**Intuitions’ Linguistic Sources:**

**Stereotypes, Intuitions, and Illusions**

Eugen Fischer and Paul E. Engelhardt

Final draft version of paper accepted (28/01/2015) by *Mind and Language*

Intuitive judgments elicited by verbal case-descriptions play key roles in philosophical problem-setting and argument. Experimental philosophy’s ‘sources project’ seeks to develop psychological explanations of philosophically relevant intuitions which help us assess our warrant for accepting them. This paper develops a psycholinguistic explanation of intuitions prompted by philosophical case-descriptions. For proof of concept, we target intuitions underlying a classic paradox about perception (‘argument from illusion’), trace them to stereotype-driven inferences automatically executed in verb comprehension, and employ a forced-choice plausibility-ranking task to elicit the relevant stereotypical associations of perception- and appearance-verbs. We obtain a debunking explanation which resolves the philosophical paradox.

The ‘warrant project’in experimental philosophy seeks to determine what warrant thinkers have for accepting philosophically relevant intuitions. The arguably most ambitious strand of this project, variously known as *‘the sources project’* (Pust 2012) or *‘cognitive epistemology’* (Fischer 2014), seeks to do so by developing experimentally testable psychological explanations of relevant intuitions which help us assess the *evidentiary value* of these intuitions, i.e., help us assess whether the mere fact that given thinkers have these intuitions, as and when they do, speaks for the intuitions’ truth. This project is not restricted to card-carrying members of the experimental philosophy movement. It also falls, e.g., into the remit of ‘iceberg epistemology’, whose basic ‘idea is that the conscious and consciously accessible aspects of belief formation are just the tip of a larger iceberg of epistemologically relevant psychological factors’ (Henderson and Horgan 2011, 196); these crucially include cognitive processes that are automatic and go on ‘below the waterline’ of conscious awareness. One promising approach is to look for *‘GRECI explanations’* (as we will call them) that trace intuitions back to largely automatic cognitive processes which are *generally reliable* but predictably engender *cognitive illusions*, under specific circumstances (e.g. Nagel 2010, 2012, Fischer 2014, Fischer et al 2015). The immediate aim is to provide either validating or debunking explanations of specific intuitions that are philosophically relevant. The ultimate goal is to develop ‘epistemic profiles’ of the underlying cognitive processes, which set out under what circumstances we may trust their deliverances, and when and where we ought to beware (cp. Weinberg 2015).

Extant contributions to the sources project seek to explain targeted intuitions by drawing on concepts and findings from cognitive or social psychology (e.g. Fiala et al. 2011, Nagel 2011, 2012, Fischer 2011, 2014, Nahmias and Murray 2010). The present paper will break new ground by building, instead, on findings from psycholinguistics (section 1), to develop an experimentally testable GRECI explanation of intuitions that are elicited by verbal case-descriptions which are used in philosophical argument (sections 2-3). Then we will introduce and apply an experimental paradigm from psycholinguistics, to experimentally test the key hypotheses the explanation builds on (section 4). Thus confirmed, the explanation will debunk intuitions at the root of notorious paradoxical arguments.

1. **Linguistic Sources of Philosophical Intuitions**

In the aetiological sense dominant in cognitive psychology and required for the purposes of the sources project, intuitions are deliverances of predominantly automatic cognitive processes.[[1]](#footnote-1) According to a common definition (Bargh 1994, Moors and De Houwer 2006), cognitive processes are *automatic*, rather than ‘controlled’ to the extent to which they are effortless (the key property) as well as unconscious, non-intentional, and autonomous. These properties are gradable and operationally defined. E.g., a process is *effortless* to the extent to which it is independent from working memory, and thus requires no attention or other limited cognitive resource, so that performance is not impaired by multi-tasking (simultaneously keeping in mind long numbers, etc.).[[2]](#footnote-2) Automatic cognitive processes generate also, e.g., perceptual and memory judgments. We therefore define, more specifically:

*Intuitions* are judgments which are based on *‘automatic inferences’* (Kahneman and Frederick 2005, 268, cp. Evans 2010, 314), i.e., on largely automatic cognitive processes which duplicate inferences governed by normative or heuristic rules.[[3]](#footnote-3),[[4]](#footnote-4)

Since effortlessness serves as a meta-cognitive cue for plausibility, the intuitive judgments delivered by effortless processes tend to strike thinkers as plausible, regardless of whether or not they go on to reflectively endorse them (Alter and Oppenheimer 2009, Thompson et al. 2011). Where they are not so endorsed, it is tempting to characterise them as ‘inclinations to assent’ (cp. Sosa 2007 and Earlenbaugh and Molyneux 2009).

Most intuitions that shape debates in analytic philosophy are elicited by verbal descriptions of hypothetical cases (such as Gettier or trolley cases). They may be used as evidence for or against philosophical theories (of knowledge, etc.) or as starting-points of philosophical arguments. Psycholinguistic research 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 sense defined. Such processes routinely go on in comprehension and production of text and speech (see below). We should therefore explore whether some of the intuitions philosophers have about cases when stating or reading their verbal descriptions are generated by these routine comprehension-related processes. A good starting-point is provided by the simplest case-descriptions, viz. single-sentence descriptions that serve as initial premises of philosophical arguments and prompt intuitions that drive those arguments.

Where GRECI explanations trace intuitions of the relevant kind back to a generally reliable process that predictably goes astray only under specific circumstances, debunking explanations of particular intuitions of this kind ‘only’ have to show, in addition, that these particular intuitions arise under such vitiating circumstances. Validating explanations, by contrast, need to identify and exclude a potentially wide variety of such circumstances, to validate particular intuitions.[[5]](#footnote-5) When making first steps towards the kind of explanation the sources project seeks, we should therefore aim for simpler debunking explanations. Philosophically, such explanations are particularly productive where they debunk intuitions involved in paradoxical arguments that engender philosophical problems: in philosophical paradoxes which we can seek to resolve by showing that we lack warrant for accepting (some of) the intuitions they rely on.

We will now review key psycholinguistic findings about associative processes in semantic memory (section 1.1), to explore how they can contribute, first, to assessment-facilitating (GRECI) explanations of intuitions and, second, to the resolution of paradoxes and the philosophical problems they engender (section 1.2). We will thus motivate two programmatic claims which we will then substantiate through a case-study on a notorious paradox, the ‘argument from illusion’, whose ‘puzzling power’ (Smith 2002, 21) has attracted comment (sections 2-3). The underlying intuitions about apparent visual ‘illusions’ will turn out to be cognitive illusions.

*1.1 Psycholinguistic Background*

Semantic memory is our memory for facts and ‘general world knowledge’, as opposed to personally experienced or ‘episodic’ events (McRae and Jones 2013, Tulving 2002). It is commonly conceived as a semantic network which doubles as information-storage and inference-engine. 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). When a concept is *‘activated’* it is more likely to be used by several cognitive processes, crucially including processes involved in utterance-comprehension (from word-recognition to disambiguation).[[6]](#footnote-6) Simultaneous activation of concepts can lead to the activation of a proposition made up of those concepts. An activated concept or proposition becomes conscious if – and only if – it is activated above a threshold and more strongly than all competitors, at a certain point in the course of processing. The associative process of 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 (classical review: Neeley 1991),[[7]](#footnote-7) the constant evolution of structure and links within the network ensures that information about the world is brought to bear on the process: The co-occurrence of features (things and their common properties, wholes and their common parts) and events (causes and typical effects, etc.) forges links between the respective nodes which grow stronger upon frequent activation and atrophy upon disuse. The more frequently we encounter tomatoes that are red (in the supermarket) or Germans who are nasty (in war movies), the stronger the links between the respective concepts become, the more activation gets passed on from the stimuli ‘tomato’ and ‘German’, respectively, to nodes representing ‘red’ and ‘nasty’, respectively. These concepts thus come to be *stereotypically associated* with the words: They are activated most rapidly and strongly, and come to mind first, when we encounter the words. We then spontaneously make *stereotype-driven* inferences to the conclusions that the fruit and people talked about are red and nasty, respectively.

This is nicely illustrated by a riddle that students find notoriously difficult even in classes on feminist linguistics (Giora 2003, 13):

(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 upon reading ‘surgeon’ you automatically attributed several stereotypical properties of surgeons to the speaker – including the stereotypical gender (male). When reading ‘He’s my son’, you then leaped to the further conclusion that the surgeon is the patient’s father – who has supposedly just been killed. Whence the difficulty.

In addition to nouns standing for objects and people, also nouns for actions and events (Hare et al. 2009) as well as verbs can be associated with stereotypical features (Harmon-Vukic et al. 2009, Ferretti et al 2001, McRae et al. 1997). Just complete the following sentences with the first word to come to mind:

1. She was sewing the socks with a \_\_\_\_\_\_
2. The man was arrested by \_\_\_\_\_\_
3. Joe is good at manipulating people, he is so \_\_\_\_\_\_\_
4. They arrested a\_\_\_\_\_\_\_
5. Jack keeps getting manipulated; he is so \_\_\_\_\_\_\_

The most frequent responses are: (1) ‘needle’, (2) ‘the police’ (3) ‘cunning’, ‘shrewd’, or ‘clever’ (4) ‘criminal’, ‘crook’, ‘suspect’, and (5) ‘naïve, gullible, stupid’. This reflects the fact that some actions are typically performed with specific instruments (sewing – needle), some are typically performed by certain kinds of agents (arrest – police) or brought off by agents with particular qualities (manipulate – cunning), and some are typically performed on certain kinds of things or people (called *‘patients’*) (arrest – criminal) or individuals with certain traits or properties (manipulate – naïve). Due to such co-occurrence these action/event-, agent-, and patient-features come to be stereotypically associated with these verbs.[[8]](#footnote-8) Such features jointly form *stereotypes* (aka ‘generalised situation schemas’) associated with the relevant verb (‘sew’, ‘arrest’, ‘manipulate’, etc.). They have us leap from the latter to conclusions about the denoted event or action, or its agents or patients, which attribute to them features stereotypically associated with the verb (e.g., the person who let herself be manipulated must have been a bit naïve).

