Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

12 Intuitions and illusions

From explanation and experiment to assessment

Eugen Fischer, Paul E. Engelhardt, and Aurélie Herbelot

The perhaps most ambitious strand of current experimental philosophy—variously characterized as 'the sources project' (Pust 2012) or the research programme of 'cognitive epistemology' (Fischer 2014)—seeks to develop experimentally testable explanations of philosophically relevant intuitions that facilitate the assessment, first, of their evidentiary value and, second, of our warrant for accepting them.[1] While prominently advocated (e.g., Knobe and Nichols 2008; 8), this programme is as yet actually pursued only by a handful or two of philosophers (including Fiala *et al.* 2011; Fischer 2014, 2015; Nagel 2010, 2011, 2012; Nahmias and Murray 2010; Nichols and Knobe 2007). Some contributions pursue it with the aim of showing that we may trust intuitions that are commonly adduced as evidence for philosophical theories (see critical review by Kornblith 2015 [this volume, Chapter 6]). Others (including the present chapter) target paradoxical or conflicting intuitions that are relevant as a source of philosophical problems and seek to resolve these problems by showing for at least some of the underlying intuitions that we lack the right to accept them. That is, the immediate aim of these efforts is to provide validating or debunking explanations of specific intuitions. The ultimate aim is to develop 'epistemic profiles' of the underlying cognitive processes which tell us under what circumstances we may trust their deliverances and when and where we ought to beware (Weinberg 2015 [this volume, Chapter 7]).

Extant contributions to cognitive epistemology have pursued both these aims by developing '*GRECI explanations*', as we might call them, which trace specific intuitions back to cognitive processes that are *generally reliable* but predictably engender *cognitive illusions*, under specific circumstances (Fischer 2014, 2015; Nagel 2010, 2012). To do so, these contributions have built on research from cognitive and social psychology. In targeting apparently paradoxical intuitions from the experimentally still almost virgin territory of philosophy of perception (Section 1, pp. 260–265), the present chapter will break new ground by using instead, concepts, findings, and experimental paradigms from computational linguistics and psycholinguistics, namely, to develop (Section 2, pp. 265–274) and experimentally

[1] These efforts also belong to the 'submarine part' of 'iceberg epistemology' (Henderson and Horgan 2011), which examines epistemologically relevant psychological factors below the waterline of conscious awareness.

test (Section 3, pp. 274–280) a fresh GRECI explanation. This fresh explanation traces intuitions back to a cognitive process that is generally involved in verb comprehension and potentially relevant in *all* areas of philosophy. On this fresh basis, we will address a, perhaps *the*, key methodological challenge to experimental philosophy which is clearly recognized by its practitioners (e.g., Knobe and Nichols 2008, 7–8): the challenge of developing strategies that allow us to move from experimental findings to epistemological assessments, without naturalistic fallacy (see Fischer and Collins 2015 [this volume, Introduction, Section 5, pp. 23–27]). To tackle this challenge, we will explore two strategies that allow us to use the explanation proposed, to assess the evidentiary value of (i) the spontaneity of the intuitions explained and (ii) the subjective confidence they inspire (Section 4, pp. 280–286).

1 An intuitive paradox

Many characteristically philosophical problems are engendered by philosophical paradoxes. Persuasive arguments that lead to a conclusion *q* which is apparently at odds with some common-sense conviction *p* motivate several philosophical questions of the form ‘How is it possible that *p* (given that *q*)?’ (Fischer 2011; see also Papineau 2015 [this volume, Chapter 1]). We can try to resolve such problems by resolving the underlying paradoxes: by showing that we lack warrant either for accepting the conclusion that *q* or for assuming that it is at odds with *p*. Where the arguments for *q* rely on intuitive premises that are not themselves supported by evidence or (non-circular) argument, cognitive epistemology becomes potentially relevant: Such arguments appear to bring out a conflict between these intuitions and common-sense convictions. We are justified in accepting claims at odds with common-sense convictions only if we have positive reasons in their support.[2] Where a thinker accepts those intuitive premises in the absence of evidence or (non-circular) argument, the only positive reason she is in a position to adduce is the mere fact that she or others have these intuitions; her warrant for accepting the intuitions will depend upon their *evidentiary value*: upon whether the mere fact that thinkers have these intuitions, as and when they do, speaks for the intuitions’ truth. GRECI explanations that let us determine the evidentiary value of particular intuitions—and show they have none—thus allow us to resolve intuitive paradoxes and the philosophical problems they engender.

A case in point is what is now (since Smith 2002) known simply as ‘the [*sic*] problem of perception’. This classical problem has been a linchpin of Western philosophy of perception, from the mid-eighteenth to the mid-twentieth century, and has again become of a focus of debate (Brewer 2011; Crane 2011; Fish 2009; Robinson 2001; Smith 2002). The problem is raised by a number of related but

[2] This is the comparatively uncontroversial ‘first half’ of the default-and-challenge model of justification (e.g., Williams 2001, 25): The present argument does not rely on the model’s further—and more controversial—claim that endorsement by common sense makes acceptance the *appropriate* default response to a claim, in the absence of positive reasons for doubt.

distinct arguments—known as 'arguments from illusion', 'from hallucination', etc.—which lead from different premises to the apparently paradoxical conclusion that when we use our five senses we are never aware, or at any rate never directly aware, of physical objects or public events, but only of private ideas, perceptions, or sense-data. Its apparent clash with the common-sense conviction that we see tables and books, hear concerts and explosions, and smell burnt milk motivates the question of how any of these perceptual achievements is as much as possible: How is it possible for us to perceive (or correctly say that we perceive) physical objects and public events (given that all we are aware of, in perception, are private perceptions)? (See, e.g., Ayer 1940.)[3]

The historically most influential of the philosophical paradoxes that give rise to this problem are *arguments from illusion* which were at their most influential within analytic philosophy and in the first half of the twentieth century (Ayer 1940, 1956; Broad 1923; Moore 1918–19; Price 1932; Russell 1912; see also Martin 2003). They proceed from premises that set out mostly familiar cases of non-veridical perception (misleadingly called 'illusions') where physical objects look or otherwise appear to have a size, shape, colour, or other property they do not actually possess. In their seminal statement, Hume spontaneously leaps from the description of such a case (below: 'The table … no alteration') to the conclusion that we are aware of an 'image', rather than the physical object:

> The table, which we see, seems to diminish, as we remove farther from it: but the real table, which exists independently of us, suffers no alteration: it was, therefore, but the image, which was present to the mind. These are obvious dictates of reason.
>
> (Hume 1748/1975, 152)

Current textbook reconstructions of the argument break this decisive 'sense-datum inference' (Smith 2002, 25) up into two steps, and derive its conclusion with the so-called 'Phenomenal Principle' (2, below) and Leibniz's Law (4, below) (e.g., Robinson 2001, 57–58):

(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 (e.g., whenever something looks elliptical to them), they are (directly) aware of something that actually has that shape, size, or colour (e.g., they are aware of an elliptical patch of colour). Hence:

(3) When subjects view a round coin sideways, they are (directly) aware of something that actually is elliptical.

(4) If b has a property that a lacks, $a \neq b$. Hence:

[3] The distinctively philosophical problem is, more specifically, that of reconciling that common-sense conviction with as much of the paradoxical argument as the querist accepts, in the light of the apparent clash. Thinkers who accept only the premises wonder, e.g., how sense-perception is possible, given that there are cases of non-veridical perception or 'illusion' (e.g., Smith 2002; Crane 2011).

(5) When subjects view sideways a coin that actually is round, they are (directly) aware not of the coin but of something else (an image, perception, or sense-datum).

Different versions of the argument then generalize in different ways from this negative conclusion about one kind of case to all cases of perception, non-veridical and veridical alike.

The persuasiveness of this argument, as of directly and indirectly related arguments (arguments from hallucination and sceptical arguments like the argument from dreaming), arguably depends to a large extent on the appeal of intuitive conceptions of the mind as an inner realm of perception (see Fischer 2011, 2014, and 2015 for a first explanation). In this chapter, however, we will consider intuitions that are specific to the argument from illusion and inquire whether they can serve to justify its conclusion.

Commentators unhesitatingly characterize the argument as resting on an ‘appeal to intuition’ (e.g., Robinson 2001, 54). We will explore whether this is true also in the strict aetiological sense of the word dominant in cognitive psychology and required for the purposes of cognitive epistemology. ‘Intuitions’, in this sense, are judgments generated by largely automatic cognitive processes. As commonly defined, cognitive processes are *automatic*, rather than ‘controlled’ to the extent to which they are effortless, unconscious, non-intentional, and autonomous (Bargh 1994; Dijksterhuis 2010). These properties have operational definitions: A process possesses, e.g., the key property of *effortlessness* to the extent to which it is independent of working memory and thus requires no attention or other limited cognitive resource, so that performance is not impaired by multitasking (simultaneously keeping in mind long numbers, etc.).[4] Automatic cognitive processes generate also, e.g., perceptual and memory judgments. We therefore define:

Intuitions are judgments which are

(1) based, more specifically, on ‘*automatic inferences*’ (Kahneman and Frederick 2005, 268; see also Evans 2010, 314), i.e., on largely automatic cognitive processes which duplicate inferences governed by normative or heuristic rules,[5]

(2) and accompanied by high levels of *subjective confidence* (Thompson *et al.* 2011), regardless of whether or not the thinker endorses them upon reflection.[6]

[4] Further, a process is *unconscious* to the extent to which the subject is unable to report its course as opposed to express its outcome (judgment, decision, etc.), *non-intentional* to the extent to which its initiation is insensitive to the aims or goals the subject pursues, and *autonomous* to the extent to which the subject is unable to end the process or alter its course, once initiated.

[5] Whereas *normative rules* (logic, probability theory) determine or constrain what is right or reasonable, *heuristics* are rules of thumb which yield reasonably accurate judgments in most relevant contexts, without determining or constraining what is to count as correct. A process ‘duplicates’ an inference where it leads from the same inputs to the same judgments.

[6] In particular when they are not so endorsed, it is tempting to characterize them as ‘inclinations to assent’ (see Sosa 2007 and Earlenbaugh and Molyneux 2009). But high levels of subjective confidence tend to pre-empt (further) reflection (Thompson *et al.* 2011).

These two defining features are empirically linked: Effortlessness serves as a metacognitive cue (Alter and Oppenheimer 2009) for spontaneous assessments of plausibility (Kelley and Lindsay 1993) and confidence (Thompson *et al.* 2011). Judgments delivered by effortless automatic processes therefore tend to inspire confidence (see below, Section 4.3, pp. 285–286).

Whether a judgment is an intuition in the aetiological sense explained can be established only by a successful psychological explanation that traces it back to automatic cognitive processes which duplicate rule-governed inferences. But we can plausibly start to look for such an explanation where three diagnostic conditions are met:

(1) Thinkers spontaneously leap to non-perceptual judgments about verbally described cases. In particular where the cases described are highly unusual or recherché, and when the description prompts a judgment at odds with common-sense judgments or generally acknowledged facts, the judgment cannot be based on mere remembering of familiar facts about cases of that kind. If not random but recurrent and shared, the judgment then is probably due to an inference from premises contained in the vignette, or a process duplicating such an inference.

(2) These judgments strike the thinkers as highly plausible (evidence of effortlessness).

(3) Thinkers cannot explain how they arrived at these judgments or how one might justify the judgments with rules they master (evidence of the unconsciousness of the underlying cognitive process and of the rule-governed inferences duplicated).

