Metacognition and Inferential Accounts of Communication

Nicholas Allott

This is a postprint (author's final draft, after refereeing). The paper was published as:

Allott, N. (2020). Metacognition and inferential accounts of communication. In T. Chan & A. Nes (Eds.), *Inference and Consciousness* (pp. 125–148). London: Routledge.

Please send any comments or questions you have to: nicholas.allott@gmail.com

Abstract:

Utterance interpretation is widely seen as an inferential achievement, and utterance interpretation resembles 'system 2' reasoning in some respects: the inferences are generally warranted and unencapsulated. But in normal, smooth communication, they are quick and seemingly effortless and thus more akin to paradigm unconscious 'system 1' inferences. Resolving this tension is a goal for any cognitively realistic account of utterance interpretation. This chapter argues on theoretical and empirical grounds for a minimalist kind of metacognition whereby a mental process is unconsciously monitored and controlled by another, perhaps without the latter metarepresenting the former. Cognitively realistic inferential theories of utterance interpretation require there to be such feedback, even in normal smooth communication. Also, two separate sets of experiments, reviewed here, show that feedback occurs in comprehension without hearers being aware of it. A potential objection concerns levels of explanation (in David Marr's sense): that some process is inferential seems to be a claim at Marr's functional level, while claims about the way that feedback works in utterance interpretation are at the algorithmic level. This chapter notes that facts at the algorithmic level constrain facts at the functional level and suggests that processes for implementing abductive inference must have monitoring and control.

Keywords: pragmatics, heuristics, metacognition, inference, utterance interpretation

1. Introduction

This chapter aims to show that metacognitive processes of monitoring and control play a role in utterance interpretation, even when the process is smooth, automatic, and unreflective. More precisely, I argue that unconscious monitoring and control – what Joelle Proust (2013) has referred to as 'procedural metacognition' and Shea et al. (2014) 'system 1' metacognition – is a central feature of utterance interpretation. I support my claim with theoretical arguments and evidence from two empirical paradigms.

This is intended as a step towards understanding how utterance interpretation can both be an inferential achievement (Grice, 1989; Wilson and Sperber, 2012), and a largely subliminal, automatic process, phenomenologically very different from full-blown reasoning. In the terms of dual-processing accounts of cognition (Sloman, 1996; Stanovich and West, 1998; Evans, 2003; Evans and Frankish, 2009), utterance interpretation resembles 'system 2' reasoning in some respects: the inferences are generally warranted, and apparently unencapsulated. But in normal, smooth communication, they are typically quick and seemingly effortless, thus in this respect more akin to paradigmatic unconscious, 'system 1' inferences. A good theory of utterance interpretation should shed light on this tension and in my view this will require an account of the role played by unconscious, automatic feedback mechanisms, for reasons I explain in Section 2, below.

Feedback here is understood in terms of processes that monitor and control other processes, and monitoring and control are the essential ingredients of what has become known as metacognition.[1] Metacognition, often loosely characterized as 'thinking about thinking', has been studied in several separate literatures in psychology over several decades which have recently attracted attention from philosophers. I briefly describe one classic experimental paradigm in Section 2.3.

Several lines of research suggest that metacognitive feedback often shapes behavior unconsciously (Kentridge and Heywood, 2000; Spehn and Reder, 2000; Paynter, Reder, and Kieffaber, 2009; Shea et al., 2014). This view may have a paradoxical air if one sees metacognition as bringing first-order thought processes to second-order awareness by metarepresenting them. However Proust has forcefully argued that metacognition does not require metarepresentation, and that the basic form of metacognition is what she terms procedural metacognition (Proust, 2013). That is, there are mental processes that are dedicated to monitoring and controlling other mental processes without representing them, or at the least without representing them as such. Crucially for my purposes here, it is plausible that much of this procedural metacognition is unconscious. Dedicated subpersonal metacognitive processes track how 'first-order' mental processes are doing in their tasks, so that subsequent processing can be guided by that performance, even when there is no awareness at the personal level that anything like this is happening.

Shea et al. (2014) also argue that unconscious metacognition exists, although unlike Proust they define metacognition as representational. They argue for a distinction between system 1

metacognition, which is unconscious and is dedicated to the control of processes within one agent, and system 2 metacognition, which brings properties of processes to conscious awareness so that this information can be shared with con-specifics.

What is important for current purposes is that on either of these views, it is likely that many processes that have not been regarded as metacognitive because they lack conscious metacognitive phenomenology will turn out to involve metacognition.[2] Here I argue that this includes normal, smooth utterance interpretation.

Various rival accounts of utterance interpretation are current in linguistic pragmatics. The most prominent share two central assumptions: i) that utterance interpretation is a type of inference to the best explanation about certain intentions of the speaker; and ii) that it is performed by, or according to, a specialized heuristic or heuristics.

Paul Grice's well-known work on meaning (Grice, 1957) and on conversation (Grice, 1975; Grice, 1978) suggested an inferential model of communication (Sperber and Wilson, 1986, pp. 21ff.). This was a major shift from most previous views of communication, which focused on the role of language seen as a code for transmitting thoughts.

In a pure coding/decoding model, communication is simply the transmission of a meaning – the message – by encoding it in language or some other code. The idea is that the transmitter encodes and transmits the message as a linguistic signal, which the receiver then decodes. There is some truth to this. Language is a code in the sense that the relationship between word-types and what they mean is (mostly) arbitrary. So linguistic parsing is indeed a form of decoding.

However, it is now well-established that what a speaker conveys by an utterance ('speaker meaning' in the prevailing terminology) is not determined by the linguistic material that she utters. The inferential model accounts for this by treating the parsed linguistic form of each utterance as merely an input to the hearer's inference about what the speaker meant by uttering it. The inference in question is from some observed behavior (an utterance) to an explanation for that behavior in terms of the speaker's intentions to convey something (the speaker's meaning). This inference may draw on the linguistic and extra-linguistic context and on personal and cultural background knowledge. On all these points, the major tendencies in linguistic pragmatics and philosophy of language are in agreement – that is, Griceans such as Kent Bach, neo-Griceans such as Levinson, and relevance theorists (who are best understood as post-Griceans) such as Sperber and Wilson.[3] A second key assumption, shared by neo-Griceans and relevance theorists, is that the process can be seen as the operation of a heuristic or heuristics. (Sperber and Wilson, 1986, p. 45; Levinson, 2000, pp. 30ff.) This chapter's thesis is a claim about properties of this utterance interpretation heuristic (or suite of heuristics): namely that it involves subliminal metacognition.

