Accepted manuscript, forthcoming in *The Journal of Philosophy*
Please cite published version

Metacognition of Inferential Transitions

Nicholas Shea

Abstract

A reasoning process is more than an unfolding causal chain. Although some thoughts cause others in virtue of their contents, paradigmatic cases of personal level inference involve something more, some appreciation that the conclusion follows from the premises. Both first-order processes and second-order beliefs have proven problematic or inadequate to account for the phenomenon. Thus, here I argue for an intermediate position, according to which an epistemic feeling, a form of procedural metacognition, plays this role. Extensive psychological research has shown that epistemic feelings are involved in monitoring many kinds of cognitive process, affecting how the processes unfold. Inferences may be no different. Inferences are also plausibly accompanied by an epistemic feeling, in particular a feeling of reliability or unreliability. Such a feeling accounts for the phenomenological datum. It can also play a significant epistemic role for the thinker.

Sections

(1) Introduction

The Representational Theory of Mind (RTM) is based on the powerful idea that causal transitions between representational vehicles can be configured so as to respect representational contents. Implemented in electronic computers, this principle underpins much of modern technology. Implemented in people, it offers a compelling account of the subpersonal mechanisms giving rise to many of our mental capacities, for example our capacity to recognize and track objects, or to learn from rewards. We also make transitions between representations at the personal level (personal level: premises and conclusion are conscious and/or potentially known by the person). These *inferences* look to be something more than causal transitions between contentful vehicles. The thinker does not just find themselves being caused to form a conclusion. They have some kind of appreciation that the conclusion follows from the premises.

What is the extra ingredient? It is something both phenomenologically and epistemically distinctive. Phenomenologically, the thinker has some sense of what is going on. Rather than a conclusion just being caused to appear in the mind, the thinker appreciates, in some way, that the conclusion follows from what came before. Epistemically, this appreciation plays a role in securing the epistemic status of the conclusion, for example in furnishing justification or entitlement.

Some theorists account for the phenomenon in terms of a second order ingredient. For Paul Boghossian, for example, what characterises cases of inference proper is that the thinker takes the conclusion to follow from the premises (Boghossian 2014, 2016, 2018, 2019). The difficulty is to satisfy this *taking condition* without requiring an explicit second order belief, which both seems overly intellectual and threatens a regress. Others attempt to capture the difference in purely first order terms. Hilary Kornblith, for example, argues that inferences exhibit more flexibility than do mere dispositions to make causal transitions (Kornblith 2012, 2016). Here I show that there is no need to choose between these two camps. An intermediate possibility has been overlooked. The ingredient I will point to is not a matter of intellectual reflection, but it is more than merely first order.

My inspiration is the psychological literature on metacognition. We now have extensive evidence that people have epistemic feelings about many cognitive processes: perceptual decisions, memory recall, action execution and so on. These forms of 'procedural' metacognition concern a cognitive process, and affect its results, without requiring the thinker to deliberate on or conceptualise the process (Proust 2012). Hence procedural metacognition does not require the thinker to deploy (or possess) concepts of memory, belief, truth, accuracy, entailment or confidence. Nor does it require the capacity for reflection or reflective self-ascription. At the same time, procedural metacognition is more than merely first order since an epistemic feeling concerns the cognitive process to which it attaches, the process which it regulates.

The philosophical literature on the nature of inference has not considered procedural metacognition. I will argue that procedural metacognition is a good

candidate for the missing ingredient. A certain kind of procedural metacognition – what I will call a 'feeling of reliability' – could be the salient difference between causal transitions and cases where it seems to the thinker that the conclusion follows from the premises. It is plausible that metacognitive feelings attach to inferential transitions, feelings that roughly track the reliability of a pattern of inference. That would explain the distinctive phenomenology of inference – the less-than-fully-reflective sense the thinker has that the conclusion follows from the premises. The feeling alone will not be sufficient to turn a causal transition into something the thinker does, a mental action, but it plausibly forms part of an account of how mental agency is exercised when performing inferences, both psychologically and epistemically. The feeling does epistemic work for the thinker, since it is a means by which their dispositions to make transitions between representations, and to endorse conclusions, are adapted to reflect whether the conclusion is reliable and does indeed follow from the premises.

I start by characterising the phenomenon (section 2) and giving an overview of the debate about the additional ingredient (section 3). I then introduce some well-studied examples of procedural metacognition from the psychological literature (section 4). This motivates the idea that an epistemic feeling attaches to inferential transitions – a feeling which accounts for the phenomenological contrast and plays an epistemic role even for an unreflective thinker (section 5). Feelings of reliability and unreliability are also likely to be part of the way reflective thinkers comply with the demands of epistemic responsibility (section 6).

(2) The Phenomenon

A series of mental representations unfold in thought because of their contents. It is because of what I wanted that I formed a particular intention. Whether that is compatible with RTM is an important issue in its own right, but that is not our issue here. The contrast we are interested in arises between pairs of cases both of which involve transitions in virtue of content. The phenomenon is something further. My aim is to account for this further ingredient, not to offer an account of how transitions occur in virtue of content.

Consider Boghossian's case of the highly strung character who, whenever he thinks to himself, *I'm having so much fun*, finds himself disposed to conclude, *but there's so much suffering in the world* (Boghossian 2014). The second thought is a conclusion he draws. It arises because of the content of the first thought. But it doesn't seem – to him – to follow from the first. Although the thinker is moved by content, what is absent is any sense that the second thought makes sense in the light of the first.

If our highly strung depressive were simply free-associating, then we could dismiss the case as involving a special kind of mental activity. When free-associating, one thought follows another in all sorts of ways for all kinds of reasons. When I say 'salt' you may think *pepper*. That doesn't mean that, when you're following a recipe

that calls for salt, you'll decide to add pepper. But the contrast we are interested in arises when people are not just free-associating. It occurs when they are engaged in theoretical or practical inference. I'm weighing up the pros and cons of a particular American president's term in office. After a series of cons, I try to counter any bias by finding some pros. *The economy grew strongly*, I add to my mental list, before finding myself thinking, *but Trump's an idiot*. I am in fully 'factual' mode, not free-associating, and the unfolding chain of thought is explained by content. But at this particular juncture a further ingredient is lacking. It doesn't seem to me that the conclusion follows from the premises. There is a clear phenomenological contrast with ordinary inferential transitions where, in a way that is usually unremarked on, one thought leads smoothly to the next.

The point is not that the inference is invalid or unreasonable. The thinker can have this feeling when making an inference that is in fact invalid (e.g. affirming the consequent). A thinker can also lack this feeling when making an inference that is in fact valid. Michael Dummett offers the example of a mathematical proof. One can understand the meaning of each line of the proof and yet, at some step, fail to see why the next line follows from what came before (Dummett 1991, p. 198). Even if the disposition to make such a transition had been drilled into the thinker by an 'International Academy of Logic', they would make the transition without appreciating its validity. They would lack any sense of the conclusion's following from the premises.

Bill Brewer captures it nicely: 'such beliefs would come as a succession of mere hunches, wholly unsubstantiated for me by the de facto validity of the argument propelling my endorsement of them' (Brewer 1995, p. 242). What is lacking in following the argument is, 'some appreciation of why I am right in believing its conclusion', 'some grip on how I thereby know the conclusion' (p. 242). As Kornblith puts it, the thinker has a 'feeling of endorsing the last proposition on the basis of earlier ones' (Kornblith 2012, p. 143).

These quotations bring out two aspects of the crucial ingredient. First is the phenomenological datum that (some / most / all) inferences are accompanied by some kind of feeling that the conclusion follows from the premises. This feeling is lacking in cases where a thought is nonetheless caused by the content of earlier thoughts. (There may also be a contrary feeling in such cases, perhaps of uncertainty or puzzlement.) Second, this feeling plays an epistemic role: it is part of why the thinker is justified or entitled to draw the conclusion.

My aim is to use procedural metacognition to account for this feeling. The phenomenological difference between the cases where it does or does not seem to the thinker that one thought follows from others is plausibly explained by the presence and absence of a feeling of reliability. This feeling also does important epistemic work for the thinker. It allows them to differentiate bad transitions from good, without requiring them to stop and reflect at every step. Although this relatively crude unstructured signal falls short of an intellectual understanding of the validity of the inference, as we will see in the next section, it may be a mistake in these cases to expect the kind of justification that is available through explicit self-directed epistemic reflection.