'*Phenomenal judgments*' like (3) above, which commentators (e.g., Robinson 2001, 54) have regarded as intuitions all along, fit this bill. (i) These judgments are at variance with what competent speakers ordinarily say about the cases at issue (see Section 4.2, pp. 282–285) and are not perceptual or introspective in nature: Proponents of the argument consider verbal statements, which typically state a general fact (such as 'When viewed sideways, round things look elliptical') (see Ayer 1940, 3; Price 1932, 28), and then ask themselves 'what is to be inferred' (Ayer 1940, 4). They then (ii) find the phenomenal judgments they 'infer' 'plausible' (Broad 1923, 240) or even 'as plain as can be' (Moore 1918–19, 21). Typically, however, (iii) they are unable to say anything informative about why they confidently 'infer'—or leap to—phenomenal judgments from those premises: They often do not state the Phenomenal Principle (see p. 261) at all; when they do, they notoriously fail to provide any supporting argument beyond appeal to the plausibility of the conclusions obtained.[7] Since the 'principle' merely states the

[7] The one *apparent* exception is Broad (1923). But, *pace* Robinson (2001, 37), the statement he quotes from Broad (1923, 239) states not the Phenomenal Principle but the 'sensum theory' of perception that Broad bases on phenomenal judgments, and the one argument he appears to give for the principle (Broad 1923, 240f.; see Smith 2002, 36–37) explains why it may strike thinkers as plausible but is patently circular when uncharitably read as a justificatory argument.

Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

general form of these inferences without providing any reason why they should preserve truth, this suggests an inability to explain them and justify the plausible non-perceptual judgments at issue.

'*Negative judgments*' or attributions of non-awareness ('When viewing a coin sideways, subjects are not aware of the coin they look at') supposedly are consciously inferred from prior phenomenal intuitions. Some textual observations, however, suggest that also these supposed conclusions are intuitions, in the aetiological sense: In key texts by Russell (1912, 1–3), Broad (1923, 236, 240), Price (1932, 3), and Ayer (1940, 1–4), we find statements of the argument, or of the intuitive lines of thought underlying it, in which negative judgments immediately follow relevant case-descriptions and *precede* phenomenal judgments. Those authors find the paradoxical negative judgments they (i) ostensibly leap to from verbal descriptions (ii) so plausible they immediately seek to accommodate them in the light of apparently inconsistent facts, and (iii) comment on how difficult they found it to transform the underlying lines of thought into acceptable (rule-governed) arguments (see Ayer 1956, 89, and Price 1932, 27).

These potential intuitions (e.g., 'When looking at it sideways, the viewer is not aware of the penny') then prompt spontaneous protests:

> When I look at a penny from the side I am certainly aware of *something*; and it is certainly plausible to hold that this something is elliptical in the same plain sense in which a suitably bent piece of wire, looked at from straight above, is elliptical. (Broad 1923, 240)

The phenomenal judgment ('this something is elliptical') is spontaneously offered to characterize the 'something' Broad protests he *is* aware of. This move is consistent with the common practice of picking out things about whose identity we are unsure, by otherwise non-committal descriptions of how they look to us, from here, now, as when a rambler points into the valley and asks, 'Do you see that small, red patch? Might that be our car?'—or as Broad might, once his prior spontaneous judgment that he is not aware of the coin leaves him unsure about the identity of the thing of which he manifestly *is* aware (see below, Section 4.2, pp. 282–285).

Let's therefore provisionally assume that both phenomenal and negative judgments are intuitions in the relevant aetiological sense, and that negative prompt phenomenal intuitions in intuitive lines of thought, before their order gets turned around in efforts to construct acceptable arguments. We will now build up to a psychological explanation that will (1) vindicate this working hypothesis and (2) trace the targeted intuitions back to an automatic cognitive process that is generally reliable but predictably engenders cognitive illusions, under specific circumstances—which obtain in the formulation of arguments from illusion. We will thus try to show that influential intuitions about supposed optical illusions are genuine cognitive illusions.

If we succeed, we stand to resolve the paradox along the lines indicated at the outset, since its proponents do not offer non-circular arguments for the intuitive judgments on which their arguments rely: Current textbook reconstructions infer what we called 'negative judgments' from phenomenal judgments, but fail to provide justification for *these* beyond an appeal to their intuitive plausibility. Earlier authors argue, conversely, that the viewer is not aware of, e.g., the round coin that looks elliptical, but is aware of something—let's characterize it as an 'elliptical patch'. Here, the negative intuitive judgment that serves as premise remains unsupported, as authors merely leap to it from the initial case-descriptions. Thinkers' warrant for accepting these paradoxical intuitions hence depends upon the intuitions' evidentiary value. A debunking explanation of these intuitions therefore stands to resolve the intuitive paradox; analogous treatment of arguments from hallucination, etc., stands to resolve the problem of perception they jointly engender.

2 A psychological explanation

The non-perceptual 'negative' judgments at issue are prompted by brief verbal descriptions of cases of non-veridical perception—typically single sentences that serve as premises of those arguments. Psycholinguistic research has studied the inferences we spontaneously make in speech- and text-comprehension/production. It has uncovered automatic association processes in semantic memory (e.g., Neely and Kahan 2001) which duplicate inferences governed by heuristic rules—and can thus generate intuitions in the aetiological sense defined. Let's therefore explore whether the apparently intuitive judgments prompted by the present single-sentence case-descriptions can be explained by routine comprehension-related processes, and turn to psycholinguistics. One such routine process is that of *stereotype-driven amplification.* We will now present this cognitive process and show that it is generally reliable but capable of generating cognitive illusions.

2.1 A pertinent process

Many nouns (Hare *et al.* 2009) and verbs (Harmon-Vukić *et al.* 2009; Ferretti *et al.* 2001) are associated with stereotypes. *Stereotypes* are sets of properties which come to mind first, and are easiest to process, when we hear those nouns or verbs. Take the verb "to manipulate", and complete the following sentences with the first word that comes to mind:

Joe is so easily manipulated. He is so___
Jack is really good at manipulating people. He is so___

The most common answers are "gullible", "naïve", "stupid", and "cunning", "shrewd", "clever", respectively. Accordingly, cunning, shrewdness, and cleverness are subject-properties stereotypically associated with the verb "to manipulate", while gullibility, naivety, and stupidity are stereotypically associated patient-properties (where a verb's *patients* are the referents of its direct object). Many verbs are stereotypically associated with action-, subject-, and patient-properties.

Stereotypes guide both the spontaneous classification under object- or event-categories (see Kahneman and Frederick 2002), and spontaneous inferences from those expressions. These inferences are governed by a pragmatic heuristic derived from Grice's second Maxim of Quantity, 'Do not say more than you must!' (see Grice 1989, 26): According to the neo-Gricean *I-heuristic*, 'what is expressed simply is stereotypically exemplified' (Levinson 2000; see also Garrett and Harnish 2007). As a production rule, this heuristic has us do two things: It has us (i) skip mention of stereotypical properties in descriptions of stereotype-consistent situations (Brown and Dell 1987), while (ii) making explicit all deviations from the stereotype ("He managed to manipulate even shrewd Joe"). As a comprehension rule, it has us amplify the utterance content by assuming—in the absence of such indications to the contrary—that events, agents, and objects have the properties stereotypically associated with the given, say, verb. E.g., when we haven't been explicitly told anything to the contrary about the two protagonists, the heuristic has us infer from "Jack manipulated Joe" that Jack is cunning and Joe gullible.

Such *stereotype-driven inferences* are routinely made in text comprehension and can account, e.g., for this riddle:

(R) A young man and his father had a severe car accident. The father died, and the young man was rushed to hospital. The surgeon at the emergency room refused to operate on him, saying, 'I can't. He's my son.'—How is this possible?

If you find this question difficult, chances are that you leaped from the word "surgeon" to the conclusion that the speaker has the stereotypical features of surgeons—including the stereotypical gender (male) that is ruled out by prior context ('The father died'). This illustrates that stereotype-driven inferences are automatically made regardless of contextual information; to prevent them, the deviation from the stereotype must be made explicit ("female surgeon") (Givoni *et al.* 2013).

Inferences with the I-heuristic are duplicated by automatic association processes in semantic memory. Such *association processes* are studied through priming experiments (Lucas 2000): Participants are first presented with a stimulus or 'prime' (word, sentence, or short text) and then a 'probe' word or letter string, and have to either read out the word or decide whether the string forms a word. Under certain conditions, researchers infer from shorter response times (e.g., for "bank" and "money" than for "bank" and "honey") that the prime ("bank") activated the probe concept ('money'), i.e., increased the likelihood of its use by several cognitive processes, which include the interpretation of utterances employing the verb, and inferences from them (Peleg and Giora 2011). Strength of activation is inferred from the size of the response-time difference.

Semantic memory is our memory for facts as opposed to personally experienced or 'episodic' events (McRae and Jones 2013; Tulving 2002). It is standardly conceived as a semantic network. Such a network consists of nodes representing concepts and links between them that can automatically pass on activation from stimuli, verbal and other, along several pathways simultaneously (Allport 1985). Priming experiments serve to trace the pathways (Lucas 2000). Simultaneous activation of

Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

concepts can amount to the activation of a proposition made up of those concepts. An activated concept or proposition becomes conscious if—and only if—its activation exceeds a threshold as well as the activation levels of competitors. Spreading activation can therefore duplicate inferences, by spreading in sufficient strength from nodes representing one proposition to nodes that jointly represent another.

According to standard conceptions of semantic memory (Neely 1991; Kahneman 2011), the network constantly evolves in accordance with three principles: (1) The co-occurrence of features (things and their common properties, wholes and their common parts) and of events (causes and typical effects, etc.) forges links between the respective nodes. (2) These links grow stronger upon frequent activation. (3) They atrophy upon disuse. For example: The more frequently we encounter tomatoes that are red (in the supermarket), the stronger the links between the respective concepts become, the more activation gets passed on from the 'tomato'- to the 'red'-node, and the more strongly the verbal stimulus ("tomato") that activates the former activates the latter concept. This has two consequences: 'red' will be ever more likely to be among the first concepts to come to mind when we hear "tomato" (i.e., to become stereotypically associated with the word), and the more likely we are to infer that the thing talked about is red (i.e., stereotype-driven inferences result from basic processing principles of semantic memory).

These *stimulus-driven* (bottom–up) processes are fully automatic and entirely determined by current linguistic stimuli and immediately preceding words. Their conclusions are subsequently integrated with the outputs of similarly automatic *top–down* processes that occur in parallel to stimulus-driven processes but are sensitive also to earlier linguistic input and the deliverances of other cognitive processes, including visual perception (review: Giora 2003). Where, however, contexts are comparatively uninformative and deviations from stereotypes remain implicit, the input through parallel top–down processes is minimal, and stereotype-driven inferences are liable to go through. This was the case with vignette (R), where we were not told about the speaker's female gender and did not get to see or hear her. Crucially, it is liable to occur with brief philosophical case-descriptions that are unaccompanied by visual information. The underlying processes have been shown to occur not only in the comprehension but also the production of speech and text (Levelt 1989; Pickering and Garrod 2013; Stephens *et al.* 2010; see also Giora 2003, 134–136). They are hence set to duplicate inferences with the I-heuristic not only in interpersonal communication but also in the sort of subvocalized cognition characteristic of abstract philosophical thought.