There are two prongs to my argument. One, dealt with later in Section 2, is that very general considerations about processing suggest that utterance interpretation must be steered by

feedback of this sort. This line of argumentation depends on some general assumptions about mental computation and heuristics and on framing utterance interpretation as inference performed by a heuristic, which I motivate briefly; I cannot mount here a full-dress defense of inferential pragmatics or of basic assumptions about the computational character of mental processes.

The other prong of the argument is provided by two sets of experimental results that I discuss in Section 3. I argue that they show that tacit feedback occurs in utterance interpretation. The way I proceed is to set out some ways that subliminal metacognition could work in communication and then present the empirical results. These suggest that there is subliminal metacognition in i) suppression of activated word senses; ii) slowing of reading speed when there are problems integrating the interpretation of an utterance with the model of the context.

I also discuss whether these effects could be accounted for without postulating metacognition. Here I consider Recanati's bottom-up accessibility-driven account of recovery of what is said, which is an attempt to show that this component of speaker's meaning can be arrived at without abductive inference. I argue that his account is non-metacognitive and also highly implausible.

This discussion raises the question of how claims about low-level properties of a heuristic are relevant to questions about what kind of task the heuristic performs, and I make some tentative suggestions, appealing to David Marr's distinction between different levels of explanation in cognitive science.

2. Theoretical Discussion

The inferential processing involved in utterance processing is not in general full-blown reasoning or reflective inference, thus differing from what at least some philosophers mean when they use the word 'inference'. In his recent paper on inference, Paul Boghossian writes:

> By 'inference' I mean reasoning with beliefs. Specifically, I mean the sort of 'reasoned change in view' . . . in which you start off with some beliefs and then, after a process of reasoning, end up either adding some new beliefs, or giving up some old beliefs, or both. (Boghossian, 2014, p. 2)

He adds:

> I am interested in reasoning that is person-level, conscious and voluntary, not sub-personal, sub-conscious and automatic, although I shall not also assume that it is effortful and demanding. (Boghossian, 2014, pp. 2–3)

Utterance interpretation differs from this in that it is typically involuntary and subconscious: we do not have any choice about whether we perform it, and we are not occurrently aware of

doing so, only of its result.[4] At least, that is so for cases where everything goes smoothly, which I will be focusing on for two reasons.

The first reason is that I take that to be the normal case. What we seem to be aware of typically is that the speaker is stating p and/or implying q, promising to do r and so forth. The speech sounds and words that are uttered are also available to awareness, although not – at the personal level – represented in the detail, or with the structure that they are parsed to have by subpersonal processes.

We are typically not aware of having to infer what the speaker meant from the sounds she made, although this must be happening, since the input to the process is a stream of speech sounds, and a) the sound stream does not possess either linguistic or speaker meaning intrinsically, and b) one sound stream typically corresponds to many possible linguistic meanings and an open-ended number of speaker meanings.

We generally interpret utterances without noticing that we have (for example, and inter alia) assigned reference to indexicals, chosen senses for ambiguous expressions and reconstructed what was meant by the use of degree adjectives and possessives. For example, a utterance of the sentence in (1a) might be an assertion of something like the proposition in (1b), and in the right context that proposition could and generally would be arrived at without the hearer noticing that there was inferential work involved.

(1a) Mary: His book is too long.

(1b) The book that John wrote is too long for Mary to expect her students to read.

My second reason for focusing on utterance interpretation that is phenomenologically effortless is that it is a harder case for the view that I am arguing for. It would be no great surprise to find metacognition involved in reflective, voluntary, phenomenologically effortful utterance interpretation. It is much less obvious that it plays a central role in normal, smooth utterance interpretation. That is therefore the more interesting claim.

A distinction that is relevant here is between what Sperber and Mercier call intuitive and reflective inference.[5] The latter is processing whose purpose is to provide a person with (consciously available) reasons, in which, '[y]ou are paying conscious attention to the relationship between argument and claim, or premises and intended conclusions' (Sperber and Mercier, 2012, p. 375).

Utterance interpretation does not in general involve reflective inference, so understood. Hearers are arguably able to become aware of the way that their conclusions about speaker meaning are supported by what the speaker uttered, but such inferential links are not something that hearers typically attend to or become aware of.

One might wonder what is left of the notion of 'inference' once awareness and reflection have been stripped away. I suggest that utterance interpretation is inferential in approximately (and at least) the following sense:

> An inference, as the term is used in psychology, is a process that, given some input information, reliably yields as output further information that is likely to be true if the input information is. (Sperber and Mercier, 2012, p. 371)

In sum, utterance interpretation is one type of (mostly) involuntary 'change in view', a process that adds beliefs about utterance content by drawing warranted, but non-demonstrative, conclusions from the input.

2.1 Utterance interpretation seems ill-structured

Utterance interpretation, like other inference to the best explanation, is prima facie an ill-structured problem. That is, very roughly, it is a problem for whose solution there is apparently no failsafe algorithm (Simon and Newell, 1958; Simon, 1973; Sperber and Wilson, 1986, p. 45; Simon, 1997, p. 128; Allott, 2008). I would argue that it is ill-structured in at least two ways.

The first applies to abductive inference in general. It is unclear what information is relevant, so it is hard to see how an algorithm could decide with certainty what information to consult: in computational terms, the search-space is indefinitely large. This is what Jerry Fodor calls isotropy, and it is one reason that he has argued that there is no theory of central cognition (Fodor, 1983; see also Allott, 2019, on this argument applied to pragmatic theory). Just about anything could turn out to be relevant in inference to the best explanation. Drawings of rabbits on ancient pots may provide evidence about the astrophysics of supernovae, to take a well-known example (Robbins and Westmoreland, 1991; Antony, 2003).

Pragmatic inferences are responsive in principle to just about any information (Sperber and Wilson, 1996), as illustrated by examples like those in (2):

(2a) John was arrested by a policeman yesterday; he had just stolen a wallet. (Recanati, 1993, p. 265)

(2b) John was arrested by a policeman yesterday; he had needed one more arrest to qualify for an end-of-year bonus.

(2c) John was arrested by a policeman yesterday; he had just taken a bribe.

The examples in (2) illustrate the intricate dependence on world knowledge of the assignment of referents to indexicals. The hearer will probably take 'he' to be anaphoric on 'John' in (2a) and on 'a policeman' in (2b). The assignment of reference to 'he' in (2c) could go either way, depending on the hearer's estimate of the relative honesty of John and the local police force.