These leaps duplicate inferences with the neo-Gricean *I-heuristic* (derived from Grice’s second Maxim of Quantity, ‘Do not say more than you must’, cp. Grice 1989, 26): ‘What is expressed simply is stereotypically exemplified’ (Levinson 2000, 37). This heuristic instructs us to facilitate or find interpretations that are positive, stereotypical, and highly specific (Levinson 2000, 118):[[9]](#footnote-9)

* It tells *speakers/writers* to be economical: ‘Skip mention of stereotypical features when talking about situations which conform to the relevant stereotypes; on the other hand, make deviations from the stereotype explicit (“female surgeon”).’ (Cp. Brown and Dell 1987)
* The rule tells *hearers/readers* to amplify the content of an oral or written utterance: ‘In the absence of explicit indications to the contrary, assume that, according to the speaker/writer, actions, agents and patients possess the features stereotypically associated with the noun or verb.’

Our explanation of philosophically relevant intuitions (in section 3) will build on the fact that spreading activation in semantic memory duplicates such subtle pragmatic inferences. These infer attributions of stereotypical properties while taking into account contextual cues (‘indications to the contrary’). This duplication is achieved through the interplay of automatic and attentional processes (for a review, Giora 2003).

Fully automatic *stimulus-driven processes* are wholly determined by current and immediately preceding linguistic input. According to the empirically well supported graded-salience hypothesis(Giora 2003, Peleg et al 2004, Peleg and Giora 2011), each verbal stimulus (word, collocation, or idiomatic expression) activates *all* semantic and stereotypical features associated with that expression, in its different senses.[[10]](#footnote-10) Crucially, speed and strength of such activation are a function of the frequency with which a given sense is used and the strength of the stereotypical association of the property with that sense. Speed and strength of activation depend, more specifically, on how *‘salient’* the given sense is, i.e., on how often the given speaker/hearer encounters or uses it, and how familiar it thus is to her (Giora 2003, 16-22). Stereotypical associates of salient senses are activated independently of context. E.g., the ambiguous stimulus ‘mint’ activates the probe ‘candy’ rapidly and strongly, even where it is used in a less frequent sense (prime: ‘All buildings collapsed except the mint’) (Simpson and Burgess 1985, Till et al. 1988).

Partially attentional *expectation-driven processes*, by contrast, are influenced also by the outcomes of other cognitive processes (such as simultaneous visual perception) and by previous stimuli (earlier parts of the sentence or paragraph), which previously activated associated stereotypical features, so that activation gradually builds up to concepts expected in the context. Stimulus- and expectation-driven processes initially run in parallel. Their outputs are subsequently integrated: The activation of some concepts is automatically enhanced, as they receive activation from others through different links to shared semantic or stereotypical features, while other concepts’ activation decays (Park and Reder 2004). Explicit marking of infrequent senses (through such riders as ‘figuratively speaking’) can further enhance the activation of relevant concepts, which might otherwise be side-lined by preferential activation of concepts associated with dominant uses (Givoni et al. 2013). Where the occurrence of concepts is inconsistent with expectations, its activation is ‘inhibited’ and/or delayed. In addition, the activation of concepts which are contextually inappropriate or irrelevant to the subject’s goals can be lowered through effortful ‘suppression’ (Gernsbacher and Faust 1991, Faust and Gernsbacher 1995, Williams 1992). In these different ways, the initial preferential activation of stereotypical associates can be mitigated in the light of contextual cues and explicit indications of deviation from relevant stereotypes.

The interplay of these processes duplicates inferences governed by the I-heuristic. These processes have been shown to occur not only in utterance and text comprehension but also in speech/text production (Levelt 1989, Pickering and Garrod 2013, Stephens et al. 2010, see also Giora 2003, 134-136). They are hence set to duplicate inferences in line with the I-heuristic not only in interpersonal communication but also in the sort of sub-vocalised cognition characteristic of philosophical thought.

*1.2 Philosophical Prospects*

The research we reviewed holds promise for the kind of explanation experimental philosophers seek in pursuit of the sources project: *GRECI explanations* which trace intuitions about verbally described cases back to automatic cognitive processes which are *generally reliable* but predictably give rise to *cognitive illusions* under specific circumstances. The links in semantic memory evolve or adapt in response to degree of exposure (section 1.1): The more red tomatoes I see, or hear, read, or think about, the stronger the link between the concepts ‘tomato’ and ‘red’ becomes. If, by contrast, I start to be exposed to mainly (unripe) green tomatoes, the link between ‘tomato’ and ‘green’ will be strengthened at the expense of the link between ‘tomato’ and ‘red’, and the former will activate the latter concept less and less strongly. As a result, the strength of stereotypical association is sensitive to the co-occurrence frequencies a person is exposed to.[[11]](#footnote-11) Unless the frequencies are seriously skewed by biasing media (war movies where all Germans are nasty), stereotypes therefore tend to be reasonably accurate and get gradually, albeit somewhat lengthily, modified where and when they have become inaccurate. Outside periods of rapid change (and topics notoriously attracting misrepresentation), inferences to attributions of stereotypical features in line with the I-heuristic are, by and large, reasonably reliable.

On the other hand, we can specify circumstances under which such inferences are bound to lead to *cognitive illusions*: to predictable wrong intuitions which strike us as plausible even once we know they are wrong. For example, suppose speakers *unwittingly* start to give a well-established word a new rarefied use, which allows for the word’s application to situations which do not conform to the stereotype associated with the word’s familiar and dominant use. Then they will not make the deviation from the stereotype explicit, and unmarked stereotype-inconsistent applications of the word will trigger stereotype-driven inferences to wrong conclusions which contextual integration need not correct. In other words, unwitting violations of the production-part of the I-heuristic will trigger contextually inappropriate inferences in line with its comprehension-part. Due to the effortlessness or fluency of such stereotype-driven inferences, the resulting judgment will strike us as plausible, irrespective of reflective endorsement (cp. Alter and Oppenheimer 2009).

The philosophy of perception provides a relevant example. Philosophers like to argue at a particularly general or abstract level. Sometimes, they consciously look for suitably generic terms.[[12]](#footnote-12) More frequently, they simply recruit well-established words which already have a more specific or subtly different remit in ordinary language, and thus endow familiar words with rarefied new uses, without noticing their novelty. Thus, philosophers who wish to argue at one go about all our five senses generally employ the verb ‘to perceive’ as mere shorthand for ‘to see or hear or smell or taste or feel’, without finding it necessary to explain this. Many use ‘to be aware of’ in an equally unexplained but yet more generic sense, namely, to speak simultaneously about our five senses and associated experiences, e.g., both about sounds of the kind we hear *with* our ears and of the kind Tinnitus patients constantly hear *in* their ears. In these generic uses, these expressions are not meant to carry any implications of knowledge. But they are in fact strongly associated with epistemic subject-properties, in their ordinary uses:

to perceive = to apprehend with the mind or senses, and

to be aware of = to have cognizance.[[13]](#footnote-13)

Second, and in the same generalising vein, philosophers of perception use the verbs ‘seem’ and ‘appear’ to speak simultaneously about how things look or sound or smell or taste or feel to the subject of perception or awareness.[[14]](#footnote-14) They often take themselves to be using these verbs in a purely phenomenal sense, in which they imply nothing about what people are inclined to judge or believe, as and when things look, seem, or appear thus-and-so to them. But, as we shall see (in section 3.1), appearance-verbs are strongly associated with such doxastic patient-properties. Philosophers of perception may hence correctly apply such verbs in their new technical uses to situations in which the patients of appearances lack pertinent inclinations to believe. Where philosophical authors don’t notice the need to explicitly mark this deviation from the stereotypes associated with the dominant uses of these verbs, they and their reader are then liable to automatically attribute such doxastic inclinations to patients, despite the contextual impropriety. We will follow this up for philosophical arguments (‘from illusion’) which base conclusions about what viewers are aware of, on premises about how things appear or seem to them.

Many characteristically philosophical problems arise from apparent conflicts between uncontroversial or common-sense beliefs, brought out by paradoxes: by arguments which proceed from premises articulating some uncontroversial or common-sense beliefs, and lead to a conclusion apparently inconsistent with another such belief, that p. They thus motivate questions of the form (Fischer 2011, cp. Papineau 2009):

How is it possible that p (viz. given that q)?

where q typically is the conclusion of the paradoxical argument.

A good example is ‘the [*sic*] problem of perception’ (as it is generally known since Smith 2002). Several paradoxes, including ‘arguments from illusion’, converge on the conclusion that, when using our five senses, we become (directly) aware only of subjective sense-data. This conclusion [q] appears to clash with such common-sense convictions as that [p] we see tables and chairs, hear shouts and explosions, and smell burnt milk. This clash motivates the question:

How is it possible that [p] we see and otherwise perceive physical objects and public events, given that [q] all we are ever (directly) aware of in perception are subjective sense-data?

This classical problem (lynch-pin of Western philosophy of perception from the 18th to the mid-20th century) has again become a focus of debate (Brewer 2011, Crane 2011, Fish 2009, Smith 2002, Robinson 2001). With problems of this form, it is worthwhile to explore whether the acceptance of q is due – only – to one or more automatic inferences to intuitive judgments, in the underlying paradoxical arguments.