As noted, links between concept-nodes not only grow stronger with frequent use but also weaker upon disuse: The more unripe green tomatoes one is exposed to instead of red ones, the weaker the association between 'tomato' and 'red' becomes. The strength of stereotypical association is thus sensitive to co-occurrence frequency in the sample to which the subject is exposed.[8] Unless this sample is

[8] Strength of association is also influenced by prototypicality (Rosch 1978), and hence also depends upon cognitive principles of abstract categorical organization (Giora 2003).

seriously skewed by biasing media (British tabloids continually vilifying the EU), stereotypes therefore tend to be reasonably accurate and get gradually modified where they have become inaccurate. Outside periods of rapid change (and topics notoriously attracting biased comment and motivated misrepresentation), stereotype-driven inferences are, by and large, reasonably reliable.[9]

We now build up to one set of circumstances under which this generally reliable process will predictably lead to cognitive illusions. According to the well-supported *graded salience hypothesis* (Giora 2003), where a word has several distinct senses, its utterance will activate most rapidly and strongly the concepts that are stereotypically associated with the most frequent or dominant use of the word (e.g., "mint"), and activate these *dominant stereotypical associates* ('candy'), regardless of context ("All buildings collapsed except the mint", Simpson and Burgess 1985; Till *et al.* 1988). In particular in uninformative contexts, we are then liable to spontaneously infer the presence of dominant stereotypical associates also where the word is used in an infrequent sense. To prevent such inappropriate inferences from ambiguous terms, speakers often explicitly mark uses of infrequent senses through such riders as "figuratively speaking" which can enhance the activation of relevant concepts that might otherwise be sidelined by preferential activation of concepts associated with dominant uses (Givoni *et al.* 2013).

Now suppose speakers *unwittingly* use a well-established word in a new sense, which licenses application to situations which do not conform to the established stereotype, i.e., to the stereotype that is associated with the word's well-established and dominant use. These speakers will then apply the word to such stereotype-inconsistent situations, without making the deviation from the stereotype explicit. They will thus unwittingly violate the production part of the I-heuristic. Its comprehension part is then liable to lead us astray: In particular in uninformative contexts, such violations will prompt stereotype-driven inferences to wrong conclusions about the stereotype-inconsistent situations, namely, to conclusions that wrongly attribute properties that are stereotypically associated with the dominant use of the word. Due to the effortlessness or fluency of stereotype-driven inference, the resulting judgments will strike thinkers as plausible (Kelley and Lindsay 1993), irrespective of reflective endorsement (Thompson *et al.* 2011).

2.2 A key hypothesis

All this happens in our arguments from illusion, whose initial premises employ the verbs "appear", "seem", and (infrequently) "look", more or less interchangeably,[10] and infer intuitive conclusions about what subjects are 'aware of', 'in perception'. Philosophers of perception often want to cover all five senses

[9] E.g., the inference in (R) will most frequently lead to true conclusions, in most Anglophone countries: In England, for instance, only 9.2 per cent of surgeons were female in 2012 (Royal College of Surgeons, <http://surgicalcareers.rcseng.ac.uk/wins/statistics>, accessed 17 September 2014).

[10] "appear", e.g.: Ayer (1940, 3), Robinson (2001, 57), Russell (1912, 2), Smith (2002, 25); "seem" e.g.: Ayer (1940, 3), Broad (1923, 239–240), Crane (2011, 3), Moore (1918–19, 21–23), Russell (1912, 2).

simultaneously and draft verbs that already have well-established uses in ordinary language into service as generic terms, without realizing the novelty of their use. They thus use the verb "to perceive" as mere abbreviation of "to look or hear or smell or taste or feel" and use "to be aware of" even more generically, to talk simultaneously about our five senses and associated experiences, without having to commit themselves to what kind of thing—'physical or psychical or neither' (Price 1932, 3)—we are seeing or hearing (e.g., sounds of the kind we hear *with* our ears or of the kind tinnitus patients constantly hear *in* their ears). In the same generalizing vein, "seem" and "appear" are then used to speak at one go about how things look or sound or smell or taste or feel to the subject of perception or awareness.[11] Finally, proponents of the argument from illusion take themselves to be using "look" and its cognates—as well as the supposedly merely more generic "appear" and "seem"—in a purely phenomenal sense, in which they imply nothing about what subjects are inclined to judge or do believe.

We will now, however, build up to the hypothesis that in their dominant pre-philosophical uses, "*x* appears *F* (to *S*)", "*x* seems *F* (to *S*)", and "*x* looks *F* (to *S*)" are applied in situations in which the (often implicit) patient *S* is at least inclined to judge, think, or believe that *x* is *F*, so that 'appearance-verbs' enjoy strong stereotypical association with this doxastic patient-property. Berit Brogaard (2013, 2014) has argued that, in their intransitive uses ('Joe looks dirty', as opposed to 'Joe looks dirtily at her'), "look", "appear", and "seem" are subject-raising verbs (Postal 1973) which are semantically unrelated to their grammatical subject ('Joe') and serve not so much to attribute any property from their complement (dirtiness) to the grammatical subject's referent (Joe) as to indicate an experiential, doxastic, or epistemic attitude of the patient to a content (Joe is dirty).

All three verbs are used not only in visual but also in non-perceptual contexts ('This plan looks/appears/seems clever').[12] In visual contexts, their doxastic implication (that the patient is inclined to judge that *x* is *F*) is generally defeasible. In many non-perceptual contexts, it is not (as witnessed by the anomaly of 'The risks still appear/seem/look manageable to the analyst, but she no longer thinks they are'). Take "to appear". The *Oxford English Dictionary* explains two related intransitive uses: In one use ('to be in outward show, or to the superficial observer', with both visual and non-perceptual examples), it carries a strong but defeasible doxastic implication (the acute observer need not fall for the outward show); in the other ('to be in one's opinion; to be taken as'), the doxastic implication is indefeasible. Either way, the implication is unaffected by *suggestio falsi*: Even when the speaker suggests that things are not what they seem, look, or appear to

[11] E.g., Broad (1923, 236, my italics): 'When I judge a penny *looks* elliptical … ' but 'This *seems* to me elliptical, or red, *or hot*' when covering different sense-modalities. Since Chisholm (1957) until today (Brogaard 2014), all seven verbs are jointly categorized as 'appear words' differing, basically, only in degree of generality or the perceptual sense invoked.

[12] These uses are not happily captured by familiar philosophical explications of 'phenomenal' and 'epistemic' uses of "looks", etc. (Chisholm 1957; Jackson 1977; Maund 1986; Brogaard 2014), which we therefore set aside.

someone, she still suggests that this patient is inclined to judge that they are (whence the lack of *suggestio falsi* for first-person present-tense statements).[13]

The hypothesis that, in their relevant uses, all three verbs are strongly associated with doxastic (rather than experiential or epistemic) patient-properties can be supported by distributional semantics analysis. While the reader has to be referred elsewhere for an explanation of distributional semantics (e.g., Erk 2012; Turney and Pantel 2010) and for proper presentation of the most relevant results (Fischer *et al.* in prep.), the basic idea here is to infer conclusions about the meanings and stereotypical associates of words from the linguistic contexts in which they are used—more specifically, from their 'distribution': Two predicates (say, "seem" and "find") have a *similar distribution* in a corpus to the extent to which they co-occur in the corpus with the same other words (say, "probable", "odd", "irritating", etc.) as arguments, in the same proportion(!).[14] If "seem(to)" and "appear(to)" have a distribution highly similar to that of "think", "believe", and "find" (in its doxastic sense), they stand to be used interchangeably with these doxastic verbs in a variety of prototypical contexts (after argument swapping), and to be stereotypically associated with doxastic patient-properties. If they are distributionally more similar to those doxastic verbs than to experiential terms or epistemic verbs like "know" and "realize", then "seem" and "appear" probably are more strongly associated with doxastic than with either experiential or epistemic attitudes.

These tentative conclusions are suggested by an analysis of the intransitive use of "seem", "appear", and "look" in a vast corpus, namely Wikiwoods (Flickinger *et al.* 2010). The Wikiwoods corpus is a Wikipedia snapshot parsed with the English Resource Grammar (ERG: Baldwin *et al.* 2004).[15] The ERG output distinguishes between the uses of ambiguous words in different senses (whence such entries as 'find(mental)' in our results). Our analysis identified the 'nearest neighbours' of "seem(to)", "appear(to)", and "look(to)": those verbs that are distributionally significantly more similar to them than any others.

The results are given in the Appendix to this chapter (p. 287), for the reader's perusal. Note that the doxastic verbs "believe", "think", and "find(mental)" are among the nearest neighbours of all three targets. Indeed, the five distributionally most similar nearest neighbours of "seem(to)" and "appear(to)" include all three verbs, while those of "look(to)" include "find(mental)". By contrast, the nearest

[13] See the usage note for "seems", *Oxford English Dictionary* (2nd ed. 1989, online version June 2012). In this first-person use (with often implicit reference 'to me'), preference of "seems *F*" over the simpler "is *F*" warrants the inference of 'doubt-and-denial conditions' (Grice 1961).

[14] For example, the verbs "seem (*x*, *Fx*, *y*)" and "think (*y*, *Fx*)" are classified as distributionally similar because the same expressions *F* are found to fill the second argument-slot, in the same proportions. The adjectives in this slot generally characterize abstract notions and people; these strong selectional preferences ensure that values for *x* are drawn from the same semantic categories.

[15] For the purpose of our experiments, it was converted into a so-called Dependency Minimal Recursion Semantics (Copestake 2009) format, giving us a representation of the text akin to a set-theoretic formalization (Montague 1974), albeit in an underspecified form. Our results were obtained from the DISSECT toolkit (Dinu *et al.* 2013).

Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

neighbours we identified do not include any clearly experiential terms; and while they do include epistemic verbs ("know", "realize"), these enjoy a significantly lower distributional similarity to all three target words than the doxastic "believe", "think", and "find(mental)".

In conjunction with our previous considerations, these results motivate the working *hypothesis (H)* that "*x* seems *F*", "*x* appears *F*", and possibly "*x* looks *F*" serve primarily to indicate a doxastic, rather than an experiential or epistemic, attitude of patients, and are strongly stereotypically associated with at least the weak doxastic patient-property '*S* is inclined to judge that *x* is *F*'. We will now see that stereotype-driven inferences which exploit the hypothesized associations and proceed from standard formulations of the initial premises of arguments from illusion can explain intuitive judgments to the effect that the subject is not aware of, say, the coin she views (Section 2.3, pp. 271–274). Then we will experimentally test the working hypothesis on which the explanation relies (Section 3, pp. 274–280, below).

2.3 The explanation

Most arguments from illusion proceed from perfectly familiar cases of non-veridical perception, in which no adult is inclined to judge that the thing that looks *F* (elliptical, small, yellow) under the circumstances (from this perspective or distance, or in this light) actually is *F*. In this respect, these cases are inconsistent with the stereotypes associated with the dominant uses at least of the verbs "seem" and "appear". Even so, proponents of the argument typically use precisely these verbs to formulate their initial premises (see note 10, p. 268) and fail to explicitly mark these stereotype-inconsistent uses (say, through riders like "in a purely phenomenal sense"). They thus violate the production-side of the I-heuristic.

They arguably do so inadvertently and because they do not realize the dominance of those doxastic uses and do not bear in mind the strength of the doxastic implications at issue.[16] Even recent commentators who are perfectly familiar with the distinction between 'phenomenal' and 'epistemic' senses of 'looks' (made popular by Chisholm 1957), think they need to advert to 'those uses of the language of appearing that may be called "merely evidential"' only long after stating the argument and feel they can then 'set aside' these uses without further ado (e.g., Smith 2002, 37).

In the absence of relevant textual markers, readers—and authors—are prone to make stereotype-driven inferences in line with the comprehension-side of the I-heuristic, even when these inferences lead to contextually inappropriate conclusions (Section 2.1, pp. 265–268). These are particularly likely to go through without essential correction in uninformative contexts, like those of typical statements of arguments from illusion, where neither prior text nor simultaneous visual

[16] By contrast, Austin (1962, 36) stresses the doxastic features of the dominant use of "seem" and "appear" in discussing the 'root idea behind the use' of these verbs and criticizing the argument. But even Austin dismisses the doxastic uses of "looks" as irrelevant (*ibid.*, n. 1).

Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

stimuli provide further information whose integration would lead to their suppression (Gernsbacher and Faust 1991; Faust and Gernsbacher 1996; Williams 1992).

By our hypothesis (H), the I-heuristic licenses inferences from standard formulations of premises like

(P) A round coin seems elliptical when viewed sideways

to conclusions like

C_0 The viewer is inclined to judge that the object viewed is elliptical.

This stereotype-driven inference is initially duplicated by stimulus-driven association processes. When equally automatic top–down processes integrate C_0 with further information from the sentence-context, a potentially conscious judgment results: (P) tells us that the object viewed is round. Hence the judgment the viewer is inclined to make is false. The viewer hence lacks the confident true belief required for knowledge and sports either the indecision or false belief stereotypical of ignorance. The integration of C_0 with this other information from the sentence-context thus yields

C_1 The viewer is ignorant of the shape of the coin/object viewed.

The activation, by (P), of C_0 corresponds to the inference from the accident vignette (R) (in Section 2.1, pp. 265–268) to the conclusion that the surgeon is male. The present leap all the way to C_1 is facilitated by the stereotype of 'ignorance', whereas, in (R), contextual integration into the accident scenario required deviating from the activated stereotype—and accordingly proved too difficult for effortless processes.

According to standard conceptions of semantic memory as a spreading-activation network (Section 2.1, pp. 265–268), nodes representing properties, relations, or other 'semantic features' are linked to nodes representing their typical bearers, and thus provide an indirect link between nodes representing 'semantically similar' concepts (as psycholinguists put it in *their* idiom; e.g. Oostendorp and Mul 1990, 36–37):

> A *concept* is *semantically similar* to another, for a subject S, to the extent to which S takes the things (individuals, stuffs, properties, etc.) they stand for to share the same attributes or to stand in the same relations.

The more semantically similar two concepts are, the more activation will be passed on, through an increasing number of shared 'property-nodes', from (nodes representing) one to (nodes representing) the other, when activated. Along with the concept it stands for, a verbal stimulus therefore strongly activates concepts standing for near synonyms. Activation is particularly strong and sustained when a word is expected in the context (Barnhardt *et al.* 1996). When a near synonym is

Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

significantly more expected in the context than an initially activated concept, its node may be activated more strongly and the near synonym may be the word to come to the thinker's mind when hearing or formulating the judgment (Park and Reder 2004).

In its dominant ordinary use, the word "unaware" is a near synonym of "ignorant" (in C_1), as the *OED* reminds us:

> to be aware of = to have cognizance, know, viz. have knowledge as obtained by observation or information.[17]

Philosophers typically consider premises like (P) in the light of a guiding question. This is not the question of what subjects know, but the question of what subjects are aware of in the cases described. Hence attention is directed towards 'aware' and its antonym, and 'unaware' is set to become more strongly activated than 'ignorant', when a philosopher leaps from (P) to C_1. He will then most likely make the resulting judgment in the shape of:

C_1* The viewer is unaware of the shape of the coin/object viewed.

Arguments from illusion most frequently invoke objects with characteristic and standardized shape, size, or colour, like penny coins (Ayer 1940, 3, and 1956, 86; Broad 1923, 239ff.; Robinson 2001, 53). When (C_0) people are inclined to judge that they have a different shape, they are typically undecided or wrong not only about the shape, but also about the kind or nature of the object—C_2: They don't know it is a coin they are viewing. In particular when thinkers consider premises like (P) in the light of the more specific question of what kind of objects viewers are aware of, and of whether or not they are aware of the (say) physical coin, the processes outlined will take them to conclusions like:

C_2* The viewer is unaware of the coin viewed.

Judgments about sensible properties (like C_1*) predominate in early analytic formulations of the argument from illusion (e.g., Ayer 1940, 4; Russell 1912, 3), while current reconstructions mostly contain judgments about objects (like C_2*).

To sum up, this explanation traces the paradoxical intuition that, in a thoroughly familiar case of perspective, the viewer is not aware of the coin or its shape, to two factors: to unwitting violation of the I-heuristic, followed by automatic cognitive processing that duplicates inferences governed by it. This

[17] The same goes for the philosophical notion of 'direct awareness' which, in addition, requires that the subject acquire the relevant knowledge without—conscious—inference or other intellectual process (Price 1932, 3; Russell 1912, 4; see also Fischer 2011, 114–116). 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, 59, 61, 69), so that no inference is required to find out whether they merely appear or actually are *F* (see Broad 1923, 239–240, 248).

Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

explanation predicts that, in intuitive statements of arguments from illusion and related lines of thought, the judgments of non-awareness we have explained will frequently be accompanied by sceptical judgments to the effect that viewers don't know, and can, or should, doubt what it is they are looking at. Since the final conclusion of the argument from illusion has been used as a basis for classical sceptical arguments (review: Ayer 1956), but not vice versa, this prediction may be surprising. But it is indeed borne out by key passages from Russell (1912, 1–3), Broad (1923, 236, 240), Price (1932, 3), and Ayer (1940, 1–4), where argumentatively unsupported attributions of ignorance and expressions of possible doubt accompany attributions of non-awareness and both precede phenomenal judgments (see Section 1, pp. 260–265).

We will now shore up the proposed explanation of intuitive attributions of non-awareness (like C_2*) (Section 3, pp. 274–280), before developing an explanation of how these intuitions cause equally intuitive phenomenal judgments (Section 4, pp. 280–286). Both explanations will allow us to assess the evidentiary value of the intuitions explained—the philosophical prize at stake.

3 An experiment

The crucial first step of the proposed explanation relies on the hypothesis (motivated in Section 2.2, pp. 268–271):

(H) "*x* looks *F*", "*x* appears *F*", and "*x* seems *F*" are strongly stereotypically associated with the patient-property '*S* is inclined to judge that *x* is *F*', namely, strongly enough to support automatic stereotype-driven inferences in verb comprehension.

To develop and test this key hypothesis, we conducted the following experiment. (For more detail, and an experimental defence of the second half of our proposed explanation, see Fischer and Engelhardt forthcoming.)

3.1 Approach and predictions

We employed a forced-choice plausibility-ranking task: In this paradigm, participants are presented with *minimal pairs* of sentences or short two-sentence) texts that differed only in one *critical verb*, e.g.:[18]

29(a) To Jack, Jane looked tanned. He supposed it was just a trick of the light.
29(b) To Jack, Jane seemed tanned. He supposed it was just a trick of the light.

Participants are asked to indicate which of each pair's two constituents—(a) or (b)—strike them as more plausible, and to do so even if they have no clear-cut

[18] Here and below, numbered items are taken from the questionnaire used in our study.

preference. In our minimal pairs, the phrases or sentences containing the critical verb were followed by a sequel inconsistent with the hypothesized stereotypical associate: If "*x* seems *F*" is stereotypically associated with the patient-property '*S* is inclined to judge that *x* is *F*', then readers of the first sentence (in 29(b)) will leap to the conclusion that Jack is inclined to judge that Jane is tanned. Since the sequel ("He supposed it was just a trick of the light") is inconsistent with this conclusion, readers will then find this line (29(b)) implausible. This allows us to test hypotheses about the relative strength of stereotypical association: If the association with the patient-property at issue is weaker for "looks" than for "seems", then participants will find the look-sentence in (a) clashes less strongly with the sequel than the seem-sentence in (b), and will judge (a) more plausible than (b).

For each pair of verbs, we present several such minimal pairs, where the verbs are followed by sequels incongruent with attributions of the same event-, subject-, or patient-property *P*. Where preferences between such minimal pairs are random for a particular pair of verbs, both verbs are associated with *P* equally strongly—and, possibly, not at all. Where, however, we already have independent evidence that a verb is stereotypically associated with *P*, random choices between minimal pairs which pit it against another verb support the conclusion that also the latter has an—equally strong—stereotypical association with *P*. Distributional-semantics analyses provide us with independent evidence that both "appear" and "seem" are stereotypically associated with a doxastic patient-property (Section 2.2, pp. 268–271). We can therefore employ the forced-choice plausibility-ranking task to test three specific predictions:

P_1 We conjecture that, proportionally, "seem(to)" is used even more frequently than "appear(to)", and "look(to)" in non-perceptual contexts in which doxastic implications are indefeasible, so that its stereotypical association with the doxastic patient-property '*S* is inclined to judge that *x* is *F*' is stronger than the analogous associations of "appear" and "look". We therefore predict that appear- and look-sentences will be consistently preferred over seem-sentences, in minimal pairs (similar to 29 above).

P_2 We further conjecture that, even so, "look" has its intransitive use mainly in non-visual contexts in which its doxastic implications are indefeasible (see Section 2.2, pp. 268–271). If so, its stereotypical association with that doxastic patient-property should not be significantly weaker than the analogous association with "appear", but of roughly similar strength. We therefore predict that preferences between looks- and appears-sentences will not be significantly different from chance.

P_3 However, we expect some philosophers to perform differently: Since "appears" is infrequently used in ordinary discourse (Leech *et al.* 2001), the entirely interchangeable use of "appears" and "seems" in the philosophy of perception (Section 2.2, pp. 268–271)—which leads to their application in the same situations—should suffice to assimilate their stereotypical associations to the point where philosophers who work in the area make random choices between appear- and seem-sentences.

To infer conclusions about stereotypical associations from the confirmation of these predictions, we will of course need to exclude alternative explanations (below).

3.2 Method

3.2.1 Participants

47 volunteers participated without compensation. All were native speakers of English. They were drawn mainly from the School of Philosophy at the University of East Anglia: 8 members of teaching staff holding a PhD in philosophy, 27 students, and 2 clerical staff. 10 local extra-academic professionals (lawyers, geologists, and geo-engineers) mitigated the bias potentially inherent in the common restriction to campus-based samples.

3.2.2 Materials

A questionnaire contained 66 minimal pairs and instructed participants to first read each sentence in a pair carefully and then indicate which of the two 'strikes you as more plausible'. These pairs included six items for each of the pairings:

(1) look/appear
(2) look/seem
(3) appear/seem.

Items involving "*x* looks *F*" or "*x* appears *F*" contained sequels inconsistent with attributions of the doxastic patient-property '*S* is inclined to judge that *x* is *F*'. Perhaps surprisingly, an earlier pilot study (on 45 undergraduate philosophy students from the same university) had shown that, for look-, appear-, and seem-items, the nature of the complement (adjective vs infinitival construction, e.g., 'seems elliptical' vs 'seems to be elliptical') makes no difference to the plausibility judgments elicited. Item-types (1) to (3) all used both constructions, though always the same for both constituents of an item. They included shape, size, and colour adjectives in the complements of "looks", "appears", and "seems". Look/appear and look/seem items employed only visual contexts, while half of the appear/seem items used non-visual contexts which suggested patients would make non-perceptual judgments, such as: 'The accused appears/seems guilty. The jury foreman thinks he is innocent.' Of the fifteen visual contexts for looks- and appears-items, ten explicitly invoked familiar cases of non-veridical perception (see below) while the others left the actual facts of the matter open (like 29, above). To exclude order effects, each critical verb appeared half the time in the first constituent of the pair and half the time in the second. E.g. (presented in compressed format):

46(a/b) To Adam, the tree at the far end of the enormous park appeared/looked small. He thought it was a huge, ancient Redwood.

50(a/b) From his vantage point, the curio looked/seemed elliptical to John. He thought it was round.

60(a/b) The athletes in the arena seem/appear to be tiny from the top of the bleachers. Sitting there, Amanda thinks they are outrageously tall.

To exclude pertinent confounds, items were designed and tested (see Section 3.2.3, p. 278, below) to minimize the relevance of *extraneous background knowledge*, i.e., background knowledge about the world other than the frequency information implicit in the stereotypical associations of the critical verb. E.g., trees (in 46) come in many shapes and sizes, as do curios (in 50), in contrast with (typically round) coins, while unspecified athletes (in 60) vary in height more than, say, basketball players, and bleachers and stadiums vary greatly in size, too.