Disambiguation, enrichment, implicatures etc. are similarly sensitive to non-linguistic information.

The other way in which utterance interpretation (like much other abductive inference[6]) seems ill-structured, but which is hardly discussed in the pragmatics literature, is that there is apparently no simple test to show that a putative solution is the right one (Allott, 2008, pp. 179–180). Suppose that at some stage in the process, the hearer's pragmatic faculty has somehow generated a candidate interpretation of an utterance. How can it tell that this is the right interpretation, or at least the best one that it can generate? Here, one should compare with a well-structured problem like solving an equation in two variables, where once you have a putative solution it is simple to check whether it really is one: simply plug the hypothesized values of x and y into the equation and if the two sides come out equal then you have found a solution.

Elsewhere I have suggested that there are several properties that we should expect to be possessed by utterance interpretation given that it is a process that deals rapidly with an ill-structured problem (Allott, 2008, ch. 5). Here I focus on just one: subliminal monitoring and control, or 'procedural metacognition'.

2.2 Metacognition and communication

I think that we can distinguish at least three levels at which there may be monitoring and control in communication. My concern in this chapter is with only one of them: monitoring and feedback internal to the hearer in normal, smooth utterance interpretation.

This needs to be distinguished from monitoring that disrupts smooth processing, taking the hearer into a qualitatively different, occurrently conscious, somewhat reflective process, which feels effortful. Robyn Carston (2010) has argued that there are two different 'routes or modes of processing' in metaphor understanding, one of which is 'rapid' and 'local', the second being 'more global [and] reflective' (Carston, 2010, p. 295). This is plausibly true of utterance interpretation more generally, and intuitively, the more conscious, reflective process comes into play in various situations that roughly divide into two types:

a) Where there is more to unpack than one would normally get out of an utterance, e.g., in reading a rich text such as a Henry James novel, or when one notices a pun or a *double lecture*.

b) Where what the speaker wanted to communicate is not well packaged from the hearer's point of view, as for example, when you notice that the speaker used a wrong word or infelicitous expression. Conscious effort may then be required to arrive at even one plausible interpretation.

My concern in this chapter is to show that there is metacognition even in cases where this sort of thing does not happen. My thesis amounts to the claim that the presence of monitoring and control does not entail that we are concerned with occurrently conscious metacognition.

It is also important to distinguish between the kind of metacognition that this chapter focusses on and monitoring of the hearer's comprehension performed by the speaker. Speaker monitoring of hearer comprehension is an aspect of what Proust refers to as 'conversational metacognition', which she defines as 'the set of abilities that allow [a . . .] speaker to make available to others and to receive from them specific markers concerning his/her 'conversing adequacy' (Proust, 2008, pp. 329–330).

Speakers gauge whether hearers are paying attention to them, particularly in the normal case of face-to-face conversation, by monitoring facial expression, gaze direction and various forms of feedback such as nodding, saying 'Mm hmm', 'I see' etc. This leads them to send signals about their level of commitment to what they are saying, their wanting to 'hold the floor' or to let the other person have a turn at speaking and so on (Clark and Wilkes-Gibbs, 1986; Clark, 1994; Fox Tree and Clark, 1997; Clark and Fox Tree, 2002; Clark and Krych, 2004; Allott, 2016, pp. 501–503).

Such monitoring and feedback is surely metacognitive. It also appears to be ubiquitous in face-to-face conversation. However, it cannot be essential to verbal communication, given that this can occur in situations where such feedback is not possible, as in answerphone messages as well as almost all written communication. Here I am concerned instead with metacognition that is internal to the addressee of an utterance, which I argue is central to utterance comprehension.

2.3 Metacognition and awareness

It is necessary to illustrate in a little more detail what psychologists mean by 'metacognition'. In one common experimental paradigm for investigating metacognition, subjects are presented with a series of tasks under time pressure and can choose at each trial either to perform the task or to opt out of it. The task might be to assign a stimulus correctly to one of two previously learned categories, for instance, to say if a presented visual array is 'sparse' or 'dense'. The crucial finding is that people opt out preferentially from tasks they are less good at: in this case stimuli that are close to the boundary between the categorizations.

This metacognitive ability is often accompanied by so-called noetic feelings, which one might think of as feelings that could be informally glossed as this task is easy/difficult, or I know/don't know the answer to this one. (Note that I do not mean by these glosses to commit myself to the claim that noetic feelings have conceptual content.) Now it is very often assumed in work on metacognition that these noetic feelings are causes of (or at least causally implicated in) the behavior that paradigmatic metacognition tasks investigate.[7] It is essential for my thesis, though, that internal feedback does not always or necessarily come with such feelings; there is some fully subliminal monitoring and control.

As noted earlier, there has been some discussion of this question in the metacognition literature. One obvious logical possibility is that in some or all cases where there are noetic feelings they are epiphenomenal: the feelings and the performance are both due to the

metacognitive mechanisms, but the causal path to performance does not or need not go via the feelings. Asher Koriat argues that '[s]ubjective experience is based on an interpretation and attribution of one's own behavior, so that it follows rather than precedes controlled processes' (Koriat, 2007, p. 315).

Whatever the truth about cases where noetic feelings are present, I agree with Kentridge and Heywood when they write:

> There is nothing inherent in metacognitive regulation that demands consciousness. Metacognitive and executive processes serve to select and deploy methods for dealing with events and to assess the utility of those methods. The presence of a self-referential loop, a system which assesses its own performance and adapts accordingly, might tempt us to infer that such processes necessarily elicit awareness. Feedback loops are ubiquitous in biology and, of themselves, do not seem to be grounds for invoking consciousness. (Kentridge and Heywood, 2000, p. 308)

There is some empirical evidence that in utterance interpretation, monitoring and control is separate from reportable awareness of difficulty and anomaly, which I discuss in Section 3. First, though, I set out the case that theoretical considerations imply that utterance interpretation requires this sort of monitoring and control.

2.4 Theory-driven argument for subliminal metacognition

It seems a virtual conceptual necessity to see interpretation of verbal utterances as a suite of processes that construct an representation of utterance meaning on the basis of speech sounds. Like other pragmatic theorists, I assume that this processing can be factored into two parts:

I) A linguistic front-end which a) segments the stream of sound into phonemes and morphemes and b) assigns a syntactic structure to the utterance (parsing);

II) A conceptually distinct process or processes, 'pragmatic inference', which takes this linguistic material as input and arrives at utterance content.