For Boghossian, a transition only counts as an inference when this ingredient is present (for him, when the thinker takes the conclusion to follow from the premises, Boghossian 2014). By contrast, Susanna Siegel argues that there can be inferences that lack self-awareness or any other kind of reckoning state (Siegel 2019). I can remain neutral on whether the feeling is partly constitutive of a transition's being an inference. Our question is how to account for the datum in the cases where it arises, whether these are all the inferences or just an important subset of them.

(3) First Order or Second Order?

Our target is the phenomenologically and epistemically distinctive ingredient that accompanies some transitions between mental states (which, for Boghossian, makes the transition an inference). There is a large literature on the nature of this ingredient. I won't attempt to summarise the debate here. Instead, I will set out two prominent accounts, those put forward by Kornblith and Boghossian. I highlight these two views as representative first-order and second-order accounts, with the aim of showing that they leave space for a previously under-appreciated intermediate position.

An obvious first reaction to the phenomenon is to suggest that the thinker knows or understands that the conclusion follows from the premises, where this understanding is a piece of conceptual knowledge. The thinker reflects on the inference pattern and explicitly represents that it is valid. That belief is then a source of justification for drawing the conclusion. Kornblith objects to an account that calls for deliberate reflection, which he understands as being a matter of what psychologists characterise as 'system 2' cognition (Kornblith 2012). Cognitive processes are part of system 2 (or, better, operate in the 'type 2' way) when they are purposeful and effortful, proceed step-by-step, and are susceptible to interference by a concurrent effortful task ('cognitive load'). System 2 processes contrast with psychological processes that operate in the 'type 1' way: fast, automatic and capable of running in parallel without interference. Our ingredient cannot be just a matter of deploying system 2 cognition because system 2 processes need not be reflective (i.e. self-directed). A thinker could engage in system 2 thinking without deploying any self-directed or epistemic concepts (like BELIEF, FOLLOWS or JUSTIFIES), or without having any such concepts. The proposal must be that one directs system 2 thinking on oneself, deploying epistemic or mental-state concepts. One deliberates about the inference and forms the explicit belief that the conclusion follows from the premises. Clearly that can happen (we're doing it now), but is reflective system 2 thinking the ingredient we need?

Kornblith argues that system 2 cognition is unsuited to playing an epistemic role (Kornblith 2012). An inference may or may not be reliable, but deliberating is unlikely to make it more reliable. Kornblith points to evidence that system 2 thinking can often lead us astray, falling prey to biases and self-serving confabulation. He also points to what seem to be cases of reasoning in animals that are incapable of reflection or system 2 cognition. What characterises reasoning in animals, and distinguishes it

from merely responding to stimuli, is its flexibility. The animal can react differently to new information, tailor its behaviour to the context and adapt productively to its experience. Kornblith argues that it is this flexibility that characterises the epistemically good cases, the cases of reasoning proper, not reliance on unreliable reflective deliberation.

Kornblith might be right to question the reliability of reflective deliberation (philosophers' tool of choice), but his purely first order account leaves us without an ingredient to account for the phenomenological datum. A purely first order account in terms of flexible use of information does not generate a distinction between cases where it seems to the thinker that the conclusion follows from the premises and those where it does not. We are just left with more and less sophisticated patterns of transitions between contents.

There are other reasons to resist an account in terms of reflective deliberation. Intuitively, it is not obvious that the feature in question is an explicit belief. That certainly does not fit the intuitive descriptions offered by Dummett or Brewer. Instead, we have an 'appreciation', 'grip', 'sense' or 'feeling' that the conclusion follows. Requiring an explicit second-order belief would also deny the capacity to those who lack the relevant epistemic concepts: young children and non-human animals.[1] If the feeling is something that occurs in our thinking when we are not reflectively deliberating, we might expect that the same ingredient might also be present in the cognitive processes of thinkers who lack the capacity for reflective deliberation.

Furthermore, positing an occurrent second order belief threatens a regress. A second order belief about what follows from the premises becomes a further premise. How does the thinker appreciate that the conclusion follows from this enlarged set of premises (Carroll 1995)? Does that require a third order belief?

Boghossian blocks the regress by claiming that the relevant second order belief is merely tacit (Boghossian 2016, 2018, 2019). When people make an inference they do not at that point typically entertain an explicit belief that the conclusion follows from the premises. The tacit belief is in the background in the way that a goal can supervise a thought process without remaining explicitly in mind. A tacit belief can play an epistemic role because when the thinker is challenged as to why they drew the conclusion, they are able to make the second order belief explicit and offer it is as justification: it is because the conclusion really does follow from the premises. A tacit belief on its own is however inadequate to the phenomenological datum. If it is merely tacit, it is not something that generates a phenomenological contrast at the point when the inference is made. What is present in the good cases and absent in the merely causal cases is not a second order belief, excavated from a tacit store through

[1] Even if the concept of belief is not required (Peacocke 1996, pp. 129-130), some epistemic concepts would have to be deployed by the thinker and self-directed at an inferential transition they have made (e.g. SUPPORT, CONSEQUENCE, or EVIDENCE FOR).

reflection and made explicit, but something that is present without deliberate reflection.

Furthermore, an account in terms of tacit second order belief still seems overly intellectual. The claim is not that a thinker has the tacit second order belief simply in virtue of having the disposition to believe the conclusion in virtue of believing the contents of the premises (as in the view of Broome 2013, p. 231). A tacit belief has to be something that the thinker can make explicit. That commitment is needed for the belief to play the epistemic role Boghossian wants it to play. This calls for the thinker to possess the relevant epistemic concepts. But as we saw, it is not obvious that possessing such concepts, or having the capacity for reflective deliberation, is needed to generate our contrast, since it seems to occur in our thinking when we are not reflectively deliberating. We may well find the same phenomenon, with an associated epistemic role, in animals and young children. The capacity for reflective deliberation is epistemically more powerful, but the point of moving away from explicit second order accounts was to find a less sophisticated ingredient.

Even with mature adult thinkers, it is not clear that they have a tacit belief in the relevant sense – a belief that they could make explicit. If asked why they formed the conclusion of an inference involving IF…THEN or AND, say, it is not obvious that an ordinary reasoner would be able to give the relevant second order justification (without having had the benefit of philosophy classes).[2]

None of this amounts to a knock-down argument against either first order accounts (Kornblith) or second order accounts (Boghossian). What it does show is that there are difficulties with both, difficulties which push in the same direction – towards the middle. First order accounts are insufficiently second order, and second order accounts, to the extent that they address the phenomenological and epistemic datum, give answers that are overly intellectual. It would be nice to split the difference, but how can we have an account that is richer than first order but less than second order? This is where the large psychological literature on procedural metacognition offers a possibility which has been overlooked in the philosophical debate.

(4) Psychological Evidence of Procedural Metacognition

Research in experimental psychology and cognitive neuroscience has discovered several types of procedural metacognition. These are signals that play a role in monitoring and controlling cognitive processes without involving deliberation (type 2 / system 2 cognition) or reflection (self-application of epistemic or mental-state concepts). These signals track characteristics of a cognitive process, like its fluency or reliability. They in turn affect downstream processing. For example, a signal of uncertainty makes it more likely that the thinker will stop and reflect or gather further information. There is good evidence for procedural metacognition in non-

[2] Thanks to Michael Strevens for this point.

human animals (Hampton 2001, Middlebrooks and Sommer 2012, Kepecs, et al. 2008, Foote and Crystal 2007, Adams and Santi 2011) and in infants who lack mental state concepts (Goupil, et al. 2016, Goupil and Kouider 2016).

In philosophy, Joëlle Proust has written extensively about both procedural and analytic metacognition (Proust 2010, 2012, 2013). Many examples of the former are what she calls 'epistemic feelings'. Epistemic feelings concern a cognitive process in the sense that their functional role turns on, for example, the reliability or accuracy of the process. That is what makes them *meta*-cognitive. But they do not involve deploying mental or epistemic concepts, concepts of reliability, accuracy, belief, decision, etc. In that sense, epistemic feelings are non-conceptual. Procedural metacognition contrasts with analytic metacognition, conceptual thought about one's own mental states and processes, which requires the thinker to possess corresponding mental state concepts.