We are justified in accepting claims in conflict with common sense (like ‘In perception, we are only aware of subjective sense-data’) only if we have positive reasons in their support.[[15]](#footnote-15) While we may not infer falsity from such conflict, we may hence infer need for positive justification. Where an intuition is found to give rise to paradoxical consequences (on its own or in conjunction with uncontroversial background assumptions), its acceptance accordingly requires positive reasons. Intuitive judgments, however, go with a high degree of subjective confidence and are often accepted without efforts to provide supporting argument or evidence (Thompson et al. 2011). In this case, the only positive reason the thinker has to accept her intuition is the mere fact that she has them, as and when she does. I.e.: Her warrant for accepting the intuitions that strike her as plausible then depends upon the evidentiary value of these intuitions.

GRECI explanations allow us to assess particular intuitions’ evidentiary value by tracing them to generally reliable cognitive processes that give rise to cognitive illusions under specific circumstances. Where we find such vitiating circumstances to obtain in the formulation of a paradox, we may infer that the relevant intuitions lack evidentiary value. Where these intuitions are accepted without further evidence or (non-circular) argument we may further infer that thinkers lack warrant for accepting them. We can thus resolve a paradox and the philosophical problem it engenders by constructing debunking GRECI explanations that reveal thinkers lack warrant to accept the intuitions it builds on. These may be intuitive premises or intuitive conclusions that are prompted by the premises and required for deriving the final conclusion (that q). Either way, thinkers then lack warrant also for believing this conclusion (q) and have no good reason for raising the characteristically philosophical problem of how it is possible that p, *given that q*. We can thus show this problem *ill-motivated*. For problems (like the problem of perception) that are raised by several arguments to the same paradoxical conclusion, this requires exposing unwarranted intuitions, in each argument.

To sum up, we have argued for two programmatic claims: first, that we can obtain GRECI explanations of philosophically relevant intuitions by invoking association processes in semantic memory that duplicate pragmatic inferences with the I-heuristic (*stereotype-driven amplification*); second, that such explanations can contribute to resolving philosophical paradoxes and the characteristic problems they engender. The proof of both kinds of pudding lies in the eating – in successful application to specific examples. For proof of concept, we will now furnish a first such application: We will identify the key intuitions at the root of ‘arguments from illusion’ and show how debunking explanations of those intuitions can resolve these paradoxes (section 2). Then we will properly develop (section 3) and experimentally test (section 4) a – debunking – GRECI explanation along the lines indicated.

1. **A Paradox and its Intuitive Roots**

Arguments ‘from illusion’ constitute the perhaps most notorious family of paradoxes in the philosophy of perception. They were at their most influential in 20th century analytic philosophy (Ayer 1940, 1956, Broad 1923, Moore 1918/19, Price 1932, Russell 1912, cp. Martin 2003). They proceed from premises that set out mostly familiar cases of non-veridical perception (mislabelled ‘illusion’) where physical objects look or otherwise appear to have a size, shape, colour, or other property they do not actually possess. E.g.:

1. When subjects view a round coin sideways, the coin appears elliptical to them.

Seminal statements of the argument (e.g., Hume [1748] 1975, 152) infer directly that, in these cases, an ‘image’ (aka ‘sense-datum’) rather than a physical object must be ‘present to the mind’.

Early 20th century statements (e.g. Ayer 1940, 4, Broad 1923, 240, Russell 1912, 1-3) break this decisive ‘sense-datum inference’ (Smith 2002, 25) up into two parts. From (1) they leap to the negative conclusion:

**[Version A]**

1. When subjects view a round coin sideways, they are not (directly) aware of the round coin.

The positive conclusion that subjects are, instead, aware of an ‘elliptical sense-datum’ is then obtained from an uncontroversial response to (2):[[16]](#footnote-16)

1. When subjects view a round coin sideways, they are (directly) aware of *something*.
2. By (2) & (3), subjects are then (directly) aware of something other than the round coin (something to be called a ‘sense-datum’).

This other object is then credited with the shape, size, and colour F that the coin merely looks or appears (there and then). This yields so-called *‘phenomenal judgments’*, such as:

1. The object that viewers are then (directly) aware of is elliptical.

Current textbook reconstructions of the argument turn the order around and derive the negative conclusion (2 above, 5 below) from a phenomenal judgment (5 above, 3 below), with the explicitly stated ‘Phenomenal Principle’ (2 below) and Leibniz’ Law (4 below) (e.g. Robinson 2001, 57-8):

**[Version B]**

1. When subjects view a round coin sideways, the coin appears elliptical to them.
2. Whenever something appears a shape, size or colour F to observers, they are (directly) aware of something that actually has that shape, size, or colour. Hence:
3. When subjects view a round coin sideways, they are (directly) aware of something that actually is elliptical (an elliptical patch).
4. If b has a property that a lacks, a≠b. Hence:
5. When subjects view sideways a coin that actually is round, they are (directly) aware not of the coin but of something else (to be called ‘sense-datum’).

Different versions of the argument then generalise in different ways to a conclusion about all cases of perception.

Commentators have characterised the argument as resting on an ‘appeal to intuition’ (e.g., Robinson 2001, 54) and regarded phenomenal judgments as the key intuitions involved. Regarding these judgments as intuitions is reasonable, since they struck proponents as immediately ‘plausible’ (Broad 1923, 240) or even ‘as plain as can be’ (Moore 1918/19, 21), while being neither perceptual nor introspective in nature: Proponents of the argument volunteer these judgments not when contemplating coins from the side but when they consider verbal statements, which typically state general facts (like 1) (cp. Ayer 1940, 3, Price 1932, 28), and then ask themselves ‘what is to be inferred’ (Ayer 1940, 4). They admittedly found it difficult to explain why they made, or why one should make, the subsequent inferences (cp. Ayer 1956, 89, Price 1932, 27). Accordingly, they do not explain or justify the Phenomenal Principle,[[17]](#footnote-17) which is explicitly invoked for these inferences in Version B but merely spells out their form, without providing any positive reason to believe they preserve truth. All this suggests that the relevant phenomenal judgments are obtained through the kind of effortless and unconscious inference that makes for intuitions.

The earlier Version A of the argument, however, suggests that its first premise also elicits other intuitions: Early analytic authors leap from the initial case-descriptions directly to the *negative judgment* that viewers are not aware of the coin they view, find this so plausible they immediately accept it despite apparent conflicts with common sense, and frequently offer no explanations or arguments to back up this crucial inference (above). As a working hypothesis, we will therefore assume that the initial case-descriptions prompt, directly or indirectly, both negative and phenomenal intuitions. Philosophers may then place these intuitions into a different order and in different inferential relations, when trying to transform intuitive lines of thought into acceptable arguments. We submit that versions A and B of the argument from illusion build on the same intuitive line of thought, which involves both negative and phenomenal intuitions. In the remainder of this section, we will argue that version A of the argument more accurately reflects judgments’ order in intuitive thought: that initial descriptions of cases of non-veridical perception (like 1) prompt intuitive negative judgments (like A2) which, in turn, prompt phenomenal intuitions (like A5). Then (in sections 3-4) we will develop and test a psychological explanation that traces these negative judgments back to automatic cognitive processes – the only way in which we can rigorously establish that they actually are intuitions in the presently relevant aetiological sense.

Before introducing technical terms, many philosophers express phenomenal judgments by saying that subjects are ‘aware’ or ‘conscious of’ an ‘F patch’ (e.g. Price 1932, 3) or ‘F speck’ (e.g. Ayer 1940, 22f), where F is the size, shape or colour the physical object viewed looks (from that perspective, etc.). When talking about familiar cases of non-veridical perception (like perspective) we ordinarily use those expressions

* 1. 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?’
	2. 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 do we 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 describes things in terms of others that look similar, in a particular respect:[[18]](#footnote-18) Instead of saying, ‘The man who was un-groomed and sloppily dressed like a tramp turned out to be the CEO’, we sometimes say more pithily: ‘The tramp turned out to be the CEO’. 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 (I) and (II) 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 (I). Similarly, we may use it in talk of what others see only when we are unsure 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 (II). 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 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, 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 the further inference (with Leibniz’ Law) to ‘The subject is aware of something other than the round coin’ as every bit as poor a joke as a fellow rambler’s response to (I): ‘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 arguments from illusion, phenomenal intuitions must therefore be *preceded* by equally intuitive negative conclusions which motivate the relevant patch-talk and *pre-empt* its metaphorical interpretation: Proponents of the argument naturally fall back on patch-talk when they have concluded that the viewer is not aware of the coin she looks at and are unsure about the identity of the ‘*something’* of which she *is* aware. Since ‘elliptical patch’ is recruited as cautious characterisation of this otherwise still mysterious object assumed to be distinct from the coin, the expression is not interpreted as referring to the coin. Since the argument’s initial premise does not introduce any other object the expression could be taken to refer to in its metaphorical use, it is naturally interpreted literally: ‘The viewer is aware of a patch that is elliptical.’ To explain the ease with which phenomenal judgments are made and accepted in arguments from illusion despite the apparent violation of Gricean maxims, we assume that initial premises prompt negative intuitions which precede and facilitate phenomenal intuitions.

This priority has the perhaps surprising consequence that we can debunk phenomenal intuitions by debunking the intuitive negative conclusions that current textbook reconstructions of the argument derive from them, in efforts of ex-post rationalisation (version B of the argument): Verbal descriptions of familiar cases of non-veridical perception prompt phenomenal judgments only in conjunction with prior negative intuitions. If these enabling intuitions are cognitive illusions, then the mere fact that thinkers have these phenomenal intuitions as and when they do (namely, when in the grip of a cognitive illusion), does not speak for the truth of these intuitions. They then have no evidentiary value.