However, the less able subjects are to base their plausibility judgments on factual knowledge, the more they base them on metacognitive cues. The most important of these is *fluency*, i.e., the subjective ease we experience in processing the relevant information (review: Alter and Oppenheimer 2009). When assessing the plausibility of statements about fictitious protagonists (say, Jack and the possibly tanned Jane)—deliberately constructed so as not to engage with background knowledge—participants' judgments will largely depend on the level of fluency they experience in reading and understanding the relevant statements. Our experiment relies on the fact that this level is affected by clashes between the conclusions of stereotype-driven inferences readers automatically make in verb comprehension, and the content of further text. The level is, however, also affected by other factors, including the legibility of the text (Alter *et al.* 2007), the syntactic complexity of the sentence (Lowrey 1998), and the familiarity and pronounceability of the individual words (Oppenheimer 2006). Most of these other factors are controlled for by presenting minimal pairs (with the same syntactic structure, employing the same words but one, etc.) in the same style. But our critical verbs themselves have different frequency and, hence, familiarity: In a standard corpus (Leech *et al.* 2001), "look" appears twice as often as "seem", which is twice as frequent as "appear".

To exclude the possibility that the predicted preferences of look- over seem-sentences would be due to differences in word frequency, we constructed 30 filler items: minimal pairs whose critical verbs differ in the frequency with which they are used in ordinary discourse. Each verb-pair was used in two such items: In the *frequency-congruent* item, the text employing the more frequently used verb was also more consistent with the relevant stereotype. In the *frequency-reversed item*, stereotype-consistency and word frequency pulled in different directions. E.g.: "obey" is used more frequently than "comply", and is more strongly associated with the patient property 'S has authority of command'. Hence item 61 below is frequency-congruent and 3 below frequency-reversed:

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

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

We then used performance on these fillers to identify '*potentially frequency-sensitive*' participants who performed better (made more stereotype-consistent judgments) on the frequency-congruent than the frequency-reversed fillers, and '*frequency-insensitive*' participants whose judgements were clearly unaffected by word frequency, as their performance on frequency-reversed fillers was no worse, or even better, than that on frequency-congruent fillers. This allowed us to compare the two groups' performance on our critical items.

3.2.3 Procedure

In constructing frequency-congruent and -reversed filler items, we used word-frequency information from a British English corpus appropriate for the British participants in this study, namely Leech *et al.* (2001).

In the actual study, participants were instructed to respond as quickly as possible, as we were interested in initial responses taking less than 5 seconds—at which point effortful processing may modify otherwise intuitive judgments (De Neys 2006).

3.3 Results and discussion

The results (Table 12.1) confirm our predictions. In line with P_2, preferences between look- and appear-sentences were not significantly different from chance. In line with P_1, the preference of look- and appear- over analogous seem-sentences was significantly above chance—and of exactly the same strength. These results are consistent with the hypothesis that "look" is as strongly associated with the doxastic patient-property at issue as "appear", and that this stereotypical association is even stronger for "seem". Importantly, the change from visual to non-visual contexts did not significantly affect the preference for "appear" over "seem" (0.70 vs 0.64, $t(45) = -0.96$, $p > .05$). The difference between these kinds of contexts hence either makes no difference to the use of these verbs, or affects both to the same extent. In other words, our findings are inconsistent with the suggestion that, in visual contexts, these verbs are understood in a 'purely phenomenal' sense in which they are both devoid of doxastic implications, while being understood—in non-

Table 12.1 Means, standard deviations, and *t*-tests for the full sample ($N = 47$)

Look/appear	0.52 (0.32)	$t(46) = 0.40$, $p > 0.05$
Look/seem	0.67 (0.23)	$t(46) = 5.11$, $p < 0.01$
Appear/seem	0.67 (0.23)	$t(46) = 5.74$, $p < 0.01$
Frequency-congruent filler	0.82 (0.11)	$t(46) = 20.67$, $p < 0.01$
Frequency-reversed filler	0.83 (0.12)	$t(46) = 19.63$, $p < 0.01$

perceptual contexts—in a different ('epistemic') sense in which they have doxastic implications of different strengths. In contrast, our results are consistent with the hypothesis that "appear" and "seem" have roughly similar doxastic implications in both visual and non-perceptual contexts—"seem" consistently more strongly and indefeasibly than "appear".

Participants made stereotype-consistent judgments about filler items as often when stereotype-consistency and word frequency pulled in the same direction (0.82) as when the two pulled in different directions (0.83) ($t(46) = -0.46$, $p > .65$). Ostensibly, word frequency did not influence their plausibility judgments. For closer investigation, we identified 18 potentially frequency-sensitive and 29 frequency-insensitive participants (see Section 3.2.2, pp. 276–278). Both groups clearly reproduced the response pattern we had found for the overall sample (Table 12.2).

These findings eliminate the key confound of word frequency and allow us to interpret our results as not merely consistent with, but supportive of hypotheses about strength of stereotypical associations.

The responses of the eight professional philosophers in our sample were consistent with our specific prediction P_3 about philosophers of perception. These participants were highly stereotype-sensitive in their judgments: They were even better than other participants at making stereotype-consistent judgments about filler items (overall means 0.88 (philosophers) vs 0.82 (non-philosophers), respectively), and yet significantly ($t(45) = -2.35$, $p < .05$) better at making such judgments about frequency-reversed filler items (0.92) than about frequency-congruent items (0.84). Word frequency evidently did not affect their judgments at all. These philosophers made random choices not only between appear- and looks-sentences (0.42, $t(7) = 0.51$, $p > .05$), like everybody else, but, as predicted, also between appear- and seem-sentences (0.54, $t(7) = -0.94$, $p > .05$). Their preference for look- over seem-sentences was even more pronounced than that of other participants (0.92 vs 0.62, $t(45) = -3.64$, $p < .01$). This suggests a twofold conclusion: first, for these philosophers, the stereotype differences between "looks" and "appears", and "appears" and "seems", respectively, are too slight to affect plausibility judgments even in a forced-choice task. Second, these two subthreshold differences add up to a difference just large enough to consistently affect these philosophers' rankings of looks- vs seems-sentences. The first conclusion is consistent with the hypothesis that philosophers' non-standard and interchangeable

Table 12.2 Means and standard deviations based on sensitivity to word frequency

	F-sensitive (18)	*F-insensitive (29)*	*Paired-sample t*-test
Look/appear	0.52 (0.34)	0.52 (0.31)	$t(45) = 0.03$, $p > 0.05$
Look/seem	0.71 (0.22)	0.65 (0.24)	$t(45) = 0.91$, $p > 0.05$
Appear/seem	0.71 (0.16)	0.65 (0.23)	$t(45) = 0.97$, $p > 0.05$
Frequency-congruent filler	0.88 (0.08)	0.79 (0.11)	$t(45) = 3.26$, $p < 0.01$
Frequency-reversed filler	0.74 (0.11)	0.89 (0.08)	$t(45) = -5.44$, $p < 0.01$

use of "seems" and "appears" (Section 2.2, pp. 268–271) has assimilated the stereotypes they associate with these verbs. The second conclusion is consistent with the finding that the philosophers among our participants were particularly good at making stereotype-consistent judgments.

To sum up, our results confirm the hypothesis (H) that "seem", "appear", and "look" all have sufficiently strong stereotypical associations with the doxastic patient-property '*S* is inclined to judge that *x* is *F*' to support stereotype-driven inferences of the kind identified at the root of the argument from illusion (Section 2.3, pp. 271–274). We have also obtained first evidence (if from a very small sample) that in philosophers' minds these associations are of roughly equal strength, so that these inferences cannot be prevented by mere substitution of the verb in the formulation of the argument's premises.

3.4 Future research

A well-established battery of psycholinguistic tests can be used to further confirm that people make stereotype-driven inferences relying on the verbs' association with doxastic patient-properties, and make such inferences also in inappropriate contexts (where, e.g., familiar cases of non-veridical perception, as in items 46, 50, and 60 above, constitute even more inappropriate contexts than 'non-committal' items like 29). Inferences to sequel-incongruent conclusions lead to slowdowns in reading (longer fixation-times for sequels) and signature electrophysiological responses (N400s). To test hypotheses about automatic inferences, psycholinguists therefore use materials which have been normed to exclude possible confounds, e.g., through plausibility ranking tasks like ours, for further studies, including reading-time measurements (Klin *et al.* 1999; Harmon-Vukić *et al.* 2009) with eye-tracking (Patson and Warren 2010) and electrophysiological measurements of event-related brain potentials (Kutas and Federmeier 2000, 2011). Explanations of intuitive judgments in terms of stereotype-driven inferences in language comprehension are thus capable of rigorous confirmation (or disconfirmation) through a series of complementary experiments.

4 From explanation to assessment

We now turn to the philosophically decisive step from explanation to assessment and will explore how our explanation can help us determine the evidentiary value of the intuitions explained. A thinker's intuition has *evidentiary value* to the extent to which the mere fact that she has this intuition, as and when she does, speaks for its truth. In line with the bipartite aetiological definition of 'intuition' (Section 1, pp. 260–265), we can distinguish two elements of such evidentiary value: First, the mere fact that a thinker spontaneously makes (i.e., automatically infers) the judgment at issue, under the relevant circumstances, may speak for its truth. Second, the fact that the thinker feels sure or confident in making it may speak for its truth. We will now explore how psychological explanations can help us assess the evidentiary value, first, of the spontaneity of judgments (Section 4.1, pp. 281–282) and, second, of the subjective confidence they inspire (Section 4.3, pp. 285–286). To do so, we will revisit our explanation of negative intuitions (Section 4.1, pp. 281–282) and build up to an explanation of phenomenal intuitions which various philosophers confidently declared 'obvious' (Section 4.2, pp. 282–285).

4.1 The evidentiary value of spontaneity

Cognitive illusions are predictable wrong intuitions which can be modified or even completely corrected by effortful reflection, but strike us as plausible, even once we know they are wrong (see Pohl 2004, 2–3). We have traced the negative 'unawareness-intuitions' prompted by the initial premises of arguments from illusion back to a cognitive process, namely stereotype-driven amplification, which is generally reliable but engenders cognitive illusions under specific circumstances (Section 2, pp. 265–274). This GRECI explanation would seem to warrant according stereotype-driven intuitions in general some evidentiary value, as a default.[19] Once we have a comprehensive epistemic profile of the process and have identified all vitiating circumstances, we can demonstrate, more specifically, that a particular intuition of a particular thinker has evidentiary value by showing that it has been generated by this process in the absence of any vitiating circumstances. Debunking explanations which show that a particular intuition has no evidentiary value are less involved: We 'only' need to show that it has been generated by the process at issue under one set of vitiating circumstances that we have already identified.

In this chapter, we have identified one set of circumstances under which stereotype-driven amplification can engender cognitive illusions: When authors unwittingly introduce non-dominant uses of a word, and make them without explicit marking, the authors and their readers are liable to infer the presence of stereotypical features associated with the dominant use, in insufficiently rich contexts, regardless of whether these are appropriate (stereotype-consistent) or inappropriate (stereotype-inconsistent) contexts. Where this happens, automatic inferences relying on the stereotype cease to be reliable. The finding that speakers have unwittingly introduced such a non-dominant sense—such as philosophers' phenomenal sense of appearance-verbs—therefore provides an 'undermining defeater' (Pollock 1984) for intuitions traced back to stereotype-driven inferences from the relevant words, in *any* context: It shows that the fact that competent speakers have these intuitions, as and when they do (which now is in appropriate and inappropriate contexts alike), no longer *eo ipso* speaks for their truth.