I will argue that there is a form of procedural metacognition directed at transitions between conscious thoughts, a *feeling of reliability* of the transition. There have been extensive studies of how procedural metacognition arises for several kinds of psychological processing. These offer us a model of how procedural metacognition of inference is likely to work. I will give three examples.

First off, in perceptual processing. In a typical task participants are shown a stimulus and asked for a judgement about a property of the stimulus. Shown an array of randomly scattered dots in motion, they are asked whether the preponderant direction of motion is to the left or to the right. Having reported their decision, say with a button press, they are then asked how confident they are in that decision (Moran, et al. 2015). Reported confidence is a broadly reliable guide to the accuracy of a perceptual decision (Rademaker, et al. 2012, Bona and Silvanto 2014), although the confidence reported is also affected by rewards and strategic considerations (Hertz, et al. 2017). Perceptual confidence seems to draw on dedicated neural mechanisms (Fleming, et al. 2010). Intervening on these neural mechanisms systematically perturbs the confidence people report (Rounis, et al. 2010, Cortese, et al. 2016). The feeling of confidence in a perceptual decision is a form of procedural metacognition. As well as affecting reported confidence, it affects whether the subject will pay a cost to get more information (Pescetelli, et al. 2021), and how they will gamble on their choice (Moreira, et al. 2018). The feeling of high or low confidence plays a functional role in cognition that depends on its tracking, at least roughly, the accuracy of the decision.

A second example is metacognition of memory. When a memory is retrieved it is accompanied by a feeling of certainty or uncertainty (Koriat, et al. 2006). Thinkers rely on this feeling when deciding whether to act on the information that has been retrieved (Koriat and Helstrup 2007, Proust 2013). A metacognitive feeling also arises prospectively, before recalling a memory. Thinkers have a 'feeling of knowing' which predicts whether they will be able to recall information accurately (Koriat and Levy-Sadot 2001, Dokic 2014). In the absence of a feeling of knowing the thinker may attempt alternative strategies for recalling the information (Arango-Muñoz 2014). We find similar processes at work in young children (Goupil, et al. 2016, Goupil and

Kouider 2016) and non-human animals (Hampton 2001, Middlebrooks and Sommer 2012, Kepecs, et al. 2008, Foote and Crystal 2007, Adams and Santi 2011).[3] These feelings affect how cognitive processes unfold, without the thinker having to form self-referential conceptual thoughts.

A third area where procedural metacognition has been investigated is in reasoning (Ackerman and Thompson 2017). This work gets close to our target phenomenon but does not quite capture it. Research has focused on people's confidence in the conclusion of a piece of reasoning. For example, participants will be given a syllogism and asked to select which sentence follows from the premises. Their 'feeling of rightness' about the conclusion (Thompson, et al. 2011, Thompson, et al. 2013, Thompson and Johnson 2014) can then be probed in various ways, directly by asking them to report their confidence (De Neys, et al. 2011), or indirectly through implicit measures like skin conductance response or whether they avail themselves of the opportunity to reconsider (De Neys, et al. 2010).

The feeling of rightness has been found to play an interesting role in problems that have an intuitively obvious but incorrect answer. To answer correctly the thinker must avoid giving the fast, automatic answer and instead engage in some deliberate reasoning (Frederick 2005).[4] Researchers have found that the feeling of rightness affects how thinkers tackle the problem. The degree to which the first, automatic answer feels right predicts how much the thinker will engage in further reasoning about the question (Ackerman and Thompson 2017). That is helpful, but only partially, since feelings of rightness are an imperfect guide to whether the answer is actually correct (Shynkaruk and Thompson 2006, Markovits, et al. 2015). They are largely based on how fluently the initial answer occurs (Thompson, et al. 2011, Thompson, et al. 2013), and the familiarity and consensuality of the response (Bajšanski, et al. 2019).

The feeling of rightness is sensitive to whether the inference is valid. Where there is a conflict between the intuitive answer and the conclusion that would follow from applying a logical rule, that conflict seems to be registered by the thinker (De Neys, et al. 2008, De Neys, et al. 2010, De Neys, et al. 2011, De Neys 2012). Being the result of a valid inference also seems to confer a stronger feeling of rightness on the conclusion (Markovits, et al. 2015). The feeling of rightness attaches to the conclusion of an inference, not to the process of inference itself as I am proposing. Nevertheless, this substantial body of research establishes the plausibility of the idea that epistemic feelings are involved in the way people perform inferences.

---

[3] Whether these forms of procedural metacognition are feeling-based is not firmly established, particularly in animals.

[4] An example from the Cognitive Reflection Task: 'A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?'.

(5) The Proposal

5.1 A Feeling of Reliability

My proposal is that a transition between mental representations in thought is accompanied by an epistemic feeling. This feeling roughly tracks the reliability of the pattern of inference. I have called it a 'a feeling of reliability' but this should not imply that the thinker must know that the feeling concerns reliability (or that they have a concept of reliability). The label reflects its functional role. The feeling affects the way the thinker exercises their capacity for inference: whether they will rely on the inference or whether they will be more cautious, for example, gathering further information or engaging in deliberation. In the good cases, where reasoning unfolds smoothly and the thinker is disposed to go on, the transition is accompanied by a feeling of high reliability. The contrast cases lack that feeling. They may instead have a feeling of low reliability – a kind of uncertainty or disfluency. These two feelings may well lie on a common, graded scale. The feeling of low reliability may be the same as or related to the feeling of error that participants report when something goes wrong in a calculation or reasoning task (Fernandez Cruz, et al. 2016). The feeling of high reliability largely goes unnoticed, except when we are invited to reflect on the contrast with cases where it is absent, where a content just appears causally in the train of thought (as with the anxious depressive).

The feeling of reliability is generated by and attaches to a transition – a process. It is separate from the just-mentioned feeling of rightness, which attaches to the conclusion of a piece of reasoning. It is likely that the feeling of reliability affects the feeling of rightness (the interim confidence the thinker has in the conclusion). That would explain why the conclusion attracts a stronger feeling of rightness when the inference is valid (Markovits, et al. 2015). However, the feeling or rightness is affected by many other factors, notably the plausibility of the conclusion. When people draw a conclusion which, although logically entailed, is unfamiliar or conflicts with the consensus, they will feel uncertain about the conclusion. It will have a low feeling of rightness. The goodness of the pattern of inference is just one of many factors bearing on the thinker's intuitive feeling of whether the conclusion is likely to be true.

To get a sense of the feeling of reliability, compare the following inferences:-

(1) If the card is green at the top then it is red at the bottom.
(2) The card is green at the top.
∴ (3) The card is red at the bottom.

(4) If the card is blue at the top then it is yellow at the bottom.
(5) The card is not yellow at the bottom.
∴ (6) The card is not blue at the top.

Both are instances of valid forms of inference but, for most, the first inference is more fluent than the second. I suggest that we have a feeling of reliability that is high in the first case and lower in the second case. This explains why, in the first case, as the conclusion appears, it seems to us to follow straightforwardly; whereas in the

second case we do not get that sense so strongly.[5] Modus ponens is typically fast and automatic (Lea, et al. 2005). Modus tollens is less so, and generally attracts lower confidence (Markovits, et al. 2015).

The fluency of an inference makes the thinker less likely to engage in reflection about the answer (Thompson, et al. 2011, Thompson, et al. 2013). This suggests that fluency makes a strong contribution to whether an inference feels reliable. This is likely to depend on experience. Simple valid inferences are used often in everyday reasoning. Nothing goes wrong as a result so we have considerable experience that the reasoning pattern is reliable. For example, the following pattern is very familiar and unproblematic in use:

(7) p v q
(8) ¬p
∴ (9) q

(10) Maya is either a doctor or a lawyer.
(11) She is not a doctor.
∴ (12) She is a lawyer.

Compare what would happen when reasoning with the putative logical concept TONK (Prior 1967). TONK obeys the following introduction and elimination rules:

p	p tonk q
∴ p tonk q	∴ q

A thinker who started to use this concept in reasoning would rapidly encounter contradictions. There is no way that inferences involving TONK would build up an associated sense of reliability. Experience would teach that things do not go smoothly when using this pattern of inference.