In version A, arguments from illusion depend on negative intuitions at odds with common-sense convictions, which they give us no positive reason to accept. Whether proponents of the argument are justified in accepting these negative intuitions hence depends upon their evidentiary value. In version B, these arguments depend upon phenomenal intuitions. Since these intuitions are at odds with judgments competent speakers ordinarily make about the relevant cases, they, too, are in need of positive reasons in support. The Phenomenal Principle with which they are officially derived provides no such reasons. Hence thinkers are justified in accepting these phenomenal intuitions only to the extent to which they have evidentiary value. But they have such value only if the negative intuitions that precede and facilitate them are no cognitive illusions. A debunking explanation that exposes negative intuitions as cognitive illusions hence stands to debunk also phenomenal intuitions and resolve *both* versions of the paradox simultaneously.

1. **Explaining the Key Intuitions**

We will now explain these key intuitions as outcomes of automatic association processes in semantic memory which duplicate the I-heuristic. Our explanation will build on two experimentally testable hypotheses: The appearance-verbs used in the initial premises of arguments from illusion, namely,

H1 ‘x appears F to S’, ‘x seems F to S’, and ‘x looks F to S’ are strongly stereotypically associated with the doxastic patient-property *S is inclined to judge that x is F*.

Second, the expressions ‘S is aware of an F’ and ‘S perceives an F’ used in the negative intuitions prompted by those premises really stand for epistemic concepts. This translates into a testable hypothesis about association in semantic memory:

H2 ‘S is aware of an F’ and ‘S perceives an F’ are strongly associated with the epistemic subject property *S knows it is an F she is aware of/perceives*.

We will now informally motivate these hypotheses(section 3.1) and explain how stereotype-driven inferences which exploit the hypothesised associations can lead from the initial premises of arguments from illusion to the negative intuitions to be explained (section 3.2). Then we will present an experimental study that will provide the decisive – empirical – argument for the two hypotheses on which this explanation relies (section 4).

*3.1 Doxastic and Epistemic Associates of ‘Appearance’ and ‘Perception Verbs’*

Berit Brogaart (2013, forthcoming) 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 subjects (‘Joe’) and serve not so much to attribute any property from their complement (dirtiness) to those subjects’ referents (Joe) as to indicate an experiential, doxastic, or epistemic attitude of the often implicit patient to a content (Joe is dirty).

In ordinary discourse, we use ‘looks’ in visual and non-perceptual contexts alike, mostly to attribute doxastic attitudes to people: ‘The car looked quite new to the customer’, ‘The risks look manageable to our analysts’, etc. The customer may still have qualms about the car’s age, and the analysts need not unequivocally endorse the risk assessment; the doxastic attitude attributed amounts to an inclination to judge. Where the (often implicit) patient is the speaker, self-attributions can therefore express hedged judgments.

The uses in visual and non-perceptual contexts are clearly related:[[19]](#footnote-19) More often than not, observers are inclined to judge that what visually looks F *is* F. The use in visual contexts hence forges a stereotypical association between ‘x looks F [to S]’ and the doxastic patient-property *S is inclined to judge that x is F*. In contexts involving the patient’s use of her eyes, this doxastic implication is defeasible (‘From far away, the tower looked round to us, but we already knew it was square’). Through pragmatic strengthening and related processes (Traugott 1989, cp. Traugott and Dasher 2005) we often metaphorically extend the use of words to make them stand for stereotypically associated properties (‘Achilles is a lion’ for ‘Achilles is strong/ courageous/ noble’). This presumably motivates the use of ‘looks’ in attributing doxastic inclinations to patients, regardless of whether or not they make any use of their eyes. In this arguably dominant use, the doxastic implication is not defeasible (as witnessed by the anomaly of ‘The strategy still looks good to them, but they no longer think it is’). To sum up, the use of ‘x looks F [to S]’ in ordinary language forges a robust stereotypical association between the verb and the doxastic patient-property *S is inclined to judge that x is F*.

Turning to ‘appear’, the *Oxford English Dictionary* recognises two intransitive uses: ‘to be in one’s opinion; to be taken as’ and ‘to be in outward show, or to the superficial observer’. In the former use, it characterises a strong doxastic attitude of the patient: He is not merely inclined to judge, but judges that… In the latter use, for which the dictionary provides both visual and non-perceptual contexts, at any rate this strong doxastic implication is defeasible (the acute observer need not fall for the outward show). While ‘appears’ is often used with implicit reference to circumstances affecting judgment, ‘x seems (to be) F’ is mostly used with implicit reference to some evidence supporting the judgment that x is F (Austin 1962, 37), which is available to, and taken seriously by, the patient (as witnessed by the anomaly of ‘The accused seemed guilty to the jurymen, but they were not inclined to think her guilty.’) Thus, ‘x seems F [to S]’ is even more strongly associated with at any rate the weaker patient-property *S is inclined to judge that x is F*. For all the verbs at issue, the association holds regardless of any *suggestio falsi*: Even if the speaker suggests that things are not in fact as F as they look, appear, or seem, he still suggests that the patient is inclined to judge that they are F. (Whence the lack of *suggestio falsi* for first-person present-tense statements.)[[20]](#footnote-20)

The intuitive claims we seek to explain typically employ the verbs ‘to be aware of’ (e.g. Robinson 2001, 58, Smith 2002, 25) or ‘to be directly aware of’ (Ayer 1940, 4, 25, Broad 1923, 239, Crane 2011, 3) or ‘directly perceive’ (Ayer 1940, 2), though it stands to reason that the adverb gets inserted only as an afterthought. Occasionally, we also find ‘see’ (e.g. Ayer 1940, 4). All these verbs ordinarily carry the implication that the subject knows who or what it is she is seeing, perceiving, or aware of.

In the case of ‘see’, this epistemic implication is perfectly defeasible in visual contexts, as in:

1. At the reunion, I immediately saw my old classmate John, but, embarrassingly, recognised him only much later.

The stereotypical association is, however, strong enough to have motivated the metaphorical extension of ‘see’, analogous to that indicated for ‘looks’, to a purely epistemic use, as in ‘I see your point’ (Fischer 2014), where the knowledge-implication (I know what your point is) is indefeasible.

The association, we hypothesise, is even stronger for ‘perceive’ and ‘aware of’: The *Oxford English Dictionary* explains their ordinary uses as ‘to apprehend with the mind or senses’ and ‘to have cognizance’, respectively (see section 1.2). Both mark the same epistemic achievement – which need not be brought off through use of our five senses: ‘knowing by observation *or information*’.[[21]](#footnote-21) I can, e.g., become aware of the sunny weather in Malaga either by looking out the hotel window or by reading an international weather report.Accordingly, the implication that the subject of perception or awareness knows what it is she perceives or is aware of, would not seem to be defeasible even in visual contexts: Just try to replace ‘saw’ in (S) by ‘perceived’ or ‘was aware’.

The philosophical notions of ‘direct perception’ and ‘direct awareness’ share the implication of knowledge and impose the further requirement that the subject acquire the relevant knowledge without – conscious – inference or other intellectual process (Price 1932, 3, Russell 1912, 4, cp. Fischer 2011, 114-16).[[22]](#footnote-22) When hedging a conclusion they leaped to on the strength of those epistemic implications by adding ‘directly’ to ‘aware’ or ‘perceive’, thinkers hence do not retract the knowledge requirement on which their leap relied.

*3.2 Explaining Negative Intuitions*

Arguments from illusion tend to proceed from perfectly familiar cases of non-veridical perception (perspective, distance, yellow lighting) in which well-nigh nobody is tempted to judge that something *is* elliptical, small, or yellow, when it *looks* elliptical, small, or yellow. Proponents of the arguments explicitly acknowledge that, in these cases, viewers confidently judge that things actually have some shape, size, or colour distinct from the one they look under the circumstances (e.g. Ayer 1956, 88, Broad 1923, 236-7, 241, cp. Price 1932, 27). Accordingly, they should regard the use of verbs with strong doxastic implications (like ‘look’, ‘appear’, and ‘seem’) as misleading, and either refrain from their use or take precautions. In contrast with ‘seem’, ‘look’ and ‘appear’ have uses in which their doxastic implications are clearly defeasible, even if these are not their most frequent or salient uses. The customary precaution we take when using words in a less salient sense is to use them with an explicit marker that draws attention to that sense (Givoni et al. 2013) or to explicitly cancel the implication in question. One should therefore expect that proponents of arguments from illusion avoid the use of ‘seem’ in stating their initial premises and either preface ‘look’ and ‘appear’ with an introductory formula like ‘In a purely phenomenal sense’, or fully spell out, ‘When viewed sideways, round coins look elliptical, though nobody is tempted to think they are that shape.’

But nothing of the sort happens: Statements of arguments from illusion employ mainly ‘appear’ (e.g. Ayer 1940, 3, Robinson 2001, 57, Russell 1912, 2, Smith 2002, 25) and ‘seem’ (e.g. Ayer 1940, 3, Broad 1923, 239-40, Crane 2011, 3, Moore 1918/19, 21-3, Russell 1912, 2), with, if anything, a slight preponderance of the latter. In no statement of the argument is their use accompanied by explicit marking, even where doxastic uses of appearance-verbs are mentioned in related discussion (e.g. Robinson 2001, 52, Smith 2002, 37). The authors at issue thus violate the production-part of the I-heuristic, by using ‘looks’, ‘appears’, and ‘seems’ without making the deviation from the relevant stereotypes explicit. They arguably do so unwittingly and because, in setting out the argument, they neither realise the dominance of the doxastic uses nor bear in mind the strength of the doxastic implications at issue. Thus, even philosophers using 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 then ‘set aside’ these uses without further ado (e.g., Smith 2002, 37).

Recall (from section 1.1), however, that all semantic and stereotypical features associated with a word get activated upon encounter, that the features most strongly associated with the word on its dominant use get activated most rapidly and strongly, that the most strongly activated concepts determine how we interpret the claim and to which conclusions we leap from it, and that such stereotype-driven inferences may be made even in inappropriate contexts. Contextual integration may prevent inappropriate inferences from going through, where the context is sufficiently informative and rich (e.g. simultaneously provides visual and verbal information). In addition, explicit marking of less salient senses speeds up activation of the relevant concepts and brings it above threshold (Givoni et al. 2013). But absent such marking or rich context, inappropriate stereotype-driven inferences from non-dominant uses in stereotype-deviant cases are liable to go through unchecked.