Pragmatic inference is described here as (merely) 'conceptually' separate from linguistic parsing because it is widely assumed that in practice there are interactions between parsing and pragmatics, including 'top down' effects. One such is suppression, which I discuss in Section 3.1.

As discussed earlier, utterance interpretation is typically fast and automatic. Therefore, given very general assumptions about costs of computation, it seems reasonable to assume that there is limited information search (to use a term from the literature on simple heuristics: e.g., Todd and Gigerenzer, 2000, pp. 729–730): a great deal of information that might be relevant is not processed and not even recalled from memory.

A further reasonable assumption is that the system is not calculating for each item of information that could be processed whether it would be worth considering. That approach, called 'optimization under constraints', will often be more computationally expensive. In general, to calculate for each piece of information whether it is worth processing and to what depth is 'a more complex . . . procedure that includes the basic decision problem plus the problem how many costly resources to allocate to that original problem.' (Vriend, 1996, p. 278. See also Todd and Gigerenzer, 2000, pp. 729–730; Allott, 2008, pp. 170–172.)

Such considerations strongly suggest that there is a kind of metacognition that 'opts out' from lines of thought that are not progressing well, and opts in to just one or a few lines of thought that seem more promising. This would (on average) steer pragmatic processing towards recall and processing of information that would be cognitively worthwhile, and towards processing it in ways that would be profitable. Given the speed and seamless phenomenology of (much) utterance interpretation this metacognition must normally operate below the level of consciousness.

This kind of model is fundamental to relevance theoretic pragmatics, although as far as I am aware the term 'metacognition' has not been used in this literature until now. One of relevance theory's fundamental aims is 'to describe how the mind assesses its own achievements and efforts from the inside, and decides as a result to pursue its efforts or reallocate them in different directions' (Sperber and Wilson, 1986, p. 130; see also Sperber and Wilson, 1996; Sperber and Wilson, 2002; Allott, 2008). Something similar is implicit in the use of the term 'heuristic' in neo-Gricean pragmatics (Levinson, 2000, pp. 30ff.), although there has been less attention in that tradition to the details of cognitively realistic theories of utterance interpretation.

3. Types of Monitoring and Control in Utterance Interpretation

In this section I give empirical evidence that two types of monitoring and control, suppression and guided resource allocation, do indeed take place. One way that monitoring and control could feature in utterance interpretation is suppression of senses. That is, when an interpretation is beginning to be favored, rival candidates are actively demoted. I discuss evidence for this in Section 3.1.

A second way is preferential allocation of resources guided by monitoring of the success or failure of the ongoing interpretation process. An obvious possibility is that more effort and time is put into interpretation when no overall interpretation is successfully reached. There is considerable evidence for this, and some evidence for it happening subliminally, as I discuss in Section 3.2.

3.1 Metacognition and suppression of word senses

There is evidence that suppression of unintended senses of words occurs in utterance interpretation. In this section I first explain the phenomenon and then present evidence that suggests that such sense suppression is an unconscious metacognitive process.[8]

It is known that word-senses and core meaning features of words are activated regardless of whether the sense/feature coheres with the context. This is known as 'priming' and is seen in experiments on the effect of utterances of ambiguous words and metaphors. Classic examples are in (3) and (4):

(3) The man found several bugs in his room.

(4) My lawyer is a shark.

Hearing (3), both senses of the word – covert listening device and small invertebrate – are activated, as we know from experiments which test how fast participants are to respond with word or non-word to related words such as 'spy' and 'ant'. Crucially, both are primed even in contexts where only one sense of 'bug' is plausible (Meyer and Schvaneveldt, 1971; Schvaneveldt and Meyer, 1973). Similarly, the example in (4) is a metaphorical use of 'shark', but it is known that core features are activated, e.g., in this case <FISH>, even when they are incompatible with the metaphorical reading.

It is also known that activation of a word-sense or feature is typically followed by decay of that sense. This can be shown by probing at different times after the initial activation. The priming effects that indicate activation gradually decrease. But it has been shown that the drop-off in activation is faster than in normal decay for both the non-target sense in disambiguation cases and the feature that clashes with the correct interpretation in metaphor. This is standardly interpreted by researchers in this field as suppression of the unrelated feature or word sense (Neely, 1976; Tanenhaus, Leiman, and Seidenberg, 1979), as the following summary of the literature describes:

> The results of these experiments showed an early activation of target words related to both meanings of the homonym, which was interpreted in terms of an automatic, exhaustive process of spreading activation of associates. However, the activation of the contextually inappropriate meaning dropped as early as 200–300 ms from the offset of the ambiguous word. This pattern of results was interpreted as showing active suppression of the irrelevant reading of the ambiguity, given that passive decay should take considerably longer. (Rubio Fernández, 2007, pp. 353–354)

It is important to see that while we can probe this activation and suppression in cases of ambiguity and metaphor, it probably occurs in all utterance interpretation, and perhaps in thought more generally. Some evidence comes from work on schizophrenia. Schizophrenic patients 'often jump from one subject to another based on the sounds or associations of words

they have uttered' (Covington et al., 2005, p. 87). This has been linked to excessive priming or impaired control of priming (Kuperberg, 2010, pp. 582–3). Such problems with priming may also be connected with the loss of control of the train of thought which is a primary symptom of schizophrenia: patients with thought-disorder have been found to have increased priming relative to non-schizophrenic controls (Pomarol-Clotet et al., 2008). To the extent that these problems are due to lack of control of activations of senses they support the claim that such control is a feature of normal language processing and perhaps of thought more generally.

Priming and suppression of senses are certainly not conscious processes, neither occurrently nor in the sense of being available. This is obvious introspectively: we only know that this sort of thing is going on because of the experimental evidence. Moreover both the activation and the suppression are too fast to be under conscious control:

> since controlled, attentional processes take 400–500 ms to operate . . . although the meaning selection process must be context-sensitive (unlike the early spreading activation phase), it operates in an almost automatic way . . . . This would explain why hearers are usually unaware of having encountered a homonym in a disambiguating context. (Rubio Fernández, 2007, p. 353)

Suppression of a sense seems metacognitive. Why think so? The argument is that we know that there is a natural outcome: decay. We assume that is what would happen in the absence of control. When we see suppression rather than decay, this is therefore a sign of control. What is more, the control seems to be directed by monitoring of the way that the process is going, since it is unintended senses that are being shut down.

A possible objection is that the experimental results could be accounted for by a suppression process that operates automatically once a sense is selected. But I think that this objection is misconceived because such a process would be metacognitive. It would involve monitoring and control, in this sense: there would have to be sensitivity to the success of the first-order process (monitoring) and then as a result, changes to the first-order process (control).