Evidence for the role of experience in shaping inferential dispositions can be found in people's disposition to affirm the consequent. Those who have taught logic classes will know that, at the outset, most students have a disposition to affirm the consequent, apparently without being aware of the risk of drawing false conclusions. This may be because they are deploying a form of reasoning, inference to the best explanation, that is mostly reliable in many contexts (see also Oaksford and Chater 1994):

(13) If wood is less dense than water, then wood will float on water.
(14) Pieces of wood float on water.
∴ (15) Wood is less dense than water.

---

[5] For the cognoscenti who find (4)-(6) no different than (1)-(3), try:
(4*) If the card is not blue at the top then it is yellow at the bottom.
(5*) The card is not yellow at the bottom.
(6*) The card is blue at the top.

As students study logic, the way they reason changes. Some inferences become more automatic, especially as a result of doing exercises and getting feedback (e.g. modus tollens, (4)-(6) above). Other transitions become less fluent, for example affirming the consequent. Students become better at the syllogism, and better able to judge the accuracy of their reasoning (Prowse Turner and Thompson 2009). It seems very likely that the feeling of reliability or unreliability experienced by the reasoner when performing these types of inferences changes accordingly.

My aim has been to account, both phenomenologically and epistemically, for the contrast between normal inference and the fun/suffering transition described by Boghossian (2014). How, though, does the feeling of reliability fit into a wider account of mental agency? Performing an inference is a mental action. A mental action is something the thinker *does*, as a whole agent or person, like deciding or calculating, rather than something that happens to them, like falling asleep or feeling an injury.[6] Although I am not claiming that the feeling of reliability turns a causal transition in thought into a mental action, it will figure in an account of how mental agency is exercised. On my picture, the tokening of some premises triggers both the disposition to draw a certain conclusion and an accompanying feeling of high or low reliability. The feeling is part of why the thinker may perform the inference and draw the conclusion, or alternatively decline to do so. Relying on the feeling in this way makes sense, given the link, albeit imperfect, between the feeling and whether an inference pattern is in fact reliable.

Is high reliability needed for a transition to count as an exercise of mental agency? Plausibly not. The anxious depressive example in fact covers two cases. In one case, the thinker is caused to think, *but there's so much suffering in the world*, but does not reach that conclusion agentially. It is something that happens to the thinker. (The fact of considering the question is done intentionally, but tokening the conclusion may be a passive mental event, an intrusion on the mind.) The other case, which it seems also occurs from time to time, is where we perform an inference but do so feeling less secure about whether the conclusion follows. That then is an exercise of mental agency, and the relatively low feeling of reliability has an effect on how we exercise that agency, for example whether we decide to revisit the issue and think about it further. The thinker does get to *there's so much suffering in the world* by inference, but it doesn't seem to them to follow. If an inference has to be accompanied by a corresponding epistemic feeling, then a feeling of low reliability (unreliability) suffices.

My hypothesis is that all inferences taking place between conscious representations are accompanied by a feeling of reliability. If we allow that there can

---

[6] I stick with this simple formulation, since any richer conception of will attract some controversy, although my argument in the paper is compatible with (plausible) views according to which a mental action is: something the thinker has control over; that they could have done differently; that they can held responsible for; and that is directed towards achieving a goal.

be cases of non-conscious inference (end of §2), then there will be cases of inference which lack any feeling of reliability. But is seems unlikely that these would count as mental actions either. So the effect would be to allow the category of inference to extend beyond mental actions. For inferences that are exercises of mental agency, the feeling of reliability is a key part of how they unfold.

The feeling of reliability figures in a second, stronger thesis that I neither endorse nor reject here. One could argue that the operation of the feeling of reliability is part of what makes this kind of mental process count as an inference – that it is partly constitutive of what makes the transition an inference proper, even if not on its own sufficient to make a causal transition into an inference. The feeling would figure in a non-homuncular account of mental agency according to which the right set of mental components, operating in the right way, are what makes a mental process count as an exercise of mental agency. In the case of inference, the feeling of reliability would be one such component. However, it would take substantial further work to establish – or just to investigate – this claim.

The evidence discussed above suggests that, as a result of experience, different feelings of reliability or unreliability attach to different inferential forms. Doesn't that require the thinker to identify different forms of inference? Doesn't the thinker need to categorise the inference in order to apply the appropriate degree of reliability? No. The claim is that the feeling of reliability is generated automatically by the mechanism that causes the transition. We are concerned with dispositions to transition between representations that are self-executing: the stimulus condition for the disposition is that the corresponding premises are tokened. (Whether the conclusion is in fact tokened may depend on other factors, e.g. whether it conflicts with other beliefs.) The thinker finds the transition 'primitively compelling' (Peacocke 1992, p. 6). One way of understanding this is that the process of exercising the disposition generates an epistemic feeling (in part because of its fluency or disfluency). Different patterns of inference generate different feelings. The feeling attaching to each inferential pattern is modified by experience with that inference type, perhaps at least partly through the effect of experience on fluency. Feelings of reliability or unreliability arise automatically and are different for different types of transition.

How exactly is the feeling of reliability shaped by experience? That is an empirical question, but extrapolating from other kinds of learning it seems that at least three kinds of experience will be important. First, whether deploying the pattern of inference leads to contradictions (Millikan 1984, Shea 2023b). Second, whether the expectations that result from an inference are met or not, especially expectations about the consequences of performing actions. Third, and perhaps most important, the reactions of other people (Koriat 2008) – whether the conclusions we reach accord with others' expectations and whether others are disposed to infer in the same way. These factors may have an effect on the feeling of reliability through making the inference more or less fluent. In addition or alternatively, they may be registered by a stored representation of the reliability of the inference.

We have seen that the thinker's confidence in the conclusion is affected both by whether the inference is valid and by whether the conclusion is plausible.

Conversely, when a thinker opts for a plausible claim which does not in fact follow from the premises (exhibiting belief bias), the invalidity of the inference impacts their confidence (Evans, et al. 1983, De Neys 2012). The feeling of reliability of an inference may also affect whether the thinker draws the conclusion at all, especially where there is a conflict. In other areas of behaviour it has been shown that different sources of evidence are weighted by their relative reliability (Ernst and Banks 2002, Lee, et al. 2014, Shea and Frith 2019). The feeling of reliability attaching to inferences could have the same effect. Where an inference pattern is associated with a feeling of high reliability, that should increase the chance that the inference will be relied on and the conclusion adopted.

5.2 Beyond Logical Transitions

I have argued that there is plausibly an epistemic feeling – a feeling of reliability or unreliability – attaching to inferential transitions. The examples so far have concerned syllogistic reasoning. The inference patterns have been broadly logical. We may also be disposed to make non-logical transitions between representations. In this section I will argue that a feeling of reliability also attaches to non-logical transitions. Here is an example of a non-logical transition:

(16) x is a whale
(17) x is a mammal

Although this transition is not logically valid, it is a good disposition to have. The thinker will not go wrong in our world. Whatever singular term is substituted for x, if the premise is true the conclusion will be true too.

If there were a further, suppressed general premise, then the transition would in fact be a logical one. But a thinker could be disposed to move to (17) just in virtue of tokening (16), without any further premise. Then it would not reduce to a logical inference. Such a disposition would still be dependent on syntactic structure: introduce a 'not' into (16) and the disposition would not be triggered (cp. Quilty-Dunn and Mandelbaum (2017)). It is not simply a disposition to token the concept MAMMAL in virtue of tokening the concept WHALE. It operates at the level of complete structured thoughts.