Arguments from illusion proceed from premises employing ‘look’ or, more often, ‘appear’ or ‘seem’, such as:

P A round coin seems elliptical when viewed sideways.

According to our hypothesis H1, all three verbs are strongly associated with the stereotypical patient-property *S is inclined to think that x is F*. if this is correct, the I-heuristic licenses inferences from all standard formulations to conclusions like:

C0 The viewer is inclined to judge that the object viewed is elliptical.

The graded-salience hypothesis (section 1.1) tells us that as we encounter, say, the expression ‘seems elliptical’, a stimulus-driven automatic process will initially duplicate this inference, before the outputs of automatic expectation-driven processes are integrated and yield a potentially conscious conclusion. P itself tells us that the object viewed is round.[[23]](#footnote-23) The judgment the viewer is inclined to make (according to C0) is hence false. She lacks the confident true belief required for knowledge and sports either the indecision or false belief stereotypical of ignorance. The integration of C0 with further information from the sentence context (of P) thus leads to the potentially conscious conclusion

C1 The viewer is ignorant of the shape of the coin / object viewed.

When we read the accident vignette (R) (in section 1.1), the stimulus-driven activation of stereotypical properties had us leap to the conclusion that the speaker, introduced as ‘surgeon’, is male. The same process, followed by contextual integration, has us leap from standard premises of the argument from illusion to the conclusion that the viewer lacks basic perceptual knowledge about the object she is viewing. The leap from ‘surgeon’ to the gender-attribution corresponds to the automatic inference from P to C0. The present leap all the way to the potentially conscious judgment C1 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 (cp. section 1.1), 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):

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. (E.g. Oostendorp and Mul, 1990, 36-7)

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 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 ordinary use, ‘aware of’ means ‘to know by observation or information’ and is strongly associated with the most pertinent epistemic subject-property (according to our hypothesis H2). The word ‘unaware’ hence is a near-synonym of ‘ignorant’ (in C1). Philosophers typically consider premises like P in the light of a guiding question. This is the question not of what subjects know, but of what they are aware of in the cases described. Hence attention is directed towards ‘aware’ and its antonym, and ‘unaware’ is set to become even more strongly activated than ‘ignorant’, when a philosopher leaps from P to C1. She will thus make the resulting judgment in the shape of:

C1\* The viewer is unaware of the shape of the coin / object viewed.

Statements of the argument from illusion most frequently proceed from examples involving objects with characteristic and standardised shape, size, or colour, like penny-coins (Ayer 1940, 3 and 1956, 86, Broad 1923, 239ff, Robinson 2001, 53). When (C0) 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: (C2) 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:

C2\* The viewer is unaware of the coin viewed.

Judgments about sensible properties (like C1\*) predominate in early analytic formulations of the argument from illusion (e.g. Ayer 1940, 4, Russell 1912, 3), while current reconstructions (with the exception of Smith 2002) mostly contain judgments about objects (like C2\*). Since ‘perception’, ‘direct perception’, and ‘direct awareness’ all require knowledge, elementary reasoning leads to the common alternative formulations of the present conclusions.

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 the same heuristic. Without realising it and taking the usual precautions, philosophers make a non-dominant use of ‘seem’, ‘appear’, or ‘look’, in premises that describe situations which do not conform to the stereotypes strongly associated with these verbs in their dominant uses, and then cannot help making automatic inferences driven by those very stereotypes. This results in the paradoxical intuition that viewers are unaware of familiar objects they look at, in familiar cases of non-veridical perception, in which these things look exactly as we would expect. This conclusion will appear even more paradoxical to philosophers of perception who have taken up the generic use of ‘aware of’ as mere short-hand for ‘see or hear or smell or taste or feel’, with entirely undetermined object (see section 1.2). Such philosophers will take conclusions like C2\* to imply that the viewer does not see the coin she looks at sideways.

This apparent conflict with the common-sense conviction that we typically see coins and other such physical objects engenders the problem of perception (section 1.2). If correct, the proposed explanation doubly resolves this problem, to the extent to which it is engendered by arguments from illusion: First, it shows that these arguments’ initial premises prompt the crucial negative intuitions (like C2\*) under a set of special circumstances (viz., unwitting violation of the production-part of the I-heuristic) under which the relevant effortless cognitive processes (which duplicate inferences with its comprehension-part) predictably lead to wrong conclusions we find compelling, even so (due to their fluency). Thus, our explanation exposes these intuitions as cognitive illusions. Second, it shows us that these unwarranted intuitions are not even at odds with the common-sense conviction that we typically see the coins we look at sideways: We leap to the intuitive judgment when treating ‘unaware’ as the near-synonym of ‘ignorant’ that it ordinarily is. But of course a person can see a coin without knowing it is one. Only the undue mixing of ordinary and philosophical use, in automatic and subsequent conscious inference, creates the impression of a conflict.

1. **Experiment**

This debunking explanation builds on the two key hypotheses H1 and H2 (top section 3). To test them, we employed a forced-choice plausibility-ranking task.

*4.1 Approach and Predictions*

*4.1.1 Approach:* In our task, participants are presented with pairs of sentences or short (2-sentence) texts which differ only in one ‘critical’ word or expression (‘minimal pairs’). E.g.:[[24]](#footnote-24)

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

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

Participants are asked to indicate whether (a) or (b) strikes them as more plausible, and to make a judgment even if they have no clear-cut preference (‘forced-choice’). Hypotheses about the relative strength with which two different verbs are stereotypically associated with a given event-, subject- or patient-property can then be tested by constructing texts where the critical verb (say, ‘seemed’) is followed or preceded by a sentence which is incongruent with the attribution of the property in question. For instance, we can test the hypothesis that ‘x seems F’ is more strongly associated than ‘x looks F’ with the doxastic patient property *S is inclined to judge that x is F*, by using item 29: If the patient (Jack) is inclined to judge that Jane is tanned, he will not ‘suppose it was just a trick of the light’. If ‘seems’ is indeed more strongly associated with that doxastic patient-property, the seems-sentence (b) will engender a stronger sense of incongruence, and participants will judge the looks-sentence (a) more plausible.

For each pair of verbs, we created several minimal pairs which are all incongruent with attributions of the same event-, subject-, or patient-property P. By noting preferences for sentences with one verb over the other, across several minimal pairs, we can measure the consistency of plausibility judgments, and infer with increasing confidence which of the two verbs has the stronger stereotypical association with property P: If preferences are random (no consistency), both verbs will be associated with P equally strongly (and possibly, not at all). The more often one kind of sentence (say, with ‘look’) is preferred over the other (say, with ‘seem’), the more confident we may be that its critical verb (‘look’) is less strongly associated with P than the other (‘seem’).[[25]](#footnote-25)

*4.1.2 Predictions:* To employ this paradigm, we recast H1 and H2 as hypotheses about the relative strength of associations. We assumed as uncontroversial that ‘see’ and ‘hear’ have strong but clearly defeasible epistemic implications and thus recast H2 as the hypothesis that ‘aware’ and ‘perceive’ are associated even more strongly than ‘see’ and ‘hear’ with the epistemic subject-property *S knows it is an F she is aware of/perceives*. We thus predicted that when facing minimal pairs that are incongruent with those implications, participants would judge see- and hear-sentences more plausible than analogous aware- and perceive-sentences. In recasting H1, we assumed as uncontroversial that both ‘x appears F’ and ‘x seems F’ are strongly associated with the doxastic patient-property *S is inclined to judge that x is F*, and hypothesised that the association is significantly stronger for ‘seems’ than for ‘appears’. We then reasoned that if there is a still strong association also with ‘looks’, it will be roughly comparable in strength with ‘appears’. We thus predicted that participants’ choices between looks- and appears-sentences will be random (close to chance), and participants’ preference for looks- and appears- over seems-sentences will be significant.[[26]](#footnote-26) At this point, however, we expected philosophers to perform differently: Since ‘appears’ is a low-frequency word in ordinary discourse (cp. Fn.27), the interchangeable use of ‘appears’ and ‘seems’ in the philosophy of perception (section 1.2) should suffice to assimilate their stereotypical associations to the point where philosophers engaging with work in the area make random choices also between appears- and seems-sentences.

*4.1.3 Excluding Key Confounds:* In the forced-choice paradigm, stereotypical association is inferred from perceived incongruence which, in turn, is inferred from lower plausibility rankings. A sense of ‘semantic incongruence’ can be due to a conflict of the sentence context with different kinds of world knowledge: with frequency information implicit in stereotypical associations of the critical verbs (cp. section 1.1), and with more complex ‘scripts’ or schemas organising background knowledge about the typical sequence and structure of events (concerts, museum visits, etc.) (Abelson 1981).

Contrast, for instance, the minimal pairs X and 42:

Xa/b. The fans began to applaud/cheer because the band had come back on stage. The encore was better than the actual show.

42a/b. Some visitors spend most of their time viewing/watching the abstract artwork in the top floor of the museum.

The verb ‘cheer’ might be stereotypically associated with fans as fillers of the subject-role, and more strongly so than ‘applaud’. But comparative plausibility judgments about X might well be influenced by participants’ knowledge about what typically happens when, at pop (as opposed to classical) concerts. So X would be useless for our purposes. By contrast, knowledge about what typically happens when, during museum visits, does not help decide between 42a and 42b – which we designed to elicit judgments relying on the fact that ‘watch’, but not ‘view’, is stereotypically associated with events as objects (we don’t ‘watch the paint’, though we may ‘watch the paint dry’; Husband, 2012).