A second possible objection is more cogent. A critic could argue that what those in the field call 'suppression' is actually a bottom-up effect of context (perhaps acting via activation of concepts) interacting with the activation effects caused by the words in the target sentence. The idea is that the activations caused by words happen first, followed by an inhibition from context. On this view the accelerated decay of activation in cases of poor fit with context could be accounted for without any need to postulate monitoring and control.

This is very much like the view François Recanati has advocated as an explanation of pragmatic 'garden-path' effects. These (which are a theoretical possibility rather than a well-established phenomenon) are cases where the first interpretation constructed is not the one ultimately accepted (Recanati, 2004, pp. 32ff.). Suppose a speaker uses the word 'bank' in a context in which the financial-institution sense is highly accessible, but where only interpreting the word as river bank makes sense. It is plausible that an interpretation containing the financial sense is

constructed and then rejected or superseded. If so, this is a pragmatic garden path. Dan Sperber argued that such cases would show that interpretation is not driven only by accessibility of senses: the most accessible interpretation can be rejected if it leads to an overall interpretation that is unsatisfactory. Recanati's reply is that such cases could be accounted for solely in terms of accessibility, if we assume (e.g.,) that lexical priming is faster than activation from the context: then an initial interpretation could be superseded by a competitor that simply takes longer to emerge, and there is no need to postulate any top-down evaluation of the interpretations.

This line of argument raises some difficult questions which I return to in Section 5. Here I offer two responses. First, such an opponent would be asking us to believe in miracles. That is, he would be asking us to accept that in successful communication all the activations (from the words in the utterance, plus features of the context) always happen to add up to making the speaker's intended meaning the most accessible one. One can see how this might sometimes work out, but why should we think it always does? It is worth noting that even Recanati makes that claim only about recovery of what is said, and not about other pragmatic processing such as arriving at implicatures. My second response is to agree with Kentridge and Heywood's point, quoted in Section 2.3 above. Given that feedback loops are ubiquitous in biology we should expect to find that mental processes exist to keep track of the success or otherwise of other mental processes and to shut down unnecessary activation (ultimately, that is, to save energy).

3.2 Metacognition and resource allocation

There are further empirical findings that lend support to the claim that there is subliminal feedback in utterance interpretation. They come from a series of experiments that aimed to probe two abilities in development and their relation to each other: sensitivity to textual anomaly indexed by reading time and conscious awareness of comprehension difficulties (Harris et al., 1981). There were two groups of participants, aged eight and eleven years old respectively.

In the experiment the participant reveals a short story line by line as she reads it silently to herself. There are two conditions, which only differ in which of two titles is presented as the first line. In each condition, one line of the text is anomalous, but there is nothing intrinsically odd about that line. The anomaly is purely a result of encountering that line in the context of the title, as the following example materials demonstrate. The anomalous line in the first scenario is labelled (i) while the one labelled (ii) fits the context, and vice versa in the second scenario. This design allows the two conditions to be compared in order to control for all effects on reading time other than the anomaly.

Title 1: Together on the boat

Title 2: The toy boat

Charles has a sailing boat.

He shows it to his friend.

'Do you like it?' asks Charles.

'Please don't drop it'. (i)

The two boys climb aboard. (ii)

The little boat is now rolling on the water.

The wind is blowing in the sails.

Then the boat is off the shore.

(Harris et al., 1981, p. 216)

There are two crucial findings. First, both eight-year-olds and eleven-year-olds read the anomalous line more slowly than the appropriate line in all stories, with no statistically significant difference between the two groups in this respect. This indicates that, as expected, both groups were affected by textual anomaly, and in fact were affected to an indistinguishable degree. Secondly, eight-year-olds were significantly less good at picking out the problem line when asked to identify it after reading the whole text, and when successful were also slower to identify it. The authors say that this 'suggest[s] that they had not 'registered' it during their initial reading of the story.' (Harris et al., 1981, p. 219), and that if they found it at all, they typically did so by re-reading the text (which they had in front of them at this stage).

The obvious objection that the eight-year-olds' difficulty might be due to memory limitations was made less plausible by a second run in which participants were also tested for recall of the lines, including the problematic line. No significant difference was found between the age groups. Therefore the authors conclude that the eight-year-olds' poorer ability to pick out the problem line was not well explained in terms of their having lost track of which line was problematic after having noticed the anomaly during their initial reading of the story.

This experiment indicates that the time (and presumably effort) put into utterance interpretation was modulated in response to the anomaly in both age groups, and that the ability to modulate processing in this way does not depend on conscious awareness of the anomaly. However, a possible objection is that the increased reading time does not show that there is any second-order monitoring and control of the first-order comprehension process. Rather, it may be that reading takes longer in the anomalous cases because they are harder to understand.

Here is my response to this objection. Consider why it takes longer to read the anomalous line. The explanation, I suggest, is that the anomaly is detected at some level, and more resources (including, at least, longer time) are devoted to processing. Recall that the anomalous line is not anomalous in itself, but only against a particular context. If the anomaly were not detected at any level, why should the participant read more slowly? Participants could just read through the sentences at normal speed, understanding each sentence, but not integrating the meanings of the sentences at any higher level. Note that this is not a purely theoretical possibility: there is some evidence that younger (six-year-old) participants do just this (Markman, 1977). They may understand the individual words and sentences but apparently do not try to build a consistent mental model for the text as a whole.

Harris et al. conclude that

> for the age period under consideration [between 8 and 11 years old], there is evidence that the improvement in comprehension monitoring can be attributed to changes in the capacity to notice or interpret internally generated signals, rather than to any differential frequency in the generation of those signals. (Harris et al., 1981, p. 219)

That is, for both the eight-year-olds and the eleven-year-olds there was internal monitoring for anomaly in the process of 'constructive interpretation', and this monitoring resulted in changes to the first-order process (i.e. control). However, only in the eleven-year-olds did it reliably give rise to something that was available to conscious recall and report. This is evidence for subliminal metacognition in utterance interpretation in eight-year-olds.

What, if anything, can we conclude about the eleven-year-olds and about mature utterance interpretation? Earlier in this chapter I suggested that performance and noetic feelings may have a common causal basis without noetic feelings being causally responsible for spontaneous performance. Given that eight-year-olds and eleven-year-olds slow down to the same degree when they encounter anomaly, these experiments suggest that the noetic feelings that the eleven-year-olds have some access to are not what drives their spontaneous reading performance. In other words, subliminal monitoring and control takes place during utterance interpretation for everyone above a certain age, modulating reading speed. Some ability to consciously 'dip into' the internal signal stream develops with age.

