oines, L. D., & Sikos, L. (2016). Prediction during sentence comprehension is more than a sum of lexical associations: The role of event knowledge. *Language, Cognition, and Neuroscience*, 31, 597–601.
- Kim, A. E., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from anticipatory eye-movements. *Journal of Memory and Language*, 52, 205–225.
- Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, 45, 259–282.
- Kutas, M., & Federmeier, K. T. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 460–463.
- Kutas, M., & Federmeier, K. T. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, 7, 18–27.
- Laugier, S. (2013). *Why we need ordinary language philosophy*. Chicago: University of Chicago Press.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.
- Levinson, S. C. (2000). *Presumptive meanings. The theory of generalized conversational implicature*. Cambridge: MIT Press.
- Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, 6, 299–305.
- Lewinski, M. (2012). The paradox of charity. *Informal Logic*, 32, 403–439.
- Loftus, E. F. (1973). Activation of semantic memory. *The American Journal of Psychology*, 86, 331–337.
- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, 7, 618–630.

- Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior*, 23, 709–733.
- 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.
- Masuoka, T. (2002). Handan no modality [Modality of judgment]. *Nihongogaku*, 21(2), 6–16.
- Matsui, T. (2000). Linguistic encoding of the guarantee of relevance: Japanese sentence-final particle YO. In G. Anderson & T. Fretheim (Eds.), *Pragmatic Markers and propositional attitude* (pp. 145–172). Amsterdam: John Benjamins.
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 37, 913–934.
- Maud, J. B. (1986). The phenomenal and other uses of ‘looks’. *Australasian Journal of Philosophy*, 64, 170–180.
- McKoon, G., & Ratcliff, R. (1980). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior*, 19, 369–386.
- 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.
- Melinger, A., & Mauner, G. (1999). When are implicit agents encoded? Evidence from cross-modal priming. *Brain and Language*, 68, 185–191.
- 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.
- Minamide, K., & Nakamura, M. (Eds.). (2011). *Genius Japanese–English dictionary* (3rd ed.). Tokyo: Taishukan.
- Moore, G. E. (1918/19). Some judgments of perception. *Proceedings of the Aristotelian Society*, 19, 1–29.
- Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Science*, 14, 435–440.
- Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 20, 92–103.
- Murphy, T. (2014). Experimental philosophy 1935–1965. In J. Knobe, T. Lombrozo, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 1, pp. 325–368). Oxford: OUP.
- Nado, J. (2016). Experimental philosophy 2.0. *Thought*, 5, 159–168.
- Naess, A. (1956). Logical equivalence, intentional isomorphism and synonymy as studied by questionnaires. *Synthese*, 10, 471–479.
- Naess, A. (1961). A study of “or”. *Synthese*, 13, 49–60.
- Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology*, 15, 1044–1045.
- Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering* (pp. 69–93). Washington, DC: APA.
- Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, 41, 663–685.
- Nichols, S., & Pinillos, A. (2018). Skepticism and the acquisition of ‘knowledge’. *Mind and Language*. <https://doi.org/10.1111/mila.12179>.
- Nobuhara, Y. (1999). *Kokoro no Gendai Tetsugaku [Contemporary philosophy of mind]*. Tokyo: Keiso Shobo.
- O’Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 18, 777–784.

- Oden, G. C., & Spira, J. L. (1983). Influence of context on the activation and selection of ambiguous word senses. *Quarterly Journal of Experimental Psychology*, 35A, 51–64.
- Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilised irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology*, 20, 139–156.
- Ortony, A. (1993). The role of similarity in similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 342–356). Cambridge: CUP.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25, 273–285.
- Papineau, D. (2009). The poverty of analysis. *Aristotelian Society Supplementary Volumes*, 83, 1–30.
- Park, H., & Reder, L. M. (2004). Moses illusion. In R. Pohl (Ed.), *Cognitive illusions* (pp. 275–291). New York: Psychology Press.
- Patson, N. D., & Warren, T. (2010). Eye movements to plausibility violations. *Quarterly Journal of Experimental Psychology*, 63, 1516–1532.
- Peleg, O., & Giora, R. (2011). Salient meanings: The whens and wheres. In K. M. Jaszczolt & K. Allan (Eds.), *Salience and defaults in utterance processing* (pp. 32–52). Berlin: de Gruyter.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36, 329–347.
- Pollock, J. L. (1986). *Contemporary theories of knowledge*. Totowa, NJ: Rowman and Littlefield.
- Postal, P. (1973). *On raising*. Cambridge, MA: MIT Press.
- Price, H. H. (1932). *Perception*. 2nd edn., repr. 1961. London: Methuen.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Robinson, H. (2001). *Perception*. London: Routledge.
- Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57, 221–237.
- Rumelhart, D. E. (1978). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum.
- Russell, B. (1912/1980). *The problems of philosophy*. Oxford: OUP.
- Russell, B. (1964). *Tetsugaku Nyuumon in 1912* (Russell, Ed., H. Nakamura, Trans.). Tokyo: Shakaishisousha.
- Russell, B. (1965). *Tetsugaku Nyuumon in 1912* (Russell, Ed., K. Ikumatsu, Trans.). Tokyo: Kadokawa Shoten.
- Russell, B. (1967). *Probleme der Philosophie in 1912* (Russell, Ed., E. Bubser, Trans.). Frankfurt a.M.: Suhrkamp.
- Russell, B. (2005). *Tetsugaku Nyuumon in 1912* (Russell, Ed., N. Takamura, Trans.). Tokyo: Chikuma Shobo.
- Sasanuma, S. (1980). Acquired dyslexia in Japanese: Clinical features and underlying mechanisms. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 48–90). London: Routledge.
- Schaffer, J., & Knobe, J. (2012). Contrastive knowledge surveyed. *Noûs*, 46, 675–708.
- Scholze-Stubenrecht, W., & Sykes, J. B. (Eds.). (1999). *Duden-Oxford, Großwörterbuch Englisch* (2nd ed.). Mannheim: Dudenverlag.
- Searle, J. (1993). Metaphor. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 83–111). Cambridge: CUP.
- Searle, J. (2001). J.L. Austin. In A. Martinich & D. Sosa (Eds.), *A companion to analytic philosophy* (pp. 218–230). Oxford: Blackwell.
- Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18, 41–55.
- Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory and Cognition*, 34, 619–632.
- Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and non-intuitive alternatives. *Journal of Experimental Psychology: General*, 135, 409–428.
- Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 28–39.
- Smith, A. D. (2002). *The problem of perception*. Cambridge, MA: Harvard UP.
- Springer, O. (Ed.). (2000). *Langenscheidts Enzyklopädisches Wörterbuch der englischen und deutschen Sprache: "Der Große Muret-Sanders"* (Vol. 2). Berlin: Langenscheidt.