The disposition is a good one to have, given the specific contents of the predicates (*whale*, *mammal*). It would not work for arbitrary kinds F and G. The disposition is 'content-specific' in the sense of (Shea 2023a). Through experience we acquire habits of thought that effectively assume that various regularities hold in our world (e.g. *that is a dog* $\rightarrow$ *it will bark*). We get many of these habits of thought by observing how other people use terms. For example, most people learn that we can use the concept of SET like this:

(18) Something is F
(19) There is a set of all and only the things that are F

That seems innocuous, but studying philosophy one discovers that it generates contradictions. One therefore becomes more cautious about drawing the conclusion. It is not that the disposition to use SET in (18)-(19) goes away entirely. But it is more tentative and often triggers reflection. I suggest that, as a result of learning about the contradictions, this pattern of inference comes to be associated with a feeling of unreliability. Comparing (16)-(17) with (18)-(19) suggests that feelings of reliability and unreliability attach to direct, non-logical, content-specific patterns of inference.[7]

My hypothesis is that feelings of reliability attach to non-logical inferences in just the same way as with logical inferences, and that they are modified by experience in the same fashion. For theorists who argue that the phenomenon we are interested in is a matter of the thinker understanding the concepts involved in the inference, and grasping that the conclusion follows on the basis of meaning, there is a fundamental difference between logical cases like (1)-(3) and non-logical cases like (16)-(17). I am offering a simpler, more primitive account, modelled on the way procedural metacognition works in other areas. The feeling of reliability attaching to an inference is calibrated by experience of things going well or badly when using that pattern. This applies equally whether the pattern is a logical or a non-logical one. Thinkers may not discriminate between the two kinds. In neither case is the feeling understanding-based. Indeed, it is a merit of the account that the thinker can have a feeling of reliability or unreliability in respect of an inference without its needing to be based in understanding.

The issue of understanding connects to the question of whether a mental state of intuition or intellectual seeming is involved in making the transition. An intellectual seeming is supposed to be an internalist state, available to the thinker, that justifies the thinker in making the inference (or forms part of the justification) (Chudnoff 2013, Boghossian 2018, Peacocke 2021). Seeming-correct is based on the thinker's understanding of the concepts involved. Thus, an intellectual seeming is supposed to be a way of appreciating a particular inferential pattern – it seems to the thinker that this particular conclusion follows from these particular premises. (That can be false as, perhaps, when using (18)-(19) to introduce SET.) The role I am claiming for feelings of reliability falls short of that. There are not separate feelings of reliability for different inferences. If modus ponens and AND-elimination both feel reliable to a thinker, they will generate the same feeling. The feeling is not specific to an inference pattern in the way that an intellectual seeming is (or would be, if they exist).

Nor is the thinker disposed to make the transition because of the feeling of reliability. Contrast an intellectual seeming – it is supposed to be because it seems to the thinker that the conclusion follows from the premises that they are disposed to draw the conclusion. On my account, the stimulus that activates the disposition to

---

[7] A further possibility is that there are direct, content-specific transitions involved in categorisation. The thinker is disposed to move from a perceptual representation of a certain arrangement of colours, shapes and textures, to applying the concept DOG, say. Feelings of reliability may apply to these transitions. If so, they may play a role in perceptual justification. There is not scope to pursue this line of thought here.

draw the conclusion is simply tokening the premises. Activating the disposition also generates the feeling of reliability or unreliability. That in turn affects how strongly the thinker is disposed to token the conclusion, the confidence they attach to it, and whether they are likely to pause and deliberate further.

In short, feelings of reliability are like procedural metacognition in other areas. They will not be based on understanding the concepts involved and will be simpler than the intellectual seemings relied on by other theorists. Furthermore, they plausibly apply in the same way to logical and non-logical inferences.

5.3 Epistemic Role

The ingredient we are interested in distinguishes some transitions (for Boghossian, those properly called inferences) from mere causal transitions. We saw at the outset that the ingredient has a phenomenological aspect and an epistemic aspect. So far I have focused on accounting for the phenomenological aspect. The epistemic aspect is that the ingredient should give the thinker some sort of 'sense', 'grip' or 'appreciation' that the conclusion follows from the premises. Given the objections to accounting for this in terms of explicit second-order belief, we should expect this 'appreciation' to be something less than reflective knowledge, but nevertheless something that does epistemic work for the thinker. I will argue that the feeling of reliability does indeed play that role.

My claim is that the feeling of reliability provides a means for aligning thinking dispositions with reliability. Experience that a transition has been reliable (or not) thereby has an impact on the thinker's disposition to make inferences of that type. It also affects their confidence in the conclusion. We may not know (explicitly) why we are disposed to reason in some ways rather than others. But an epistemic feeling helps shape our habits of thought so as to align them with reliability. As well as giving us a feeling that at least roughly aligns with reliability, it is an ingredient whose functional role in thought serves to help our inferential dispositions to follow patterns that have proven to be reliable. (Whether the patterns followed are in fact reliable, or not, will depend on how extensive our experience is and whether we have encountered the potentially problematic cases, as with SET.) That is to say, this feeling plays a significant epistemic role for the thinker.

Does the feeling of reliability provide the basis for making the thinker the subject of epistemic responsibility? It will not deliver the same level of epistemic responsibility as exists for beliefs that have been subjected to reflective deliberation. The most stringent standards of epistemic responsibility only apply when the thinker has had the chance to reflect on relevant considerations before coming to a judgement. However, in cases where an inference proceeds without reflective deliberation, the feeling of reliability delivers some of what we want from epistemic responsibility – it gives the thinker a capacity for modifying their principles of inference (Boghossian 2016, p. 51). If contradictions or other bad consequences eventuate from the inference, since this is registered in a lower feeling of reliability, it helps align

inferential dispositions operative on subsequent occasions with epistemically good outcomes (only approximately, but helpfully nevertheless). Similarly, if the thinker is criticised for drawing a conclusion, perhaps explicitly, but more often implicitly in the reactions of others and the fluency of the exchange, this experience can slightly adjust their inferential dispositions for the future through its impact on the feeling of reliability.

Indeed, even the explicit recognition that a particular type of transition is undesirable may not be enough, on its own, to re-wire the thinker's dispositions. Why won't they just engage the same habit again, if presented with the premises without the opportunity to reflect and recall the objection? An epistemic feeling shows up when the automatic disposition is triggered, making the thinker more cautious about drawing the conclusion or making them more likely to engage in deliberate reflection before endorsing it.

It is also an ingredient that delivers some degree of epistemic responsibility in those whose capacity for reflective deliberation is reduced or absent: animals and young children, for instance. They may lack some or all of the relevant concepts or they may lack the capacity for deliberation. Nevertheless, there is more to their epistemic life than the bare fact of whether their dispositions are or are not reliable.

Feelings of reliability also offer a dimension of epistemological assessment that goes beyond a bare externalistic assessment of whether or not an inferential disposition is reliable. They also go beyond bare epistemic entitlement (Burge 2003), since the feeling is internally available to the thinker,[8] something which enters into the way they exercise their capacity to perform the inference (while not amounting to a piece of evidence). They give the thinker what is in effect a way of monitoring whether or not an inference pattern is reliable, albeit not one that the thinker conceptualises as such (hence not something that delivers reflective knowledge that an inference is reliable). This is a useful intermediate capacity, a step on the way to full epistemic responsibility. Christopher Hookway expresses the attraction of an intermediate position between bare reliabilism and reflective deliberation:

> 'It is natural to feel that we need an intermediate position, a mean between these extremes. Each emphasises a different element of our cognitive apparatus. Crude reliabilism notes that we have habits of inference and belief formation, unthinking ways of answering questions. Recognizing their value, it concludes that our epistemic position could be satisfactory if we did nothing but rely unthinkingly upon this repertoire of habits. Cartesianism is alert to the fact that sometimes these habits of belief formation are criticized; sometimes reflection shows us that they should be revisited. And it concludes that our epistemic engine is only as well adjusted as reflection reveals it to be. The problem, then, is of seeing how far reflection should extend.' (Hookway 1994, p. 215)

---

[8] Cf, Proust (2008), who argues that a feeling can entitle a subject to form a belief.

I am offering just such an intermediate position for cases where a transition between representations is direct or primitively compelling and the thinker has not or cannot reflect on it (see also Boghossian 2001, p. 28). Epistemic feelings help to align thinking dispositions with reliability. The feeling of reliability plays a role that is intermediate between bare reliability and self-critical reflection.

The feeling of reliability is in some way self-reflective, since it is concerned with the reliability of one's own cognitive processes. But it is not a form of reflective deliberation. Nor need it be subsequently subject to reflective deliberation. We can still ask: is the feeling of reliability subject to any form of epistemic assessment? My answer is that the only test of its epistemic appropriateness is its reliability, externalistically considered. The thinker herself typically does not, and often cannot, assess the reliability of these signals. However, as theorists, we can assess how well feelings of high and low reliability track the actual reliability of the inferences to which they attach. As we saw in section (4), evidence in other domains suggests that the answer will be: not very well, but well enough to be useful (Fleming, et al. 2012, Rademaker, et al. 2012, Markovits, et al. 2015).