4. Metacognition and Inference

I have already mentioned that there has been a theoretical challenge to the view that there is assessment and consequent reallocation of effort in utterance interpretation, at least for non-implicated utterance content. François Recanati has claimed that recovery of what is said, including disambiguation, reference assignment to indexicals and pragmatic enrichment, is a brute-causal, non-inferential process (Recanati, 2004, ch. 2). My interest in that view here is that the non-inferential picture that Recanati suggests for part of utterance interpretation is also a non-metacognitive one (although Recanati does not use this term).

There is general agreement that senses of words and potential referents of indexicals have accessibilities: that is, they are easier or harder to bring to mind. As noted earlier, accessibility is known to be affected by recent use of a word-sense ('priming'). It also correlates with how frequent the word-sense is in usage. Now, as briefly sketched in Section 3.1, Recanati has proposed that accessibilities in context determine the explicit utterance content reached (in normal, smooth communication).

Consider again the examples in (2). There is general agreement that there are certain 'frames' that are associated with lexical items and made accessible by tokenings of them, for instance, that 'arrest' comes with a frame that has 'slots' for an arrester, an arrestee, a crime and so forth. Here is how Recanati explains the selection of John as referent for 'he' in (2a):

> John is the subject of 'was arrested' and therefore occupies the role of the person being arrested; now that role is linked to the role of the person doing the stealing, in some relevant frame. Because of this link, the representation of the referent of 'he' as the person doing the stealing contributes some activation to the representation of the person being arrested and therefore raises the accessibility of John qua occupier of this role. John thus becomes the most accessible candidate.' (Recanati, 2004, p. 31)

Recanati's claim is that such combinations of frames and accessibility factors do the job, except of course in cases where the hearer fails to recover the intended interpretation. (Equally, we should exclude cases where there is conscious reasoning about what is said).

As noted, Recanati's concern was to develop a non-inferential account of the recovery of explicit utterance content/what is said (in contrast to recovery of implicatures that he views as inferential). In my view it is also, and connectedly, a metacognition-free account of interpretation of what is said. In other words, as I understand it, Recanati is ruling out monitoring and control. This is because his account is purely bottom-up, and bottom-up accounts are in a certain sense 'blind': the output of such a process is determined by the inputs (albeit perhaps in complex ways). This is in contrast to a process governed by metacognitive feedback, where the output of the first-order process is monitored and the first-order process may be affected in a top-down way by the monitoring process.

Recanati sketches a way of simulating effects which seem top-down, such as an influence from the general context on the sense of a word that is chosen as the intended sense. What is crucial is that in his view these arise only through activations caused by features of the input: the priming of word senses, mental frames and so on. His account rules out any kind of genuinely top-down evaluation process that gauges how well things are going and 'decides as a result to pursue its efforts or reallocate them in different directions' to quote Sperber and Wilson again.

This is brought out in Recanati's reply to a criticism from Dan Sperber. Here's the criticism:

> Sometimes the first interpretation that comes to mind (the most accessible one) turns out not to be satisfactory and forces the hearer to backtrack. According to Sperber, the

> possibility of such garden-path effects shows that success, for a candidate semantic value, cannot be equated with sheer accessibility. (Recanati, 2004, p. 32)

As discussed earlier, Recanati's response is that such garden-path effects can be understood as due to accessibility shifts during processing: e.g., lexical priming from other words in the immediate linguistic context might rapidly make one sense of an ambiguous word highly activated, but then other activation from the broader context might kick in, so that a different sense ends up most highly activated. Presumably the sense that is most highly activated at some cut-off time after the utterance is the sense that 'wins', that is, the one that features in the hearer's representation of what is said.

I have discussed the exchange here because it illustrates that Recanati, unlike Sperber and Wilson, takes monitoring and control to be outside of his framework. What is more, there seems to be a more general claim implicit in the argument, namely that purely bottom-up processing cannot amount to abductive inference. I am also inclined to endorse this claim, although I draw the opposite conclusion from it about the character of the processes involved in utterance interpretation.

4.1 Inference, metacognition, and Marr's levels

The claim that purely bottom-up processing cannot amount to abductive inference raises the general question of how, and indeed whether, facts about whether a process involves monitoring and control or 'metacognition' relate to whether that process interpretation is inferential. Here I think that it is helpful to consider the well-known distinction between different levels of description for cognitive processes, as suggested by David Marr (1982).

Marr proposed three levels of description: the functional or computational, the algorithmic, and the hardware level. A functional account is concerned with questions such as 'What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?'(Marr, 1982, p. 25). The algorithmic (or 'representational') level is concerned with questions about how the computational account can be implemented. In particular, what is the representation for the input and output, and what is the algorithm for the transformation between them? Finally, at the hardware level, which I won't be considering here, one can ask how the representation and algorithm are realized physically.

For example, at the functional level a cash register (Marr, 1982, p. 22ff.) is (among other things) an adding machine. At the algorithmic level we want to know what format its input has to be in and what is the format of the output it produces and how the computation is performed: in decimal or in binary, for example. If it can do multiplication, we want to know whether it uses look-up tables of some sort, or performs repeated addition, or something else.

Now consider the pragmatic faculty i.e. whatever suite of abilities is responsible for spontaneous interpretation of utterances. The question about whether it performs inference is at the functional level. As noted earlier, the consensus view is that the task that it performs is

inferring the best explanation for an utterance in terms of the speaker's communicative intentions.

Recanati's claims about spreading activation delivering a representation of what is said are at the algorithmic level. He does not postulate a specific algorithm, but rather a characterization of the kind of processes involved: use of the word 'police' activates a certain set of assumptions to various degrees, use of the word 'arrest' activates a certain frame which has the roles <arrester, arrestee, crime>, and similarly for other words.

How, then, is this relevant to the computational-level description as inference to the best explanation? In particular, one wants to know what it is about Recanati's spreading activation model that rules out that the correct computational level description is inferential. Why shouldn't we instead see Recanati's description as a hypothesis about how inference is performed?

That is a difficult question, and I do not pretend to have a fully satisfactory answer. Here are sketches of two possible ones. The first is that what is going on in a purely accessibility-driven system is all non-propositional or sub-propositional, so it could not connect input and output together in warrant-preserving ways. One can compare here i) spreading activation which raises the accessibility of certain nodes in a network with ii) warrant- or truth-preserving transitions between mental representations with propositional content (e.g., in a Language of Thought).