When subjects cannot base plausibility judgments on factual knowledge, they base them on metacognitive cues, in particular on the level of *‘fluency’* or subjective ease they 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 background knowledge, judgments will largely depend on the level of ‘fluency’ readers experience in understanding the relevant statements. We predict differences in plausibility which are due to different degrees of consistency between verbs’ associated stereotypes and sentence contexts, where incongruence engenders a sense of difficulty or ‘disfluency’ that leads to a lower plausibility-rating. We therefore need to exclude not merely other sources of perceived incongruence but also alternative sources of disfluency.

The overall ease or difficulty of reading and understanding is a function of the ease or difficulty with which several different processes involved are executed: phonological processing, word recognition, parsing, and integration. Manipulations of ease for each process have been found to influence participants’ judgments about statements’ plausibility, truth, or accuracy. These judgments can demonstrably be influenced, e.g., by the legibility of the font used to print the statements (Alter et al. 2007), the familiarity and pronounceability of the individual words (Oppenheimer 2006), the syntactic complexity of the sentence (Lowrey 1998), and whether subsequent words are primed by earlier expressions (Kelley and Lindsey 1993).

Most of the relevant factors are controlled for by presenting minimal pairs. But the critical words that are different (e.g., ‘see’ vs ‘perceive’) may still affect plausibility judgments about the overall sentences through their relative frequency and, hence, familiarity: e.g., see-sentences may be preferred over analogous perceive-sentences simply because ‘see’ is used approximately 66 times more often in ordinary discourse.[[27]](#footnote-27) This key confound was addressed in questionnaire design (section 4.2.2). Second, critical words may prime different concepts. We are interested in a particular form of semantic priming, namely, the differential activation of stereotypical associates (e.g. ‘see’-‘know’), but not in associative priming due to frequent co-occurrence of words (e.g. ‘see’-‘nose’, forged by the standing expression ‘see beyond one’s nose’). Relevant differences in degrees of associative priming between well-formed minimal pairs are usually reflected in judgments of which sentence ‘sounds better’ or ‘more idiomatic’. In a norming exercise, we therefore excluded or rephrased all items where even one discussion group participant explained his preference by saying that one sentence ‘sounded better’ than the other (section 4.2.3).[[28]](#footnote-28)

*4.2 Methods*

*4.2.1 Participants:* 47 volunteers participated without compensation, 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 (with BAs in other subjects). To mitigate bias potentially inherent in the common restriction to university-based samples, we also recruited 10 local extra-academic professionals (lawyers, geologists, and geo-engineers). All 47 were native speakers of English.

*4.2.2 Materials:* We administered a questionnaire with 66 minimal pairs to participants, and instructed them to read the sentences carefully and then indicate which of the pair ‘strikes you as more plausible’. The questionnaire contained 36 critical items, viz., six items for each of the six pairings:

1. look/appear
2. look/seem
3. appear/seem
4. aware vs see (3 items) or hear (3 items)
5. perceive vs see (3 items) or hear (2 items) or smell (1 item)
6. perceive/behold

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* (cp. H1).[[29]](#footnote-29) Items involving ‘aware’ and ‘perceive’ had sequels inconsistent with attributions of the epistemic subject-property *S knows what she is aware of / perceives* (cp. H2).[[30]](#footnote-30) Perhaps surprisingly, an earlier pilot study (on 45 undergraduate philosophy students from the University of East Anglia) 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 . Each of item-types (1) to (3) now used both constructions, though always the same for both constituents of an item, and 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 (e.g., ‘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 (like items 46, 50, and 60 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. Examples for (1) to (6) above, respectively (presented here in compressed format):

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

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

60a/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.

35a/b. Peter saw/was aware of the heirloom among the bric-a-brac. He took it for a piece of junk and threw it away with the rest.

26a/b. The explosion of the secret ammunitions depot could be perceived/heard by anyone within twenty miles – most of them civilians who didn’t know what the loud bang was.

38a/b. After a strenuous ascent out of the valley, I beheld/perceived the magnificent range of Mount McConnell. Unfortunately, I did not know at the time what I was looking at.

To gauge the extent to which our participants’ plausibility judgments are influenced by 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 one, the ‘frequency-congruent’ item, the text employing the more frequently used verb was also more consistent with the relevant stereotype. In the other, ‘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 over subject*. We can then compare the judgments participants make about frequency-congruent items (such as 61) and frequency-reversed items (such as 3), respectively:

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

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

If participants make judgments predominantly in line with stereotype-consistency and do not make significantly fewer such judgments about frequency-reversed than about frequency-congruent items, then this suggests that participants’ judgments are not influenced by word-frequency, but reflect different strength of stereotypical association.

To check more specifically the extent to which judgments about *critical items* (i.e. minimal pairs with the verbs ‘look’, etc., we are actually interested in) are influenced by word-frequency, we can consider those *‘potentially frequency-sensitive’* participants who performed better on the frequency-congruent than the frequency-reversed fillers, and compare their judgments about the critical items with those of the *‘frequency-insensitive’* participants whose judgments were clearly not affected by word-frequency (since their performance on frequency-reversed fillers was no worse, or even better, than that on frequency-congruent fillers).

As a final check, we paired the verb ‘perceive’ not only with the dramatically more frequent verbs ‘see’ and ‘hear’ (Leech et al. 2001), but also with ‘behold’ which is used about ten times less frequently than ‘perceive’ (cp. Fn.27) but resembles ‘see’ and ‘hear’ in having weaker epistemic implications. We wanted to determine whether participants make significantly fewer plausibility judgments in line with stereotype-consistency about the frequency-reversed perceive-behold items and, if so, whether participants’ performance on perceive-behold items correlates with their performance on frequency-reversed fillers – which would suggest an effect of word-frequency on their plausibility judgments about perceive-items and, possibly, the other verbs tested.

*4.2.3 Procedure:* In constructing frequency-congruent and –reversed filler items, we used word-frequency information from Leech, Rayson, and Wilson (2001). In contrast to the frequency-norms most commonly used in psycholinguistic research (Kucera and Fancis, 1967), this is based on a British English corpus, appropriate for the British participants in this study.

In a norming study four philosophy students (1 PhD and 3 advanced undergraduate students) responded to a draft questionnaire and were then asked to explain their responses with the first reason to come to mind, first independently on paper, then in group discussion. Since pragmatic inferences, including inferences in line with the I-heuristic, are comparatively easily recoverable (Levinson, 2000), responses even from a small group are informative. We discarded or rephrased all items in which even one participant mentioned that one formulation ‘sounded better’ than the other (cp. section 4.1.3), and included only items where at least two participants independently invoked the same stereotypical associate of the critical verb, all agreed in subsequent discussion that the verb had these ‘connotations’ or ‘associations’ for them, and no other association was mentioned by more than one student.

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 affect judgments, see De Neys 2006). The questionnaire was limited to 66 items to ensure that it could be completed within 20 minutes, which helps ensure that participants remain engaged.

Predicted or ‘correct’ responses were coded as 1, ‘incorrect’ responses as 0. Prior to the inferential analyses, we assessed both subjects and items for potential outliers on all critical items. To do so, we converted the proportions of correct responses into standardized scores (i.e., mean of 0 and standard deviation of 1). Given the size of our sample, we considered values greater than +/- 2.5 SDs from the mean to be potential outliers. For the item analysis (i.e. the means for each critical item averaged across subjects), there were no values greater than +/- 2.0 SDs in any condition. This ensures that effects are not driven by one or two items in particular. For the subjects analysis (i.e. means for each subject averaged across critical items in each condition), there was only one potential outlier. That data point was more than 4 SDs below the mean, and it occurred in the perceive/see (etc.) condition. However, the mean of this data point was .60 which is numerically higher than chance. The overall mean in this condition was .97 (see Table 1). Therefore, we felt that this particular data point was in fact not an outlier, but instead, due to a ceiling effect in this particular condition. Thus, we retained it in the final analysis, and made no adjustments to our dataset based on the outlier analysis.

*4.3 Results and Discussion*

The descriptive results and inferential statistics are reported in Table 1:

Table 1 – *Means, standard deviations, and t-tests for the full sample (N = 47)*

\_\_  *Mean (SD) One-sample t-test (0.50)*

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

Aware/see or hear 0.79 (0.20) *t*(46) = 10.07, *p* < 0.01

Perceive/see or hear or smell 0.97 (0.09) *t*(46) = 36.86, *p* < 0.01

Perceive/behold 0.81 (0.22) *t*(46) = 9.72, *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

The results confirm our predictions. Consider the preferences between look- and appear-sentences inconsistent with attributions of the doxastic patient-property *S is inclined to judge that x is F*. These preferences were not significantly different from chance. In contrast 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 (cp. H1) that ‘look’ is as strongly associated with our doxastic patient-property as ‘appear’, and 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* > 0.05). This suggests that the difference between these kinds of contexts either makes no difference to the use of these verbs, or affects both to the same extent. In other words, these results are inconsistent with the suggestion that, in visual contexts, these verbs are used in a ‘purely phenomenal’ sense in which they are both devoid of doxastic implications, while being used, in non-perceptual contexts, in a different (‘epistemic’) sense in which they have doxastic implications of different strengths. In contrast, results are consistent with the hypothesis that ‘appear’ and ‘seem’ have roughly similar doxastic implications in both visual and non-perceptual contexts – ‘appear’ consistently more weakly and defeasibly than ‘seem’.

Turning from doxastic associations of appear-verbs to the epistemic implications of perception-verbs, we find that, in sentence contexts inconsistent with attributions of the epistemic subject-property *S knows what she is aware of/perceives*, participants displayed a significant preference (see Table 1) for ‘see’, ‘hear’, and ‘smell’ over both ‘aware’ (0.79) and ‘perceive’ (0.97), and for ‘behold’ over ‘perceive’ (0.81). These findings are consistent with the hypothesis H2 that our epistemic subject-property is strongly associated with ‘aware’ and ‘perceive’, namely more strongly than with ‘see’, ‘hear’, ‘smell’, or ‘behold’, where the implication of knowledge and recognition is defeasible.