This solution exemplifies a broader pattern that has been proposed by others. Sosa distinguishes between animal knowledge and reflective knowledge (Sosa 1985, pp. 240-243). Animal knowledge arises when a true belief is formed by a reliable process. It is held only to the standards of bare reliabilism. Reflective knowledge requires second order beliefs about the belief. Sosa points out that the regress of justification is blocked if we don't require these second order beliefs in turn to meet the standard of reflective knowledge. If they need only count as animal knowledge, then their epistemic status is secured through their externally-assessed reliability. No further reflective beliefs, which would be third order, are needed.

On my account, the regress is blocked in a similar way – with the added virtue that there is a principled reason to distinguish between the two levels of justification. At the first level we have inferences between conscious, personal-level representations. That counts as a full-blown case of inference only if it is subject to a second level of assessment by way of procedural metacognition of inference. The epistemic standing of that signal is in turn secured simply through reliability (Proust 2013, pp. 205-6), not through its being evaluated by some third-order process.[9] The transition between explicit conceptual representations is subject to a form of internal monitoring and assessment whereas the feeling of reliability need not be. The right way of assessing the epistemic appropriateness or otherwise of the epistemic feeling is just to consider its reliability (Proust 2013, ch. 9). Epistemic feelings – which are probably present in non-human animals – are only held to the standards of animal knowledge. So when we consider how the thinker achieves knowledge on the basis of inference, the regress of justification is blocked in just the way that Sosa suggests.

---

[9] I am not denying that there could be further procedural monitoring applied to the signal. Or that it can be subject to reflective deliberation (as we are doing now). I simply deny that further monitoring is needed to secure the epistemic *bona fides* of the inference.

5.4 Representational Content

In this section I will argue that the feeling of reliability has representational content – a correctness condition. This conclusion is not needed to establish my claims about the phenomenological and epistemic roles of feelings of reliability. Nevertheless, I do think that the functional role played by the feeling of reliability in cognition indicates that it has a correctness condition, so it is worth setting out that case here. I follow those who argue that a representation can be metarepresentational without being conceptual (Shea 2014, Carruthers 2021, Carruthers and Williams 2022). A representation can have a correctness condition that concerns the content of another mental representation without the thinker having to deploy any concepts of mental states (including in agents who lack such concepts).

Epistemic feelings do not have the kind of content many have wanted the additional ingredient to display. As we have seen, a feeling of reliability does not make use of any epistemic or mental state concepts, nor does it require the thinker to have such concepts. It is an unstructured signal that plays a certain functional role. Since it is not composed out of concepts, it is non-conceptual according to a (vehicle-based) version of the state view of non-conceptual content (Heck 2007, the compositional nonconceptualism of Crowther 2006). More carefully, it is a *not-conceptually-compositional* representation.

An unstructured vehicle can have representational content without the thinker knowing, in any more explicit way, what it means. It may have a correctness condition and/or a satisfaction condition. I will focus on the former. A content will be meta-level if it concerns the thinker in the right way. I will take it that a not-conceptually-compositional representation is metarepresentational if its correctness condition concerns a contentful feature of a mental representation or a cognitive process (Shea 2014). That definition does not tell us what it takes to be able to have not-conceptually-compositional metarepresentations. However, note that the definition does not build in a requirement for the thinker to have self-directed concepts or be capable of reflective deliberation. It has been argued that vehicles involved in relatively straightforward computations, widespread in the animal kingdom, can acquire metarepresentational contents (Shea 2014). I will argue that the functional role of the feeling of reliability is evidence that is has metarepresentational content.

Epistemic feelings do not generally wear their significance on their sleeve. They are generated by cognitive processes and have certain downstream effects. Thinkers can learn more about their significance and how to rely on them (Heyes, et al. 2020). But that is not because the feeling, on its own, enables them to know what the feeling is about or what it is supposed to track. Nor does the signal reveal whether or not is a reliable signal. It is like the feeling of rightness that arises when there is an intuitive answer to a problem (like the bat and ball problem). In many contexts fluency is a reasonable guide to accuracy, but not for some types of problem (Ackerman and Thompson 2017). So a student can learn that, in a multiple choice test say, they should not rely on the feeling of reliability, but doubt it and take time to think again. The most

fluent answer is now taken to be a lure. (Similarly when doing psychologists' tricky problems like the Cognitive Reflection Test, Frederick 2005.)

I will assume that what a not-conceptually-compositional vehicle correlates with, together with its functional role, gives us good evidence as to its content (Shea 2013). These facts may serve to constitute the vehicle has having the content it does (Shea 2018), but I don't need to rely on that stronger claim here. We have seen that reflection on the examples, plus some psychological evidence, gives us an indication of the role played by the feeling of reliability in unfolding cognitive processes. I have argued that it will affect the thinker's confidence in the conclusion, hence how likely they are to rely on the inference; also whether they will engage in deliberation or seek further information. If the conclusion is implausible but the inference seems reliable, they may revisit the truth of the premises. The feeling is likely to be calibrated from feedback as to whether things have gone well with the inference in the past (Koriat, et al. 2006). A feeling of unreliability may make the thinker less disposed to make the inference at all. Where the inference generates contradictions or action plans that fail to succeed, that is likely to gradually erode the associated sense of reliability. Where things have gone smoothly, a less-remarked-upon fluency will prevail. So we can expect the feeling of reliability or unreliability to correlate to some extent with whether the inference to which it attaches is in fact reliable. Evidence about other epistemic feelings suggests that the correlation with reliability is unlikely to be very tight, but will still be strong enough to be useful.

This collection of functional roles suggests that the feeling is being relied upon in cognition because of the way it correlates with reliability. So is has a correctness condition along the lines of *this transition is reliable* (where reliable = the conclusion is likely to be correct if the premises are). As always with not-conceptually-compositional representations, although we are using a sentence to capture the correctness condition, the representation has none of the structure of the sentence we use to give its content. The 'this' reflects the fact that the same feeling is generated by different inferences. The correctness condition of a token feeling concerns the inference that generates that feeling. Whether this means that the content should properly be regarded as indexical, rather than non-indexical and context dependent, is not an issue that need detain us here. What I want to argue is that, if the feeling of reliability is produced, used and modified in the way I have suggested, then a token feeling plausibly represents the reliability of the inference which generates it.

(6) Role of Epistemic Feelings in Thinkers Capable of Reflective Deliberation

I have argued that the feeling of reliability has an intermediate epistemic status. It delivers some but not all of what we want from epistemic responsibility. It is a route by which a thinker's various dispositions to reach conclusions on the basis of basic or primitively compelling transitions become sensitive to whether those dispositions have proven to be reliable. This makes it an important precursor to full epistemic responsibility.

Full epistemic responsibility calls for more. Most adults are capable of deliberation (system 2 cognition). We also possess mental state and epistemic concepts, concepts of belief, knowledge, following, reliability and justification. So we are capable of reflective deliberation. It is this full suite of capacities that underpins the more demanding species of epistemic status, namely being the subject of epistemic responsibility (Smithies 2016, pp. 65-66). This could be a matter of a solipsistic thinker reflecting in isolation, but it is most obvious inter-personally. We ask one another to justify our beliefs and judgements (Mercier and Sperber 2011). Our answers are subject to praise and blame, to others' reactive attitudes. We understand those normative demands and can choose to regulate our beliefs and attitudes accordingly. That depends on our ability to use concepts of mental states, reliability and justification in our deliberation (Burge 1998, p. 262). That is, it depends on having the capacity for system 2 metacognition (Shea, et al. 2014).

Procedural metacognition does not drop out of the picture for such thinkers. I will argue that feelings of reliability and unreliability play an important role in enabling reflective thinkers to comply with the demands of epistemic norms.

For those capable of reflective deliberation, the disposition to make a transition between representations can itself become subject to deliberation. 'I back up and bring that impulse into view' (Korsgaard 1996, p. 93). This depends on the thinker being aware of the conclusion and the grounds on which they reached it (Valaris 2017). Contrast the subpersonal computations that take place within the visual system. Transitions that generally produce accurate perceptual representations may also generate illusions. The thinker cannot rationally assess transitions taking place within the visual system to work out what has gone wrong. There is little prospect of changing subpersonal transitions through reflecting on them. The ability to reflectively intercede on our inferential dispositions depends upon materials that are conscious. Since epistemic feelings are conscious, such feelings can form part of this process.