Where we further find that an intuition is produced by the process in a situation which, in a relevant respect, fails to conform to the pertinent stereotype, we obtain a 'rebutting defeater' (*ibid.*) and expose the intuition explained as a cognitive illusion. E.g., the finding that, in their dominant use, appearance-verbs enjoy a strong stereotypical association with doxastic attitudes allows us to predict that thinkers will form wrong intuitions when they unwittingly apply these verbs, in a

[19] This is a highly restricted form of 'phenomenal conservatism', namely 'dogmatism' (Tucker 2013) about stereotype-driven intuitions.

non-dominant ‘phenomenal’ sense, to perfectly familiar situations of non-veridical perception, where nobody is inclined to judge that *x* is as *F* as it ‘seems’, ‘appears’, or ‘looks’.[20] This typically happens in arguments from illusion. The intuitive attributions of non-awareness that are then generated by stereotype-driven inferences from those verbs, followed by contextual integration, are cognitive illusions. The finding that an intuition is generated in this way, under these circumstances, actually speaks against its truth.

4.2 An explanation of phenomenal judgments

Protest against the negative intuitions thus exposed as epistemologically worthless prompts phenomenal intuitions (Section 1, pp. 260–265): ‘The subject is not aware of the round coin that looks elliptical to her. But she is aware of *something*!’ (cf., e.g., Broad 1923, 240). In intuitive lines of thought, and before introducing technical terms, this ‘something’ is typically characterized as an ‘elliptical colour patch’ or ‘speck of colour’ (e.g., Ayer 1940, 22ff.; Price 1932, 3), and this characterization is spontaneously interpreted as expressing the phenomenal judgment that the viewer is aware of something (namely, a colour patch) that is elliptical—in the same ‘plain sense’ (Broad 1923, 240) in which that term applies to physical objects.

When talking about familiar cases of non-veridical perception (like perspective) we ordinarily use those expressions

(A) when we cannot tell what we are looking at, so that we can only pick it out by a description of its looks (from here, now), as when a rambler points into a valley and asks, ‘Do you see that small red patch? Might that be our car?’

(B) when we seek to convey economically how something looks for a particular protagonist, as in this passage from a novel: ‘Who was the person at the bottom of the pool? Morini could see him or her, a tiny speck trying to climb the rock that had now become a mountain, and the sight of this person, so far away, filled his eyes with tears’ (Bolaño 2009, 47).

In neither kind of case does our use of “*F* patch” or “*F* speck” imply that anything actually is the shape, size, or colour *F*: The small red patch may turn out to *be* our car, and “tiny speck” refers to a climber (whether man or woman the protagonist cannot discern). There is no suggestion that the thing (vehicle, person) we talk about *is* small or tiny (a Mini or a dwarf), merely that it *looks* small (looks the size of a small patch or tiny speck), *from here, now*.

This exemplifies a familiar metaphorical usage which has us describe things in terms of others that look similar, in a particular respect:[21] Instead of saying, ‘The person disguised as a ghost turned out to be the host of the fancy-dress party’ we

[20] While in the context of the problem of perception it is philosophically contentious what subjects are aware of, people’s inclinations to judge are uncontroversial, and the present claim can be easily confirmed experimentally, if contested. We thus avoid the ‘calibration problem’ (Leben 2014).

[21] Gibbs and Colston (2012, 48–54) provide a pertinent classification of metaphorical uses.

can say more pithily, 'The ghost turned out to be the host' (without expressing a belief in supernatural entities). In stating the argument from illusion, this usage lets us say that the viewer is aware of an elliptical silvery speck: This means that the viewer is aware of something that looks in some ways like an elliptical silvery speck, namely looks elliptical and silvery in shape and colour.

This metaphorical interpretation of 'patch talk' in or about familiar situations of non-veridical perception—like (A) and (B) above—is dictated by Grice's (1989) Maxim of Manner: 'Be as clear, precise, and brief as you can!' When we already know what we are looking at, this maxim obliges us to call a spade a spade. Hence we may only resort to the patch-idiom in talking of what we see when we cannot tell what exactly it is that we are looking at—case (A). Similarly, we may use it in talk of what others see only when we don't know what they look at, or want to avoid the suggestion that they know what they are looking at or for other reasons want to convey not what they are looking at but what that thing looks like to them—case (B). When taking authors to respect the maxim, readers hence infer from authors' preference of "elliptical speck" over "round coin" that they do not wish to suggest that the viewer knows what she views, but mean to convey that this thing looks elliptical to her.

Pragmatic maxims can be defeated, e.g., by norms of politeness or stylistic conventions. But no such defeaters seem relevant in statements of the argument from illusion. These statements tell us explicitly what object (coin, etc.) is viewed. As competent speakers, both readers and authors of these arguments should spontaneously interpret their talk of, say, an 'elliptical patch' as referring to the round coin the subject is explicitly assumed to look at, and as conveying how that—round—object looks to her (there and then). The phenomenal judgment that the subject then is aware of something that *is* elliptical should hence strike them as jarring, and any further inference (with Leibniz's Law) to 'The subject is aware of something other than the round coin' should stike them as every bit as poor a joke as a fellow rambler's response to (A): 'That small patch couldn't possibly be my car. I wouldn't ever drive anything smaller than a Bentley.' Both turn on the literal interpretation of a familiar metaphorical usage clearly relevant in the situations at issue.

In interpreting and accepting the likes of "The viewer is aware of an elliptical patch" as expressions of phenomenal judgments, champions of arguments from illusion are subject to a *semantic illusion* (review: Park and Reder 2004) akin to the famous 'Moses illusion' (Erickson and Mattson 1981): When asked, 'How many animals of each kind did Moses take onto the ark?', 81 per cent of participants who subsequently demonstrated knowledge of the ark story responded 'two', after correctly reading aloud the question and having been instructed to either answer questions *or* indicate that something is wrong with them, as appropriate. Given their biblical knowledge, these participants should have interpreted "Moses" as referring to a figure outside the ark story and rejected the question—just as readers of the argument from illusion, who have been told that the viewer looks at a round coin sideways, should have interpreted "elliptical patch" as referring to the coin and rejected the further inference with Leibniz's Law. Rather, participants

Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

spontaneously interpret the question about Moses as one about Noah, and respond accordingly—just as proponents of 'our' argument spontaneously interpret "elliptical patch" as referring to something elliptical and go along with the further inference.

The most widely accepted explanation of such semantic illusions builds on the standard conception of semantic memory (Neely 1991; Kahneman 2011) (see Section 2.1, pp. 265–268): Concept-nodes standing for things are linked to property-nodes representing their properties. When activated, a 'thing-node' passes on activation to such 'property-nodes' which, in turn, pass it on to other 'thing-nodes' they are linked to. Activation thus spreads from one 'thing-node' to nodes that represent other things which are believed to share the same properties. The more properties two things are believed to share, i.e., the more 'semantically similar' their concepts are (Oostendorp and Mul 1990, 36–37; see above, Section 2.3, pp. 271–274), the more activation is passed on from one 'thing-node' to the other. Fully automatic associative processing which exploits these features duplicates inferences with a 'simple heuristic' (Park and Reder 2004, 289) in line with Grice's (1989) Maxim of Relation or Relevance ('Make your response relevant to the subject of discussion!'). Think of the *partial match heuristic* as a straightforward search-and-match rule for determining reference (see Barton and Sandford 1993; Kamas *et al.* 1996; Park and Reder 2004): 'Pick the object of the relevant domain (e.g. protagonists of the ark story) whose concept is semantically most similar to the stimulus concept, if the similarity exceeds a threshold; otherwise, assume the expression refers to somebody or something outside the domain (no protagonist of the ark story—wrong question).'[22]

This can explain why we spontaneously give a metaphorical interpretation to 'patch talk' about familiar situations of non-veridical perception, but philosophers are literal-minded in the specific context of arguments from illusion: When in contextually embedded discussion of things seen in the environment, we are told that 'when people view a round coin sideways, it looks elliptical, and they are aware of an elliptical patch', the partial match heuristic ensures that we interpret "elliptical patch"—metaphorically—as referring to the most similar object subjects are said to view in the environment: the round coin which then looks similar to an elliptical patch, namely elliptical. But instead, once philosophers discussing objects of awareness have leaped from the sparse initial premise of an argument from illusion to the intuitive conclusion that the viewer is not aware of, say, the coin she looks at (Section 2.3, pp. 271–274), the coin has been removed from the relevant domain of discourse (the set of objects of awareness talked about). This domain then no longer contains anything sufficiently similar to elliptical patches (the premise does not mention any other objects), and the heuristic rule has us posit an object outside the domain (an object of awareness not introduced by the premises, distinct from the coin) which satisfies the relevant description ("elliptical patch") on its dominant—literal—reading.

[22] This simplifies a heuristic for determining utterance content. See Budiu and Anderson (2004).

This explanation supports our initial hypothesis (Section 1, pp. 260–265) that negative intuitions precede phenomenal intuitions in the intuitive lines of thought that inform arguments from illusions. Crucially, it also helps us assess the evidentiary value of the subjective confidence attaching to phenomenal intuitions which proponents of these arguments confidently declared to be 'obvious'.

4.3 The evidentiary value of subjective confidence

Intuitive judgments are attended by high degrees of subjective confidence (Section 1, pp. 260–265), (misleadingly) known as 'feelings of rightness'.[23] According to the now dominant 'experience-based approach' to metacognitive judgments (review: Koriat 2007), this subjective confidence does not result from deliberate reflection on the content of the judgment or of further information retrieved from memory; rather, it is immediately based on features of the process that issues in the judgment, which serve as mnemonic cues in automatic cognition. The most important of these cues is *answer fluency*, i.e., 'the ease with which this conclusion of an automatic inference comes to mind' (Thompson *et al.* 2011, 111; see also Simmons and Nelson 2006). Such fluency is frequently taken to be a function of the degree of automaticity and is mostly measured by response times in intuitive judgment tasks (Van Overschelde 2008) controlling for several other sources of (dys-)fluency (comprehensively reviewed by Alter and Oppenheimer 2009).[24] Confidence in a judgment increases with the speed with which that judgment is made (Kelley and Lindsay 1993, Robinson *et al.* 1997; Thompson *et al.* 2011) as well as with the subjective impression of effortlessness (Alter *et al.* 2007). We can hence obtain an assessment of the evidentiary value of 'feelings of rightness' attaching to intuitive judgments by combining these findings about metacognitive cues with an explanation of the intuitive judgments at issue which uncovers the relevant sources of fluency and explains why those judgments are made particularly swiftly or with particular ease: If this is due to factors which render it, respectively, more or less likely that the judgments at issue are true, their fluency does, or does not, speak for their truth, respectively.

Our explanation of the phenomenal judgments involved in arguments from illusion suggests a straightforward reason why, in this context, they are more fluent than metaphorical interpretations of the same 'patch talk': If fluency is a function of degree of automaticity, the most fluent interpretation is the one in line with the partial match heuristic that is duplicated by fully automatic association processes. Given the prior negative intuition that the viewer is unaware of (say) the coin, this yields a highly fluent literal interpretation. By contrast, metaphorical interpretation then requires rejection of this prior intuition.

[23] Psychologists measure them through rating scales that ask participants whether in offering their judgment they 'felt guessing', 'fairly certain', or 'certain I'm right' (e.g., Thompson *et al.* 2011, 114).

[24] Answer fluency is a frugal cue for more complex mnemonic cues such as self-consistency, with dependant validity (Koriat 2012). We therefore restrict our discussion to answer fluency.