- Staudacher, A. (2011). *Das Problem der Wahrnehmung*. Paderborn: Mentis.
- Stephens, G. J., Silber, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*, 14425–14430.
- Stich, S., & Tobia, K. (2016). Experimental philosophy and the philosophical tradition. In J. Sytsma & W. Buckwalter (Eds.), *Blackwell companion to experimental philosophy* (pp. 5–21). Malden: Wiley Blackwell.
- Sturt, P. (2003). A new look at the syntax-discourse interface: The use of binding principles in sentence processing. *Journal of Psycholinguistic Research*, *32*, 125–139.
- Sytsma, J., & Livengood, J. (2016). *The theory and practice of experimental philosophy*. Peterborough: Broadview.
- Tanenhaus, M. K., & Carlson, G. N. (1989). Lexical structure and language comprehension. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 529–561). Cambridge, MA: MIT Press.
- Tanenhaus, M. K., Carlson, G. N., & Trueswell, J. T. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes*, *4*, 211–234.
- Thagard, P., & Nisbett, R. E. (1983). Rationality and charity. *Philosophy of Science*, *50*, 250–267.
- Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*, 107–140.
- Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Journal of Verbal Learning and Verbal Behaviour*, *16*, 283–298.
- Traxler, M. J., Foss, D. J., Seely, R. E., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, *29*, 581–595.
- Tsai, C. I., & Thomas, M. (2011). When does feeling of fluency matter? How abstract and concrete thinking influence fluency effects. *Psychological Science*, *22*, 348–354.
- Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Urmson, J. O. (1969). A symposium on Austin's method. Part I. In K. T. Fann (Ed.), *Symposium on J.L. Austin* (pp. 76–86). London: Routledge.
- Watanabe, T., Skrzypczak, E. R., & Snowden, P. (Eds.). (2003). *Kenkyusha's new Japanese-English dictionary* (5th ed.). Tokyo: Kenkyusha.
- 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). London: Routledge.
- Weinberg, J. (2016). Intuitions. In H. Cappelen, T. Szabo Gendler, & J. Hawthorne (Eds.), *Oxford handbook of philosophical methodology* (pp. 287–308). Oxford: OUP.
- Weinberg, J. M. (2017). What is negative experimental philosophy good for? In G. D'Oro & S. Overgaard (Eds.), *The Cambridge companion to philosophical methodology* (pp. 161–183). Cambridge: CUP.
- Welke, T., Raisig, S., Nowack, K., Schaadt, G., Hagedorf, H., & van der Meer, E. (2015). Semantic priming of progression features in events. *Journal of Psycholinguistic Research*, *44*, 201–214.
- Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*, 311–334.
- Wiesing, L. (2002). *Philosophie der Wahrnehmung*. Suhrkamp: Frankfurt a.M.
- Williams, M. (1996). *Unnatural doubts*. Princeton, NJ.: Princeton University Press.
- Williams, B. (2014). *Essays and reviews: 1959–2002*. Princeton, NJ: Princeton University Press.
- Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.
- Wilson, T. D. (2002). *Strangers to ourselves*. Cambridge, MA: Harvard University Press.
- Wittgenstein (1933/2005). *The big typescript TS 213* (Ed., C.G. Luckardt & M.A.E. Aue, Trans.). Oxford: Blackwell.