When we 'bring an impulse into view', reflect on it, and come to doubt it, that will prevent us from forming the unreliable conclusion on that occasion. But reflection will have only a transient influence unless it can affect our thinking dispositions for future occasions. In most circumstances our thinking unfolds without our reflecting on the inferential dispositions on which we rely. There may be no time to do so, for example in the middle of a conversation. It is in any event a process that must stop somewhere on pain of regress. An epistemic principle would be ineffective if reflection at the point of inference were always required to comply with it. What we need is that reflective deliberation should be capable of effecting long term changes to our basic inferential dispositions. Since deliberation takes place with conscious representations, it is sensitive to the agent's goals and projects (Carruthers 2015, pp. 156-8). (Non-conscious representations are often processed in ways that are encapsulated from agent-level beliefs and goals.) When deliberation is effective in changing the agent's inferential dispositions, it is therefore a way of bringing those dispositions into closer alignment with the agent's personal-level goals and projects. To account for the intuition that flexibility is characteristic of personal level inference (Kornblith 2016), it may be that this kind of flexibility that is required.

I have argued that feelings of reliability and unreliability are one route by which our inferential dispositions are changed as a result of experience. To complete the picture, I only need the additional claim that a negative reflective assessment can impact the feeling of reliability. Just as generating a contradiction plausibly affects fluency and hence feelings of reliability on future occasions, so too it is plausible that judging that an inferential pattern is faulty will reduce the feeling of reliability that attaches to it on future occasions. As we saw, this may be a route by which philosophy students become less inclined to introduce the concept SET using transition (18) to (19). Epistemic feelings are, then, a route by which reflective deliberation can have an impact on our inferential dispositions into the future.

The negative epistemic judgement may be self-generated, but it is more likely to come from other people. We are explicitly instructed in how to reason well; we get implicit signals, arising from whether the conclusions we reach conform with others' expectations; and often we are also subject to social censure for our habits of thought. These practices of holding one another responsible for patterns of thinking include the epistemic, while extending to norms of other kinds. For example, we rightly censure someone when they make an inference that is implicitly sexist (e.g. 'when a surgeon is operating … he …'). Arguably it is engaging in these practices that makes a person the subject of epistemic responsibility (Smithies 2016). It is even arguable that the social practice of giving and assessing reasons evolved first (Mercier and Sperber 2011), and was only subsequently directed inwards (see also Carruthers 2011). Either way, the feeling of reliability is one route whereby this kind of social feedback can affect our future dispositions.

We have already seen an important influence in the other direction. Why does the thinker 'back up and bring an impulse into view'? One reason is that the transition has generated a feeling of unreliability. Epistemic feelings prompt the thinker to engage in deliberation. They provide a signal of where best to direct deliberative resources. Furthermore, for a thinker who has the concept of reliability, it is eminently plausible that they can apply the concepts RELIABLE and UNRELIABLE TO their own inferential dispositions on the basis of the corresponding epistemic feelings.

The full picture has three types of transition. First, there are cases where the thinker is moved by contents without exercising mental agency. That may be the whole story for some organisms. Second, there are the personal level inferences that have been the focus of this paper. They are accompanied by a feeling of reliability and can occur in agents who lack reflective concepts or are incapable of deliberation. Third, there are thinkers who are capable of deliberatively reflecting on their inferential transitions. At levels two and three, the feeling of reliability arises automatically from the inference process, and is calibrated diachronically by the downstream consequences of performing inferences of that type; at the third level, this includes calibration through explicit reasoning about how one ought to reason (which is often social). Feelings of reliability and unreliability affect the thinker's reasoning dispositions into the future: how likely they are to reach a conclusion or to stop and reflect on it, and the confidence they will attach to it. Epistemic feelings are an important means by which a thinker's inferential habits become aligned with epistemic norms. As well as being a precursor to epistemic responsibility in those who

are unreflective, feelings of reliability play an important epistemic role in agents who are subjects of epistemic responsibility.

(7) Conclusion

Feelings of reliability are part of a broader account of transitions that occur between mental representations and why some of them amount to an exercise of mental agency. I have put forward an account with three levels. At the first level there are content-respecting transitions. The insight behind the 'cognitive revolution' was that intelligent behaviour can be the result of causal transitions between representational vehicles which respect their contents. This is the most fundamental and most important level. It underlies all forms of reasons-responsive or broadly rational behaviour. These transitions can be set up and stabilised as organisms interact with the world, both over evolutionary time, and through learning in the lifetime of an individual. They can take place subpersonally, without being coordinated at the level of the whole agent. And the agent need have no appreciation that the conclusion of such a transition follows from the premises. They unfold algorithmically with no additional ingredient.

At the second level are inferences performed by the person (or other agent). I have argued that personal-level inferences are subject to procedural metacognition: feelings of reliability or unreliability attach to transitions between personal-level representations. These epistemic feelings roughly track the reliability of a pattern of inference. That is reflected in their functional role. Metacognitive monitoring and control mediated by epistemic feelings is likely to apply to inferences just as it does to perceptual decisions, memory and occurrent beliefs. These processes are found in animals and young children who do not possess epistemic or mental state concepts.

At the third and most sophisticated level of cognition are thinkers like us – philosophers are the paradigm – who engage in reflective deliberation about their inferential dispositions, deploying concepts of justification, consequence, reliability and belief. Because we can deliberate in this way to assess whether we are justified in the conclusions we reach by inference, and more broadly in our beliefs and choices, we become targets of epistemic responsibility, and able to assess one another's reactive attitudes. We then have some capacity to change our inferential dispositions as a matter of deliberate choice in way that reflects our goals and values.

Acknowledgements

For discussion and comments, the author would like to thank Dominic Alford-Duguid, Marc Artiga, Paul Boghossian, John Broome, Manolo Martinez, John Morrison, David Papineau, Christopher Peacocke, Declan Smithies, Michael Strevens, two referees for this journal; and audiences at the Institute of Philosophy, the Oxford Mind work-in-progress-group, the NYU Brown Bag seminar, the Association for the Scientific Study of Consciousness annual meeting, and the *Metacognition, Consciousness, and Agency*

workshop at the University of Barcelona. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 681422 (MetCogCon).

References

Ackerman, R., and V. A. Thompson. 2017. "Meta-Reasoning: Monitoring and Control of Thinking and Reasoning," *Trends in Cognitive Sciences*, **21**: 607-17.

Adams, Allison, and Angelo Santi. 2011. "Pigeons Exhibit Higher Accuracy for Chosen Memory Tests Than for Forced Memory Tests in Duration Matching-to-Sample," *Learning & Behavior*, **39**: 1-11.

Arango-Muñoz, Santiago. 2014. "The Nature of Epistemic Feelings," *Philosophical Psychology*, **27**: 193-211.

Bajšanski, Igor, Valnea Žauhar, and Pavle Valerjev. 2019. "Confidence Judgments in Syllogistic Reasoning: The Role of Consistency and Response Cardinality," *Thinking & Reasoning*, **25**: 14-47.

Boghossian, Paul. 2001. "How Are Objective Epistemic Reasons Possible?," *Philosophical Studies*, **106**: 1-40.

———. 2014. "What Is Inference?," *Philosophical Studies*, **169**: 1-18.

———. 2016. "Reasoning and Reflection: A Reply to Kornblith," *Analysis*, **76**: 41-54.

———. 2018. "Delimiting the Boundaries of Inference," *Philosophical Issues*, **28**: 55-69.

———. 2019. "Inference, Agency and Responsibility". In Jackson and Jackson, eds, *Reasoning: New Essays on Theoretical and Practical Thinking*. Oxford / New York: OUP, 101-24.

Bona, Silvia, and Juha Silvanto. 2014. "Accuracy and Confidence of Visual Short-Term Memory Do Not Go Hand-in-Hand: Behavioral and Neural Dissociations," *PLOS one*, **9**: e90808.

Brewer, Bill. 1995. "Mental Causation: Compulsion by Reason," *Proceedings of the Aristotelian Society, Supplementary Volume*, **69**: 237-53.