Rejection of intuitions is time-consuming and effortful (De Neys 2006; Kahneman and Frederick 2005). Where the initial premises of arguments from illusion trigger such negative judgments and these, in turn, prompt protests to the effect that 'we are aware of something', namely, of '*F* patches', their metaphorical interpretation is bound to be significantly less fluent than the literal interpretation that yields phenomenal judgments.

The higher fluency of phenomenal judgments is hence due to a prior negative intuition—without which the partial match heuristic would yield a metaphorical interpretation. Proponents of the argument from illusion have no warrant to accept this negative intuition (Section 4.1, pp. 281–282). When high fluency is due to interference by such an unwarranted intuition, it does not speak for the truth of the fluent judgment. The subjective confidence it engenders then has no evidentiary value. The fact that proponents of arguments from illusion confidently regard their phenomenal judgments as 'obvious' has no evidentiary value.

Conflicts between intuitive judgments and normatively correct alternatives or background beliefs lead to lower subjective confidence (De Neys *et al.* 2011) and increased reflective scrutiny (De Neys and Glumicic 2008). While the latter may well lead to *ex post* rationalization, rather than correction, of the intuition (Shynkaruk and Thompson 2006; Stanovich 2009; Wilson and Dunn 2004), the high subjective plausibility of phenomenal judgments in the face of conflict with the fact that nothing that *is* (say) elliptical is around to be seen in the scenarios that prompt them demands further explanation. This high confidence would seem to indicate that the conflict between the intuition and this fact is either not salient or regarded as merely apparent. Arguably, thinkers who find phenomenal judgments subjectively plausible take for granted the existence of a 'mind' as a complementary space of perception which can house all objects of awareness that are not around to be perceived in the environment, including those posited by phenomenal judgments. Elsewhere (Fischer 2014, 2015), one of us has developed a psychological explanation of intuitions which jointly posit such 'minds' in us and are typically accepted implicitly, without argument (Fischer 2011). This explanation suggests these intuitions too lack evidentiary value. A fuller explanation of the subjective confidence that attaches to phenomenal intuitions in the context of arguments from illusion is therefore unlikely to alter the present conclusion: It confers no evidentiary value on these intuitions.[25]

[25] Paul E. Engelhardt contributed to design and material development for the experiment reported in Section 3 (pp. 274–280), and undertook the statistical analysis. Aurélie Herbelot contributed the distributional semantics analysis reported in Section 2.2 (pp. 268–271) and the Appendix. Eugen Fischer undertook the remaining research. For helpful comments on previous drafts and closely related material, we are indebted to Rachel Giora (who first suggested the use of the forced-choice plausibility-ranking paradigm) as well as Yuri Cath, John Collins, Adam Leite, Kevin Reuter, Jonathan Weinberg, and audiences in Dublin (2013), Dortmund, Madrid, and Cambridge (all 2014).

Appendix—distributional semantics analysis of appearance-verbs

Aurélie Herbelot of the Cambridge Computer Laboratory compared the verbs “seem(to)”, “appear(to)”, and “look(to)” with the 3,000 most frequent verbs in the Wikiwoods corpus (Flickinger *et al.* 2010) for distributional similarity. The methodology is explained above (Section 2.2, pp. 268–271). The results below indicate how distributionally similar their ‘nearest neighbours’ were found to be to the three target words. The values indicate the distributional similarity between the verbs compared. The highest possible value of ‘1’ means that, in our corpus, the verbs compared co-occur with exactly the same other words as arguments, in exactly the same proportions—a word obtains this value only when compared with itself. While such similarity-values will differ depending on the corpus, the respective positions of the neighbours with regard to each target word can be expected to be replicable over any large corpus of ordinary English.

‘seem(to)’: (‘appear(to)’, 0.6549), (‘think(1)’, 0.5021), (‘believe(1)’, 0.4974), (‘be(nv)’, 0.4792), (‘find(mental)’, 0.4673), (‘say(to)’, 0.4471), (‘mean(1)’, 0.3890), (‘make(cause)’, 0.3886), (‘claim(1)’, 0.3743), (‘suggest(to)’, 0.3696), (‘argue(with)’, 0.3661), (‘state(to)’, 0.3598), (‘say(1)’, 0.3579), (‘prove(to)’, 0.3534), (‘note(to)’, 0.3518), (‘be(itcleft)’, 0.3412), (‘indicate(1)’, 0.3382), (‘feel(1)’, 0.3295), (‘know(1)’, 0.3264), (‘turn(out)’, 0.3224), (‘realize(1)’, 0.3209), (‘expect(1)’, 0.3134), (‘imply(1)’, 0.3044), (‘intend(for)’, 0.2974), (‘agree(with)’, 0.2928), (‘conclude(1)’, 0.2848), (‘show(1)’, 0.2843), (‘want(1)’, 0.2793), (‘reveal(to)’, 0.2726)

‘appear(to)’: (‘seem(to)’, 0.6549), (‘believe(1)’, 0.5181), (‘think(1)’, 0.5015), (‘be(nv)’, 0.4838), (‘find(mental)’, 0.4808), (‘say(to)’, 0.4490), (‘make(cause)’, 0.3981), (‘claim(1)’, 0.3930), (‘mean(1)’, 0.3877), (‘say(1)’, 0.3749), (‘suggest(to)’, 0.3749), (‘state(to)’, 0.3671), (‘argue(with)’, 0.3552), (‘indicate(1)’, 0.3504), (‘prove(to)’, 0.3499), (‘note(to)’, 0.3486), (‘turn(out)’, 0.3374), (‘be(itcleft)’, 0.3319), (‘know(1)’, 0.3261), (‘intend(for)’, 0.3259), (‘expect(1)’, 0.3215), (‘realize(1)’, 0.3163), (‘feel(1)’, 0.3076), (‘imply(1)’, 0.3066), (‘take(2)’, 0.2990), (‘show(1)’, 0.2986), (‘agree(with)’, 0.2955), (‘conclude(1)’, 0.2891), (‘reveal(to)’, 0.2866)

‘look(seem-to)’: (‘sound(seem-to)’, 0.3468), (‘make(cause)’, 0.3245), (‘appear(to)’, 0.2768), (‘act(seem+to)’, 0.2701), (‘find(mental)’, 0.2648), (‘grow(to)’, 0.2632), (‘seem(to)’, 0.2599), (‘feel(seem-about)’, 0.2412), (‘become(id)’, 0.1999), (‘go(state)’, 0.1695), (‘remain(1)’, 0.1694), (‘think(1)’, 0.1650), (‘make(i)’, 0.1577), (‘get(state)’, 0.1558), (‘run(prd)’, 0.1499), (‘prove(to)’, 0.1476), (‘say(to)’, 0.1405), (‘turn(out)’, 0.1286), (‘believe(1)’, 0.1278), (‘stand(1)’, 0.1211), (‘turn(prd)’, 0.1179), (‘realize(1)’, 0.1164), (‘fall(state)’, 0.1159), (‘look(like)’, 0.1142), (‘stay(prd)’, 0.1118), (‘feel(1)’, 0.1032), (‘show(1)’, 0.1030), (‘indicate(1)’, 0.1026), (‘mean(1)’, 0.1016)

References

Allport, D. A. 1985. Distributed memory, modular subsystems and dysphasia. In S. K. Newman and R. Epstein (eds.), *Current Perspectives in Dysphasia* (pp. 207–244). Edinburgh: Churchill Livingstone.

Alter, A. L. and Oppenheimer, D. M. 2009. Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13: 219–235.

Alter, A. L., Oppenheimer, D. M., Epley, N. and Eyre, R. N. 2007. Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136: 569–576.

Austin, J. L. 1962. *Sense and Sensibilia*. Oxford: Oxford University Press.

Ayer, A. J. 1940. *Foundations of Empirical Knowledge*. London: Macmillan.

——1956. *The Problem of Knowledge*. Repr. 1990. London: Penguin.

Baldwin, T., Bender, E. M., Flickinger, D., Kim, A. and Oepen, S. 2004. Road-testing the English Resource Grammar over the British National Corpus. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa and R. Silva (eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC2004)* (pp. 2047–2050). Paris: European Language Resources Association.

Bargh, J. A. 1994. The four horsemen of automaticity. In R. Wyer and T. Srull (eds.), *Handbook of Social Cognition*, vol. 1 (pp. 1–40). Hillsdale: Erlbaum.

Barnhardt, T. M., Glistky, E. L., Polster, M. R. and Elam, L. 1996. Inhibition of associates and activation of synonyms in the rare-word paradigm. *Memory and Cognition*, 24: 60–69.

Barton, S. B. and Sandford, A. J. 1993. A case study of anomaly detection: shallow semantic processing and cohesion establishment. *Memory and Cognition*, 21: 477–487.

Bolaño, R. 2009. *2666: A Novel*. London: Picador.

Brewer, B. 2011. *Perception and Its Objects*. Oxford: Oxford University Press.

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.

——2014. The phenomenal use of 'look' and perceptual representation. *Philosophy Compass*, 9(7): 455–468.

Brown, P. M. and Dell, G. S. 1987. Adapting production to comprehension: the explicit mention of instruments. *Cognitive Psychology*, 19: 441–472.

Budiu, R. and Anderson, J. R. 2004. Interpretation-based processing: a unified theory of semantic sentence comprehension. *Cognitive Science*, 28: 1–44.

Chisholm, R. 1957. *Perceiving*. Ithaca: Cornell University Press.

Copestake, A. 2009. Slacker semantics: why superficiality, dependency and avoidance of commitment can be the right way to go. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL09)*, Athens, Greece (pp. 1–9). Stroudsburg, PA: Association of Computational Linguistics.

Crane, T. 2011. The problem of perception. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2011 edition), <http://plato.stanford.edu/entries/perception-problem/>.

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. and Glumicic, T. 2008. Conflict monitoring in dual process theories of thinking. *Cognition*, 106: 1248–1299.

De Neys, W., Cromheeke, S. and Osman, M. 2011. Biased but in doubt: conflict and decision confidence. *PLoS ONE*, 6 (1): e15954, < http://www.plosone.org/article/info:doi/10.1371/journal.pone.0015954>

Dijksterhuis, A. 2010. Automaticity and the unconscious. In S. T. Fiske, D. T. Gilbert and G. Lindzey (eds.), *Handbook of Social Psychology*, 5th ed. (pp. 228–267). Hoboken: Wiley.

Dinu, G., Pham, N. and Baroni, M. 2013. DISSECT: DIStributional SEmantics Composition Toolkit. In M. Butt and S. Hussain (eds.), *Proceedings of the System Demonstrations of the 51st Annual Meeting of the Association for Computational Linguistics (ACL2013)* (pp. 31–36). Stroudsburg, PA: Association for Computational Linguistics.

Earlenbaugh, J. and Molyneux, B. 2009. Intuitions are inclinations to believe. *Philosophical Studies*, 145: 89–109.

Erickson, T. and Mattson, M. 1981. From words to meaning: a semantic illusion. *Journal of Verbal Learning and Verbal Behaviour*, 20: 540–551.

Erk, K. 2012. Vector space models of word meaning and phrase meaning: a survey. *Language and Linguistics Compass*, 6: 635–653.

Evans, J. S. B. T. 2010. Intuition and reasoning: a dual-process perspective. *Psychological Inquiry*, 21: 313–326.

Faust, M. and Gernsbacher, M. A. 1996. Cerebral mechanisms for suppression of inappropriate information during sentence comprehension. *Brain and Language*, 53: 234–259.

Ferretti, T., McRae, K. and Hatherell, A. 2001. Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44: 516–547.

Fiala, B., Arico, A. and Nichols, S. 2011. On the psychological origins of dualism: dual-process cognition and the explanatory gap. In E. Slingerland and M. Collard (eds.), *Creating Consilience* (pp. 88–110). New York: Oxford University Press.

Fischer, E. 2011. *Philosophical Delusion and Its Therapy*. New York: Routledge.