Crucially, 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)= -.46, *p* > .65). This suggests that word frequency did not influence plausibility judgments. To follow this up further, we identified 18 potentially frequency-sensitive and 29 frequency-insensitive participants. The response pattern we had found for the overall sample was clearly reproduced by both groups, as reported in Table 2.

Table 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

Aware/see or hear 0.82 (0.19) 0.77 (0.20) *t*(45) = 0.91, *p* > 0.05

Perceive/see or hear or smell 0.99 (0.05) 0.95 (0.10) *t*(45) = 1.45, *p* > 0.05

Perceive/behold 0.76 (0.25) 0.83 (0.19) *t*(45) = -1.15, *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

Accordingly, while preference for sentences with the lower-frequency verb ‘behold’ over ‘perceive’ was significantly lower than preference for sentences with the higher-frequency verbs ‘see’ and ‘hear’, differential performance on the perceive/see or hear or smell and perceive/behold items did not correlate with differential performance on frequency-congruent (*r* = 0.21, *p* > .15) and frequency-reversed fillers (*r* = 0.12, *p >* 0.40), respectively.This suggests that differential performance is due to other features of the relevant items than the relative frequency of the critical verbs. These findings eliminate the key confound of word-frequency and allow us to interpret our results as not merely consistent with, but supportive of our hypotheses about strength of associations.

Finally, the responses of the 8 professional philosophers were consistent with our specific prediction about philosophers of perception. However, due to the limited sample size, we treat these findings as having merely heuristic value. The participating philosophers were quite stereotype-sensitive in their judgments: They showed better overall accuracy on the filler items than frequency-sensitive participants. (Frequency-sensitive participants had mean accuracy of 0.81, and philosophers 0.88, cp. Table 2.) Furthermore, philosophers were better at making judgments about frequency-reversed filler items (0.92) than about frequency-congruent items (0.84).[[31]](#footnote-31) This strongly suggests that word-frequency did not affect their judgments. Crucially for our purposes, these philosophers made random choices not only between ‘appear’ and ‘look’ (0.42), like everybody else, but, as predicted, also between ‘appear’ and ‘seem’ (0.54).[[32]](#footnote-32) However, their preference for ‘looks’ over ‘seems’ was even more pronounced than that of other participants (0.92 vs. 0.62). These results suggest that, for these philosophers, the differences in strength of stereotypical association between ‘looks’, ‘appears’, and ‘seems’ are very slight: namely, that the differences between ‘looks’ and ‘appears’, and ‘appears’ and ‘seems’, respectively, are too slight to affect plausibility judgments even in a forced-choice task, but that these two sub-threshold differences add up to a difference just large enough to consistently affect the judgments at any rate of highly stereotype-sensitive subjects like these philosophers. This is consistent with the observation that, in the relevant philosophical literature, ‘seems’ and ‘appears’ are used interchangeably, and as supposedly merely more generic terms than ‘looks’ and its cognates .

To sum up, our results are consistent with the hypotheses that (H1) the appearance-verbs ‘look’, ‘seem’ and ‘appear’ and (H2) the perception-verbs ‘to perceive’ and ‘to be aware of’ have associations with doxastic patient- and epistemic subject-properties, respectively, that are sufficiently strong to play the explanatory roles we suggested for them (section 3.2).

*4.4 Future Research*

A well-established battery of psycholinguistic tests can then be used to confirm whether people actually make inferences facilitated by these associations, in inappropriate contexts (where, e.g., items 46, 50, and 60 above are even more inappropriate than item 29). Inferences to sequel-incongruent conclusions lead to slow-downs in reading (longer fixation-times for sequels) and signature electrophysiological responses (N400s). To test hypotheses about automatic inferences, psycholinguists therefore use materials like ours, which have been normed to exclude possible confounds, e.g., through plausibility ranking tasks like ours, for reading-time measurements (Klin et al. 1999, Harmon-Vukic et al. 2009) with eye-tracking (Patson and Warren 2010) and electrophysiological measurements of event related brain potentials (Kutas and Federmeier 2000, 2011).

**5. Conclusion**

This paper has identified a generally reliable cognitive process that generates cognitive illusions under specific circumstances, viz. routine stereotype-driven amplification; second, it has specified one set of such vitiating circumstances, viz., unwitting violations of the I-heuristic (as when philosophers give a new rarefied use to a well-established word, without realising the novelty) (section 1). Second, the paper developed (section 3) and experimentally tested (section 4) an explanation showing that this process, under those conditions, leads to the paradoxical negative intuitions (e.g., viewers aren’t aware of coins they view sideways) we found to be at the heart of arguments from illusion (section 2). This shows us that these intuitions have no evidentiary value. We can then infer (ibid.) that the phenomenal intuitions these negative judgments precede and facilitate have no such value, either, and that we have no warrant to accept the conclusions based on these intuitions. We have, e.g., no warrant to conclude that viewers who look at a round coin, are (directly) aware of a colour-patch or sense datum which actually is elliptical – rather than the coin which is round. The debunking explanation of those negative intuitions provides the key to the resolution of notorious paradoxes about perception. We have thus obtained first proof of concept for the idea that engagement with the psychology of language can help us resolve philosophical problems, not only in the philosophy of language, but in several areas where simple descriptions of hypothetical cases elicit philosophical intuitions.[[33]](#footnote-33)

**References**

Abelson, R.P. 1981: The psychological status of the script concept, *American Psychologist 36*, 715-729

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

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.

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

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

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

Ayer, A.J. 1956: *The Problem of Knowledge*. Repr. 1990, London: Penguin

Banaji, M.R. and Hardin, C.D. 1996: Automatic stereotyping, *Psychologyical Science 7*, 136-142

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

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

Blair, I.V. and Banaji, M.R. 1996. Automatic and controlled processes in stereotype priming, *Journal of Personality and Social Psychology, 70,* 1142-1163.

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

Brewer, B. 2011: Perception and Its Objects. Oxford: OUP

Broad, C.D. 1923: *Scientific Thought*. Repr. 2000, Abingdon: Routledge

Brogaard, B. 2013: It's not what it seems. A semantic account of ‘seems’ and seemings, *Inquiry 56*, 210-239

Brogaard, B. (forthcoming): The phenomenal use of 'look' and the contents of perception. In: P. Pagin and K. Gluer (eds.), *Philosophy Compass.*

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

Cappelen, H. 2012: *Philosophy without Intuitions*. Oxford: OUP

Chisholm, R. 1957: *Perceiving*, Ithaca: Cornell UP

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

De Neys, W. 2006: Automatic-heuristic and executive-analytic processing during reasoning: chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology 59,* 1070-1100

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

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

Evans, J.S.B.T. and Stanovich, K. E. 2013: Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science 8*, 223-241

Faust, M., & 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*, New York: OUP

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

Fischer, E. 2014: Philosophical intuitions, heuristics, and metaphors *Synthese,* 191, 569-606

Fischer, E., Engelhardt, P.E. and Herbelot, A. 2015: Intuitions and illusions. From explanation and experiment to assessment. In: E. Fischer and J. Collins (eds.): *Experimental Philosophy, Rationalism, and Naturalism*. London: Routledge, pp. 259-292

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

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: CUP

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

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, Suppl. Vol. 35*, 121-52

Grice, H.P. 1989: Logic and conversation. In his: *Studies in the Ways of Words*, Cambridge, Mass.: Harvard UP, pp. 22-40

Hanfling, O. 2000: *Philosophy and Ordinary Language*, London: Routledge

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009): Activating event knowledge. *Cognition 111*, 151-167

Harley, T. 2014: *The Psychology of Language*, 4th ed., Hove: Psychology Press

Harmon-Vukić, M., Guéraud, S., Lassonde, K.A. and O’Brien, E.J. 2009: The activation and instantiation of instrumental inferences. *Discourse Processes 46*, 467-90

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

Hume, D, 1748/1975: *Enquiries Concerning Human Understanding and Concerning the Principles of Morals.* Ed. by L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon

Husband, E.M. 2012: *On the Compositional Nature of States*, Amsterdam: John Benjamins.

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

Jaszczolt, K. 2011: Salient meanings, default meanings, and automatic processing. In K. Jaszczolt and K. Allan (eds.), *Salience and Defaults in Utterance Processing*, Berlin: de Gruyter

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

Kahneman, D. and Frederick, S. 2005: A model of heuristic judgment. In K.J. Holyoak and R. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning*. Cambridge: CUP, pp. 67-293

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

Kornblith, H. 2015: Against naturalistic defences of intuition. In: E. Fischer and J. Collins (eds.): *Experimental Philosophy, Rationalism, and Naturalism*. London: Routledge, pp. 151-168

Kucera, H., and Francis, W.N. 1967: *Computational Analysis of Present-day American English*. Providence: Brown University Press

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

Kutas, M. und Federmeier, K.T. 2011: Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology 62*, 621-647

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

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

Levinson, S.C. 1983: *Pragmatics*. Cambridge: CUP

Levinson, S.C. 2000: *Presumptive Meanings. The Theory of Generalized Conversational Implicature*, Cambridge, Mass.: 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*, Cambridge: CUP

Maund, J.B. 1086: The phenomenal and other uses of ‘looks’, *Australasian Journal of Philosophy 64*, 170-180

McCauley, C. and Thanga Velu, K. 1991: Individual differences in sex stereotyping of occupations and personality traits, *Social Psychology Quarterly 54*, 267-279

McRae, K., Ferretti, T.R., and Amyote, I. 1997: Thematic roles as verb-specific concepts. *Language and Cognitive Processes 12*, 137-176