I think that there is another reason why a Recanati-type model cannot be an implementation of inference, or at least not of abductive inference. Purely bottom-up processes without monitoring and control have no way of evaluating how well the output coheres with the input.

There is no known failsafe algorithm that, given any observation or fact, computes the best explanation for it. For this reason, in previous work I have argued that inference to the best explanation must in general be implemented as trial and error search (with various other properties): there is no alternative but to generate a trial solution and then evaluate it somehow. But in this chapter I have instead suggested that what is important is a process that has monitoring and control which checks on progress and steers processing towards better solutions. (I now think that trial and error search is a sub-category of such processes). My point here is that without some kind of steering it would be a miracle if the output happened to be the best explanation for, and warranted by, the input. Miracles may happen occasionally, but if an account relies on their occurring routinely it is defective.

5. Concluding Remarks

In utterance interpretation, any information may be relevant, but very little information can actually be processed. Therefore, I have argued on theory-driven grounds that an account of the psychology of utterance interpretation needs to explain how processing is steered towards promising lines of inquiry and away from others. There must be metacognition: monitoring and

control of the first-order processes involved. Given that utterance interpretation is normally phenomenologically 'seamless' and 'effortless', it follows that there must be subliminal metacognition, which I have compared with Proust's 'procedural metacognition' and Shea et al.'s 'system 1 metacognition'.

I have discussed two distinct experimental bodies of literature that back up this theoretical claim. The first is a considerable body of work that shows that activated word senses are suppressed when they are not needed as part of the final interpretation. The second is a study that found that time taken to read is modulated in response to contextual anomaly even in younger participants who lack consciousness of the anomaly in question.

Finally, I have tried to sketch out an explanation of how my claim that utterance interpretation involves metacognition is related to the view that utterance interpretation is inferential, appealing here to Marr's levels of description. The claim about metacognition is at the algorithmic level, while the view that a process is inferential is a functional-level claim, but facts at one level may have consequences on the other.

Utterance interpretation is not the only abductive inference task that we typically perform rapidly and without apparent effort. If it is right (following Proust and Shea et al.) that metacognitive processes can be unconscious and perhaps also non-metarepresentational, and the model I suggest of utterance interpretation is on the right lines, then a broader upshot suggests itself, namely that we can better understand how abductive inferences in general (not just ones implicated in utterance interpretation) can combine informational unencapsulation with speed, automaticity, and little awareness of execution.[9]

References

Allott, N. (2008). Pragmatics and Rationality. PhD thesis, University of London.

Allott, N. (2016). Misunderstandings in Verbal Communication. In A. Rocci & L. de Saussure (Eds.), Verbal Communication (pp. 485–507). Berlin: Walter De Gruyter.

Allott, N. (2019). Scientific Tractability and Relevance Theory. In K. Scott, R. Carston, & B. Clark (Eds.), Relevance: Pragmatics and Interpretation (pp. 29–41). Cambridge: Cambridge University Press.

Antony, L. (2003). Rabbit-Pots and Supernovas: On the Relevance of Psychological Data to Linguistic Theory. In A. Barber (Ed.), Epistemology of Language (pp. 47–68). Oxford: Oxford University Press.

Bach, K. (2006). The top 10 Misconceptions About Implicature. In B. J. Birner & G. L. Ward (Eds.), Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn (pp. 21–30). Amsterdam: John Benjamins.

Bob, P., Pec, O., Mishara, A. L., Touskova, T., & Lysaker, P. H. (2016). Conscious Brain, Metacognition and Schizophrenia. International Journal of Psychophysiology, 105, 1–8.

Boghossian, P. (2014). What Is Inference? Philosophical Studies, 169, 1–18.

Carruthers, G. (2013). Review of Foundations of Metacognition, 2012. In Michael J. Beran, Johannes Brandl, Josef Perner, & Joëlle Proust (Eds.), Notre Dame Philosophical Reviews, 2013, January 22.

Carston, R. (2010). XIII-Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images. Proceedings of the Aristotelian Society (Hardback), 110(3pt), 295–321.

Clark, H. H. (1994). Managing Problems in Speaking. Speech Communication, 15(3–4), 243–250.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in Spontaneous Speaking. Cognition, 84(1), 73–111.

Clark, H. H., & Krych, M. A. (2004). Speaking While Monitoring Addressees for Understanding. Journal of Memory and Language, 50(1), 62–81.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a Collaborative Process. Cognition, 22(1), 1–39.

Covington, M. A., He, C., Brown, C., Naçi, L., McClain, J. T., Fjordbak, B. S. et al. (2005). Schizophrenia and the Structure of Language: The Linguist's View. Schizophrenia Research, 77(1), 85–98.

Evans, J. S. B. T. (2003). In Two Minds: Dual-Process Accounts of Reasoning. Trends in Cognitive Science, 7(10), 454–459.

Evans, J. S. B. T., & Frankish, K. (Eds.). (2009). In Two Minds: Dual Processes and Beyond. Oxford: Oxford University Press.

Fodor, J. A. (1983). The Modularity of Mind: An Essay on Faculty Psychology. Cambridge, MA: MIT Press.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing 'the' as 'thee' to Signal Problems in Speaking. Cognition, 62(2), 151–167.

Grice, P. (1957). Meaning. The Philosophical Review, 66, 377–388.

Grice, P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), Syntax & Semantics 3: Speech Acts (pp. 41–58). New York: Academic Press.

Grice, P. (1978). Further Notes on Logic and Conversation. In P. Cole (Ed.), Pragmatics (pp. 113–127). New York: Academic Press.

Grice, P. (1989). Studies in the Way of Words. Cambridge, MA: Harvard University Press.

Harris, P. L., Kruithof, A., Terwogt, M. M., & Visser, T. (1981). Children's Detection and Awareness of Textual Anomaly. Journal of Experimental Child Psychology, 31(2), 212–230.

Kentridge, R. W., & Heywood, C. A. (2000). Metacognition and Awareness. Consciousness and Cognition, 9, 308–312.

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

Kuperberg, G. R. (2010). Language in Schizophrenia Part 1: An Introduction. Language and Linguistics Compass, I(8), 576–589.

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

Markman, E. M. (1977). Realizing That You Don't Understand: A Preliminary Investigation. Child Development, 48(3), 986–992.