——2014. Philosophical intuitions, heuristics, and metaphors. *Synthese*, 191: 569–606.

——2015. Mind the metaphor! A systematic fallacy in analogical reasoning. *Analysis*, 75: 67–77.

Fischer, E. and Collins, J. 2015 [this volume]. Rationalism and naturalism in the age of experimental philosophy. In E. Fischer and J. Collins (eds.), *Experimental Philosophy, Rationalism, and Naturalism* (pp. 3–33). London: Routledge.

Fischer, E. and Engelhardt, P. E. Forthcoming. Intuitions' linguistic sources: stereotypes, intuitions, and illusions. *Mind and Language*.

Fischer, E., Engelhardt, P. E. and Herbelot, A. In prep. Experimental philosophy of appearances.

Fish, W. 2009. *Perception, Hallucination, and Illusion*. Oxford: Oxford University Press.

Flickinger, D., Oepen, S. and Ytrestol, G. 2010. Wikiwoods: syntacto-semantic annotation for the English Wikipedia. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, 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.

Garrett, M. and Harnish, R. M. 2007. Experimental pragmatics: testing for implicatures. *Pragmatics and Cognition*, 15: 65–90.

Gernsbacher, M. A. and Faust, M. E. 1991. The mechanism of suppression: a component of general comprehension skill. *Journal of Experimental Psychology: Learning Memory and Cognition*, 17: 245–262.

Gibbs, R. W. and Colston, H. L. 2012. *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.

Giora, R. 2003. *On Our Mind: Salience, Context, and Figurative Language*. Oxford: Oxford University Press.

Givoni, S., Giora, R. and Bergerbest, D. 2013. How speakers alert addressees to multiple meanings. *Journal of Pragmatics*, 48: 29–40.

Grice, H. P. 1961. The causal theory of perception. *Proceedings of the Aristotelian Society Supplementary Volume*, 35: 121–52.

——1989. Logic and conversation. In *Studies in the Ways of Words* (pp. 22–40). Cambridge, MA: Harvard University Press.

Hare, M., Jones, M., Thomson, C., Kelly, S. and McRae, K. 2009. Activating event knowledge. *Cognition*, 111: 151–167.

Harmon-Vukić, M., Guéraud, S., Lassonde, K. A. and O'Brien, E. J. 2009. The activation and instantiation of instrumental inferences. *Discourse Processes*, 46: 467–490.

Henderson, D. K. and Horgan, T. 2011. *The Epistemological Spectrum: At the Interface of Cognitive Science and Conceptual Analysis*. Oxford: Oxford University Press.

Hume, D, 1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge and P. H. Nidditch. Oxford: Clarendon.

Jackson, F. 1977. *Perception: A Representative Theory*. Cambridge: Cambridge University Press.

Kahneman, D. 2011. *Thinking Fast and Slow*. London: Allen Lane.

Kahneman, D. and Frederick, S. 2002. Representativeness revisited: attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahneman (eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49–81). New York: Cambridge University Press.

——2005. A model of heuristic judgment. In K. J. Holyoak and R. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267–293). Cambridge: Cambridge University Press.

Kamas, E. N., Reder, L. M. and Ayers, M. S. 1996. Partial matching in the Moses illusion: response bias not sensitivity. *Memory and Cognition*, 24: 687–699.

Kelley, C. M. and Lindsay, D. S. 1993. Remembering mistaken for knowing: ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32: 1–24.

Klin, C. M., Guzman, A. E. and Levine, W. H. 1999. Prevalence and persistence of predictive inferences. *Journal of Memory and Language*, 40: 593–604.

Knobe, J. and Nichols, S. 2008. An experimental philosophy manifesto. In K. Knobe and S. Nichols (eds.), *Experimental Philosophy* (pp. 3–14). Oxford: Oxford University Press.

Koriat, A. 2007. Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch and E. Thompson (eds.), *The Cambridge Handbook of Consciousness* (pp. 289–326). Cambridge: Cambridge University Press.

——2012. The self-consistency model of subjective confidence. *Psychological Review*, 119: 80–113.

Kornblith, H. 2015 [this volume]. Naturalistic defenses of intuition. In E. Fischer and J. Collins (eds.), *Experimental Philosophy*, *Rationalism, and Naturalism* (pp. 151–168). London: Routledge.

Kutas, M. and Federmeier, K. T. 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4: 463–460.

——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.

Leben, D. 2014. When psychology undermines beliefs. *Philosophical Psychology*, 27: 328–350.

Leech, G., Payson, P. and Wilson, A. 2001. *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Longman.

Template: Royal A, Font: ,
Date: 05/03/2015; 3B2 version: 10.0.1465/W Unicode (Dec 22 2011) (APS_OT)
Dir: //integrafs1/kcg/2-Pagination/TandF/EXPI_COR/ApplicationFiles/9781138887251_text.3d

Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levinson, S. C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, Cambridge, MA: MIT Press.

Lowrey, T. M. 1998. The effects of syntactic complexity on advertising persuasiveness. *Journal of Consumer Psychology*, 7: 187–206.

Lucas, M. 2000. Semantic priming without association: a meta-analytic review. *Psychonomic Bulletin and Review*, 7: 618–630.

Martin, M. G. F. 2003. Sensible appearances. In T. Baldwin (ed.), *The Cambridge History of Philosophy, 1870–1945* (pp. 521–523). Cambridge: Cambridge University Press.

Maund, J. B. 1986. The phenomenal and other uses of 'looks'. *Australasian Journal of Philosophy*, 64: 170–180.

McRae, K. and Jones, M. 2013. Semantic memory. In D. Reisberg (ed.), *Oxford Handbook of Cognitive Psychology* (pp. 206–219). Oxford: Oxford University Press.

Montague, R. 1974. The proper treatment of quantification in ordinary English. In R. Thomason (ed.), *Formal Philosophy* (pp. 247–270). New Haven, CT: Yale University Press.

Moore, G. E. 1918–19. Some judgments of perception. *Proceedings of the Aristotelian Society*, 19: 1–29.

Nagel, J. 2010. Knowledge ascriptions and the psychological consequences of thinking about error. *Philosophical Quarterly*, 60: 286–306.

——2011. The psychological basis of the Harman-Vogel paradox. *Philosophers' Imprint*, 11 (5): 1–28.

——2012. Intuitions and experiments: a defence of the case method in epistemology. *Philosophy and Phenomenological Research*, 85: 495–527.

Nahmias, E. and Murray, D. 2010. Experimental philosophy on free will. In J. Aguilar, A. Buckareff and K. Frankish (eds.), *New Waves in Philosophy of Action* (pp. 189–216). Basingstoke: Palgrave.

Neely, J. H. 1991. Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner and G. Humphreys (eds.), *Basic Processes in Reading: Visual Word Recognition* (pp. 264–336). Hillsdale: Erlbaum.

Neely, J. H. and Kahan, T. A. 2001. Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger, J. S. Nairne, I. Neath and A. M. Surprenant (eds.), *The Nature of Remembering* (pp. 69–93). Washington, DC: America Psychological Association.

Nichols, S. and Knobe, J. 2007. Moral responsibility and determinism. *Noûs*, 41: 663–685.

Oostendorp, H. van and Mul, S. de 1990. Moses beats Adam: a semantic relatedness effect on a semantic illusion. *Acta Psychologica*, 74: 35–46.

Oppenheimer, D. M. 2006. Consequences of erudite vernacular utilised irrespective of necessity: problems with using long words needlessly. *Applied Cognitive Psychology*, 20: 139–156.

Papineau, D. 2015 [this volume]. The nature of a priori intuitions: analytic or synthetic? In E. Fischer and J. Collins (eds.), *Experimental Philosophy*, *Rationalism, and Naturalism* (pp. 51–71). London: Routledge.

Park, H. and Reder, L. M. 2004. Moses illusion. In R. Pohl (ed.), *Cognitive Illusions* (pp. 275–291). New York: Psychology Press.

Patson, N. D. and Warren, T. 2010. Eye movements to plausibility violations. *Quarterly Journal of Experimental Psychology*, 63, 1516–1532.

Peleg, O. and Giora, R. 2011. Salient meanings: the whens and wheres. In K. M. Jaszczolt and K. Allan (eds.), *Salience and Defaults in Utterance Processing* (pp. 32–52). Berlin: de Gruyter.

Pickering, M. J. and Garrod, S. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36: 329–392.

Pohl, R. (ed.) 2004. *Cognitive Illusions*. New York: Psychology Press.

Pollock, J. 1984. Reliability and justified belief. *Canadian Journal of Philosophy*, 14: 103–114.

Postal, P. 1973. *On Raising*. Cambridge, MA: MIT Press.

Price, H. H. 1932. *Perception*, 2nd ed. Repr. 1961. London: Methuen.

Pust, J. 2012. Intuition. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012 edition), <http://plato.stanford.edu/archives/win2012/entries/intuition/>.

Robinson, H. 2001. *Perception*. London: Routledge.

Robinson, M. D., Johnson, J. T. and Herndon, F. 1997. Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology*, 82: 416–425.

Rosch, E. 1978. Principles of categorisation. In E. Rosch and B. Lloyd (eds.), *Cognition and Categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Russell, B. 1912/1980. *The Problems of Philosophy*. Oxford: Oxford University Press.

Shynkaruk, J. M. and Thompson, V. A. 2006. Confidence and accuracy in deductive reasoning. *Memory and Cognition*, 34: 619–632.

Simmons, J. P. and Nelson, L. D. 2006. Intuitive confidence: choosing between intuitive and non-intuitive alternatives. *Journal of Experimental Psychology: General*, 135: 409–428.

Simpson, G. B. and 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 University Press.

Sosa, E. 2007. Intuitions: their nature and epistemic efficacy. *Grazer Philosophische Studien* 74: 51–67.

Stanovich, K. E. 2009. Distinguishing the reflective, algorithmic, and autonomous minds: is it time for a tri-process theory? In J. Evans and K. Frankish (eds.), *In Two Minds: Dual Processes and Beyond* (pp. 55–88). Oxford: Oxford University Press.

Stephens, G. J., Silbert, L. J. and Hasson, U. 2010. Speaker–listener neural coupling underlies successful communication. *PNAS*, 107: 14425–14430.

Thompson, V. A., Prowse Turner, J. A. and Pennycook, G. 2011. Intuition, reason, and metacognition. *Cognitive Psychology*, 63: 107–140.

Till, R. E., Mross, E. F. and 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.

Tucker, C. 2013. Seeming and justification: an introduction. In C. Tucker (ed.), *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism* (pp. 1–30). Oxford: Oxford University Press.

Tulving, E. 2002. Episodic memory: from mind to brain. *Annual Review of Psychology*, 53: 1–25.

Turney, P. D. and Pantel, P. 2010. From frequency to meaning: vector space models of semantics. *Journal of Artificial Intelligence Research*, 37: 141–188.

Van Overschelde, J. P. 2008. Metacognition: knowing about knowing. In J. Dunlosky and R. A. Bjork (eds.), *Handbook of Metamemory and Memory*. New York: Psychology Press.

Weinberg, J. 2015 [this volume]. Humans as instruments: or, the inevitability of experimental philosophy. In E. Fischer and J. Collins (eds.), *Experimental Philosophy*, *Rationalism, and Naturalism* (pp. 171–187). London: Routledge.

Williams, J. N. 1992. Processing polysemous words in context: evidence from interrelated meanings. *Journal of Psycholinguistic Research*, 21: 193–218.

Williams, M. 2001. *Problems of Knowledge*. Oxford: Oxford University Press.

Wilson, T. D. and Dunn, E. W. 2004. Self-knowledge: its limits, value, and potential for improvement. *Annual Review of Psychology*, 55: 493–518.