McRae, K. and Jones, M. 2013: Semantic memory. In D. Reisberg (ed.), *Oxford Handbook of Cognitive Psychology*, Oxford: OUP

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

Moors, A. and De Houwer, J. 2006: Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin 132*, 297-326

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

Nagel, J. 2011: The psychological basis of the Harman-Vogel paradox. *Philosophers’ Imprint 11(5)*, 1-28

Nagel, J. 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 et al. (eds.): *New Waves in Philosophy of Action*, 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.* Hillsdale: Erlbaum, pp. 264-336

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, A.M. Surprenant (eds.) 2001: *The Nature of Remembering*. Washington, DC: APA, pp. 69-93

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

Papineau, D. 2009: The poverty of analysis. *Aristotelian Society Supplementary Volume*, 83, 1-30

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

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

Peleg, O., Giora, R. and Fein, O. 2004: Contextual strength: the whens and hows of context effects. In I. Noveck and D. Sperber (eds.): *Experimental Pragmatics*. Basingstoke: Palgrave

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*, Berlin: de Gruyter, pp. 32-52

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

Postal, P. 1973: *On Raising.* Cambridge, Mass.: 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), URL = <http://plato.stanford.edu/archives/win2012/entries/intuition/>

Rayner, K. 1998: Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin 124*, 372-422

Recanati, F. 2010: *Truth-conditional Pragmatics*, Oxford: OUP

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

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

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

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, Mass: Harvard UP

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

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

Traugott, E. 1989: On the rise of epistemic meanings in English: An example of subjectification in semantic change, *Language 65*, 31-55

Traugott, E. and Dasher, R.B. 2005: *Regularity in Semantic Change*, Cambridge: CUP

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

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

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: OUP

1. They need not be ‘rational intuitions’ of the sort posited by methodological rationalists, for which Cappelen (2012) found no evidence in texts generally accepted as paradigms of intuition-based philosophising. [↑](#footnote-ref-1)
2. 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. [↑](#footnote-ref-2)
3. 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. [↑](#footnote-ref-3)
4. In another much-used sense – which we will not employ – an ‘automatic inference’ occurs whenever a concept ‘activates’ another, e.g. in priming experiments (Fn.6). [↑](#footnote-ref-4)
5. Extant proposals of validating explanations, which mainly draw on cognitive and social psychology, do not yet address this task. See Kornblith (2015) for a critical review. [↑](#footnote-ref-5)
6. Accordingly, links between concepts are investigated through *priming experiments* (Lucas 2000), where participants are first presented with a ‘prime’ word, sentence, 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. Activation – and a link to pass it on – is inferred from shorter response times for suitable prime-probe pairs (‘bank’-’money’ vs ‘bank’-’honey’). [↑](#footnote-ref-6)
7. Kahneman (2011, part I) provides an elegant introduction. Harley (2014, ch.11) offers an up-to-date textbook treatment. [↑](#footnote-ref-7)
8. The ‘cloze probability’ or frequency with which participants use a word to complete a sentence-frame, is the most common measure of the strength of association, which is additionally measured through plausibility rating, and listing tasks (McRae et al. 1997). [↑](#footnote-ref-8)
9. This rule licences a wide range of inferences, which may be duplicated by different cognitive process (Levinson 2000, 117-18), not only those discussed below. We lack the space to relate psycholinguistic findings to influential work on automatic pragmatic inferences in theoretical linguistics (Jaszczolt 2011, Recanati 2010). [↑](#footnote-ref-9)
10. At this stage, associative processing in semantic memory treats semantic and stereotypical features the same (Banaji and Hardin, 1996). The best way of drawing this important distinction at this low level, we believe, is to consider the interaction of semantic and episodic memory: Individuals with more stereotype-consistent first-hand experiences (episodic memory) tend to show stronger activation of stereotypical features (Blair and Banaji, 1996; McCauley and Thanga Velu, 1991); in contrast, semantic relations tend to be more resistant to the influence of particular experiences. [↑](#footnote-ref-10)
11. Strength of association is also influenced by prototypicality (Rosch 1978), and hence also depends upon cognitive principles of abstract categorical organisation (Giora 2003). [↑](#footnote-ref-11)
12. Cp. this recent enquiry: ‘Is there a general term covering blindness, deafness, numbness, etc.?’, Sean Power, April 10th, 2013. See <http://listserv.liv.ac.uk/archives/philos-l.html>. [↑](#footnote-ref-12)
13. *Oxford English Dictionary*, online version June 2012, entries ‘perceive’ and ‘aware’. [↑](#footnote-ref-13)
14. 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, forthcoming), all seven verbs are jointly categorised as ‘appear words’ differing, basically, only in degree of generality or the perceptual sense invoked. [↑](#footnote-ref-14)
15. This is the uncontroversial first half of the default-and-challenge model of justification (Williams 2001, 25). We will *not* rely on the model’s further, more controversial, contention that endorsement by common sense makes acceptance the *appropriate* default response to a claim, in the absence of positive reasons for doubt. [↑](#footnote-ref-15)
16. E.g. Broad (1923, 240): ‘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.’ [↑](#footnote-ref-16)
17. 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, and the one argument he appears to give for the principle (Broad 1923, 240f, cp. Smith 2002, 36-7) explains why it may strike thinkers as plausible but is patently circular when uncharitably read as a justificatory argument. [↑](#footnote-ref-17)
18. Gibbs and Colston (2012, 48-54) provide a pertinent classification of metaphorical uses. [↑](#footnote-ref-18)
19. They are not happily captured by familiar philosophical explications of ‘phenomenal’ and ‘epistemic’ uses of ‘looks’ (Chisholm 1957, Jackson 1977, Maund 1986, Brogaard, forthcoming), which we therefore set aside. [↑](#footnote-ref-19)
20. Cp. the usage note for ‘seems’, *Oxford English Dictionary* (2nd ed. 1989, online version June 2012). In this first-person use (where reference ‘to me’ may be implicit), preference of ‘seems F’ over the simpler ‘is F’ warrants the inference of ‘doubt-and-denial conditions’ (Grice 1961). [↑](#footnote-ref-20)
21. *OED*, online version June 2012, entries ‘apprehend’ and ‘cognizance’, my italics. [↑](#footnote-ref-21)
22. 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 (cp. Broad 1923, 239-40, 248). [↑](#footnote-ref-22)
23. Where the premise does not state this explicitly, we would infer it automatically, in line with the I-heuristic, from the noun ‘coin’. [↑](#footnote-ref-23)
24. Here and below, numbered items are taken from the questionnaire used in our study. [↑](#footnote-ref-24)
25. We may *not* infer, however, that the more consistent the preference of one verb over the other is, the more its strength of stereotypical association with P exceeds that of the other: Even a small difference can decide preferences about all pairs. (Though the more pronounced the difference is, the more likely it is to ‘drown out’ coincidental other differences in stereotypical association, across several sentence-contexts.) The measured variable (consistency of preference) admits parametric tests, as it is continuous, has a 0-point (50/50 preference, e.g. 3/6 items), and equal intervals on the scale represent equal differences in the property measured (a 5/6 preference for A- over B-sentences is equal to a 1/6 preference in the other direction). [↑](#footnote-ref-25)
26. Distributional-semantic analysis of the intransitive use of ‘seem’, ‘appear’, and ‘look’ provide independent evidence that these verbs – and ‘seem’ and ‘appear’, in particular – are strongly stereotypically associated with the doxastic patient-property targeted (Fischer et al. 2015). This rules out the possibility that the predicted pattern is due to ‘appear’ and ‘look’ lacking such association. [↑](#footnote-ref-26)
27. In the Leech et al (2001) corpus, ‘see’ appears 1920 times, ‘perceive’ 29, ‘aware’ 86, ‘behold’ 3, and ‘hear’ 367 times. Also our other critical verbs vary in frequency: ‘look’=1151, ‘seem’=624, ‘appear’=307. [↑](#footnote-ref-27)
28. This excludes a further confound (cp. section 2): Violations of the Maxim of Manner (Grice 1989) are similarly perceived as sounding less good or idiomatic (Levinson 1983). Lack of reference to idiomatic preference hence speaks against the suggestion that observed preferences for more precise verbs (‘see’ or ‘hear’) over more generic verbs (like ‘perceive’) is due to perceived violations. [↑](#footnote-ref-28)
29. In one look/appear- and one look/seem-item, the inconsistency was with a generally familiar background fact that remained implicit, e.g.: ‘In the dark, everything looks/seems grey’ – we take it as consensus fact that objects don’t lose their colour in the dark. Item analyses showed that these items did not behave substantially differently than items with overt cancelling sentences (see below). [↑](#footnote-ref-29)
30. In one item, the incongruent text preceded the critical verb: ‘I still don’t know what it was I perceived/saw in the twilight’. The mean for this item (.98) was nearly identical with the overall mean for perceive/see items (.97), which provides no reason to believe that cancellation order affects responses. [↑](#footnote-ref-30)
31. The difference between frequency-reversed and frequency congruent items was statistically significant *t*(7) = -2.35, *p* < 0.05, despite the limited sample. [↑](#footnote-ref-31)
32. One sample *t*-tests confirmed that philosophers showed no preferences (i.e., not different from chance) between ‘appear’ and ‘look’ *t*(7) = 0.51, *p* > .05 or between ‘appear’ and ‘seem’ *t*(7) = -0.94, *p* > .05. However, the preference for ‘look’ over ‘seem’ was significantly greater than chance *t*(7) = 9.35, *p* < .05. [↑](#footnote-ref-32)
33. Paul Engelhardt contributed to design and material development for the experiment reported in section 4, and undertook the statistical analysis. Eugen Fischer undertook the remaining research. For helpful comments on previous drafts and closely related material, he is 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, Cambridge, and Erlangen (all 2014). Both authors would like to thank Mark Curtis for assistance with data collection and entry. [↑](#footnote-ref-33)