Marr, D. (1982). Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information. San Francisco: Freeman.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in Recognizing Pairs of Words: Evidence of a Dependence Between Retrieval Operations. Journal of Experimental Psychology, 90(2), 227–234.

Neely, J. H. (1976). Semantic Priming and Retrieval from Lexical Memory: Evidence for Facilitatory and Inhibitory Processes. Memory & Cognition, 4(5), 648–654.

Paynter, C. A., Reder, L. M., & Kieffaber, P. D. (2009). Knowing We Know Before We Know: ERP Correlates of Initial Feeling-of-Knowing. Neuropsychologia, 47(3), 796–803.

Pomarol-Clotet, E., Oh, T. M., Laws, K. R., & McKenna, P. J. (2008). Semantic Priming in Schizophrenia: Systematic Review and Meta-Analysis. British Journal of Psychiatry, 192(2), 92–97.

Proust, J. (2008). Conversational Metacognition. In I. Wachsmuth, M. Lenzen, & G. Knoblich (Eds.), Embodied Communication in Humans and Machines (pp. 329–356). Oxford: Oxford University Press.

Proust, J. (2013). The Philosophy of Metacognition. Oxford: Oxford University Press.

Recanati, F. (1993). Direct Reference : From Language to Thought. Oxford: Blackwell.

Recanati, F. (2004). Literal Meaning. Cambridge: Cambridge University Press.

Robbins, R. R., & Westmoreland, R. B. (1991). Astronomical Imagery and Numbers in Mimbres Pottery. Astronomy Quarterly, 8(2), 65–88.

Rubio Fernández, P. (2007). Suppression in Metaphor Interpretation. Journal of Semantics, 24(4), 345–371.

Schvaneveldt, R. W., & Meyer, D. E. (1973). Retrieval and Comparison Processes in Semantic Memory. In S. Kornblum (Ed.), Attention and Performance IV (pp. 395–409). New York: Academic Press.

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-Personal Cognitive Control and Metacognition. Trends in Cognitive Sciences, 18(4), 186–193.

Simon, H. A. (1973). The Structure of Ill Structured Problems. Artificial Intelligence, 4(3–4), 181–201.

Simon, H. A. (1997). Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations. 4th Ed. New York: Free Press.

Simon, H. A., & Newell, A. (1958). Heuristic Problem Solving. Operations Research, 6(1), 1–10.

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. Psychological Bulletin, 119(1), 3–22.

Spehn, M. K., & Reder, L. M. (2000). The Unconscious Feeling of Knowing: a Commentary on Koriat's Paper. Consciousness & Cognition, 9, 187–192.

Sperber, D., & Mercier, H. (2012). Reasoning as a Social Competence. In H. Landemore & J. Elster (Eds.), Collective Wisdom: Principles and Mechanisms (pp. 368–392). Cambridge: Cambridge University Press.

Sperber, D., & Wilson, D. (1986). Relevance: Communication and Cognition. 2nd Ed. 1995. Oxford: Blackwell.

Sperber, D., & Wilson, D. (1996). Fodor's Frame Problem and Relevance Theory (reply to Chiappe & Kukla). Behavioral and Brain Sciences, 19(3), 530–532.

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-Reading. Mind & Language, 17(1&2), 3–23.

Stanovich, K. E., & West, R. F. (1998). Individual Differences in Framing and Conjunction Effects. Thinking & Reasoning, 4(4), 289–317.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for Multiple Stages in the Processing of Ambiguous Words in Syntactic Contexts. Journal of Verbal Learning and Verbal Behavior, 18(4), 427–440.

Todd, P. M., & Gigerenzer, G. (2000). Précis of 'Simple Heuristics That Make Us Smart'. Behavioral & Brain Sciences, 23(5), 727–741; discussion 742.

Vriend, N. J. (1996). Rational Behavior and Economic Theory. Journal of Economic Behavior and Organization, 29(2), 263–285.

Wilson, D., & Sperber, D. (2012). Meaning and Relevance. Cambridge: Cambridge University Press.

---

1. I don't claim to have a watertight definition of 'feedback'. The Oxford Shorter Dictionary offers the following reasonable characterisation: 'the modification or control of a process or system by its results or effects'. I would only add that in this paper I am concerned with effects on a mental system that come from other mental systems that are sensitive to what the first system is doing.

2. As noted by Glenn Carruthers (2013).

3. Thus, for example: 'The coded signal, even if it is unambiguous, is only a piece of evidence about the communicator's intentions, and has to be used inferentially and in a context' (Sperber and Wilson, 1986, p. 170); and 'even if what a speaker means consists precisely in the semantic content of the sentence he utters, this still has to be inferred.' (Bach, 2006, p. 24).

4. It is also worth noting that utterance interpretation is apparently a process that only adds beliefs, not one that may add some and subtract others. The hearer starts with a belief that the speaker has uttered certain linguistic material, in a certain way, in a certain context, and, if all goes well, ends up with beliefs about what the speaker intended to convey by her utterance, e.g., what she stated and what she implicated. Of course, what the speaker conveys may contradict a previous belief of the hearer's, and the hearer may end up dropping that belief as a result of the utterance. (The speaker might state or implicate that the cat is on the mat and the hearer may thereby learn that the cat is on the mat and not, as he supposed, elsewhere.) But this is arguably 'downstream' of the utterance interpretation process proper.

5. N.B. it is not entirely clear whether in Sperber and Mercier's terminology, 'intuitive inference' is simply a positive name for non-reflective inference, given that they also say that intuitive inferences are carried out by domain-specific mechanisms, but do not make it clear whether this is part of their definition or an empirical claim.

6. As Anders Nes points out (pc), one reason for this is that there will often be clashes between explanatory virtues such as simplicity, degree of fit with observation, and conservativeness and it often will not be clear how to weigh them against each other.

7. Asher Koriat writes: 'Students of metacognition not only place a heavy emphasis on subjective experience but also assume that subjective feelings, such as the feeling of knowing, are not mere epiphenomena, but actually exert a causal role on information processing and behavior' (2007, p. 293; see also pp. 315–316).

8. The claim that sense suppression is unconsciously metacognitive has previously been made in the context of schizophrenia (Carruthers, 2013), in a suggestion that brings together the idea that deficits in metacognition abilities are a central factor in schizophrenia (Bob et al., 2016) with the finding that 'studies of word recall in schizophrenia generally point toward impaired control of spreading activation' (Covington et al., 2005). See also what follows in the main text.

9. Thanks are due to Anders Nes for pressing me to be explicit about how there is an upshot for our understanding of other abductive inference processes.