Broome, John. 2013. *Rationality through Reasoning*. Oxford: Wiley-Blackwell.

Burge, Tyler. 1998. "Reason and the First Person". In Wright, Smith and Macdonald, eds, *Knowing Our Own Minds*. Oxford: OUP, 243-70.

———. 2003. "Perceptual Entitlement," *Philosophy and Phenomenological Research*, **67**: 503-48.

Carroll, Lewis. 1995. "What the Tortoise Said to Achilles," *Mind*, **104**: 691-93.

Carruthers, Peter. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

———. 2015. *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. OUP Oxford.

———. 2021. "Explicit Nonconceptual Metacognition," *Philosophical Studies*, **178**: 2337-56.

Carruthers, Peter, and D. M. Williams. 2022. "Model-Free Metacognition," *Cognition*, **225**: 105117.

Chudnoff, Elijah. 2013. *Intuition*. Oxford / New York: OUP.

Cortese, Aurelio, Kaoru Amano, Ai Koizumi, Mitsuo Kawato, and Hakwan Lau. 2016. "Multivoxel Neurofeedback Selectively Modulates Confidence without Changing Perceptual Performance," *Nature communications*, **7**: 13669.

Crowther, Timothy M. 2006. "Two Conceptions of Conceptualism and Nonconceptualism," *Erkenntnis*, **65**: 245-76.

De Neys, Wim, Oshin Vartanian, and Vinod Goel. 2008. "Smarter Than We Think: When Our Brains Detect That We Are Biased," *Psychological Science*, **19**: 483-89.

De Neys, Wim, E. Moyens, and D. Vansteenwegen. 2010. "Feeling We're Biased: Autonomic Arousal and Reasoning Conflict," *Cogn Affect Behav Neurosci*, **10**: 208-16.

De Neys, Wim, Sofie Cromheeke, and Magda Osman. 2011. "Biased but in Doubt: Conflict and Decision Confidence," *PLOS one*, **6**: e15954.

De Neys, Wim. 2012. "Bias and Conflict: A Case for Logical Intuitions," *Perspectives on Psychological Science*, **7**: 28-38.

Dokic, Jérôme. 2014. "Feeling the Past: A Two-Tiered Account of Episodic Memory," *Review of philosophy and psychology*, **5**: 413-26.

Dummett, Michael. 1991. *The Logical Basis of Metaphysics*. Harvard university press.

Ernst, M. O., and M. S. Banks. 2002. "Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion," *Nature*, **415**: 429-33.

Evans, JSBT, Julie L Barston, and Paul Pollard. 1983. "On the Conflict between Logic and Belief in Syllogistic Reasoning," *Memory & Cognition*, **11**: 295-306.

Fernandez Cruz, A. L., S. Arango-Munoz, and K. G. Volz. 2016. "Oops, Scratch That! Monitoring One's Own Errors During Mental Calculation," *Cognition*, **146**: 110-20.

Fleming, Stephen M, R. S. Weil, Z. Nagy, R. J. Dolan, and G. Rees. 2010. "Relating Introspective Accuracy to Individual Differences in Brain Structure," *Science*, **329**: 1541-3.

Fleming, Stephen M, Josefien Huijgen, and Raymond J Dolan. 2012. "Prefrontal Contributions to Metacognition in Perceptual Decision Making," *Journal of Neuroscience*, **32**: 6117-25.

Foote, Allison L., and Jonathon D. Crystal. 2007. "Metacognition in the Rat," *Current Biology*, **17**: 551-55.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making," *Journal of Economic perspectives*: 25-42.

Goupil, Louise, and Sid Kouider. 2016. "Behavioral and Neural Indices of Metacognitive Sensitivity in Preverbal Infants," *Current Biology*, **26**: 3038-45.

Goupil, Louise, Margaux Romand-Monnier, and Sid Kouider. 2016. "Infants Ask for Help When They Know They Don't Know," *Proceedings of the National Academy of Sciences*: 201515129.

Hampton, R. R. 2001. "Rhesus Monkeys Know When They Remember," *Proceedings of the National Academy of Sciences of the United States of America*, **98**: 5359-62.

Heck, Richard. 2007. "Are There Different Kinds of Content?". In Cohen and McLaughlin, eds, *Contemporary Debates in the Philosophy of Mind*. Oxford: Blackwell.

Hertz, Uri, Stefano Palminteri, Silvia Brunetti, Cecilie Olesen, Chris D Frith, and Bahador Bahrami. 2017. "Neural Computations Underpinning the Strategic Management of Influence in Advice Giving," *Nature communications*, **8**: 2191.

Heyes, Cecilia M, D. Bang, Nicholas Shea, Chris D. Frith, and S. M. Fleming. 2020. "Knowing Ourselves Together: The Cultural Origins of Metacognition," *Trends Cogn Sci*, **24**: 349-62.

Hookway, Christopher. 1994. "Cognitive Virtues and Epistemic Evaluations," *International Journal of Philosophical Studies*, **2**: 211-27.

Kepecs, Adam, Naoshige Uchida, Hatim A. Zariwala, and Zachary F. Mainen. 2008. "Neural Correlates, Computation and Behavioural Impact of Decision Confidence," *Nature*, **455**: 227-31.

Koriat, Asher, and Ravit Levy-Sadot. 2001. "The Combined Contributions of the Cue-Familiarity and Accessibility Heuristics to Feelings of Knowing," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **27**: 34.

Koriat, Asher, H. Ma'ayan, and R. Nussinson. 2006. "The Intricate Relationships between Monitoring and Control in Metacognition: Lessons for the Cause-and-Effect Relation between Subjective Experience and Behavior," *J Exp Psychol Gen*, **135**: 36-69.

Koriat, Asher, and Tore Helstrup. 2007. "Metacognitive Aspects of Memory". In Magnussen, ed, *Everyday Memory*. Hove: Psychology Press, 251.

Koriat, Asher. 2008. "Subjective Confidence in One's Answers: The Consensuality Principle," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**: 945.

Kornblith, Hilary. 2012. *On Reflection*. Oxford: Oxford University Press.

———. 2016. "Replies to Boghossian and Smithies," *Analysis*, **76**: 69-80.

Korsgaard, Christine M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Lea, R Brooke, Elizabeth J Mulligan, and Jennifer Lee Walton. 2005. "Accessing Distant Premise Information: How Memory Feeds Reasoning," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**: 387-95.

Lee, Sang Wan, Shinsuke Shimojo, and John P. O'Doherty. 2014. "Neural Computations Underlying Arbitration between Model-Based and Model-Free Learning," *Neuron*, **81**: 687-99.

Markovits, H., V. A. Thompson, and J. Brisson. 2015. "Metacognition and Abstract Reasoning," *Mem Cognit*, **43**: 681-93.

Mercier, Hugo, and Dan Sperber. 2011. "Why Do Humans Reason? Arguments for an Argumentative Theory," *Behavioral and Brain Sciences*, **34**: 57-74.

Middlebrooks, Paul G., and Marc A. Sommer. 2012. "Neuronal Correlates of Metacognition in Primate Frontal Cortex," *Neuron*, **75**: 517-30.

Millikan, Ruth Garrett. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

Moran, Rani, Andrei R Teodorescu, and Marius Usher. 2015. "Post Choice Information Integration as a Causal Determinant of Confidence: Novel Data and a Computational Account," *Cognitive Psychology*, **78**: 99-147.

Moreira, Caio M, Max Rollwage, Kristin Kaduk, Melanie Wilke, and Igor Kagan. 2018. "Post-Decision Wagering after Perceptual Judgments Reveals Bi-Directional Certainty Readouts," *Cognition*, **176**: 40-52.

Oaksford, Mike, and Nick Chater. 1994. "A Rational Analysis of the Selection Task as Optimal Data Selection," *Psychological Review*, **101**: 608.

Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.

———. 1996. "Entitlement, Self-Knowledge and Conceptual Redeployment," *Proceedings of the Aristotelian Society*, **96**: 117-58.

———. 2021. "Debating the a Priori, by Paul Boghossian and Timothy Williamson," *Mind*.

Pescetelli, Niccolò, Anna-Katharina Hauperich, and Nick Yeung. 2021. "Confidence, Advice Seeking and Changes of Mind in Decision Making," *Cognition*, **215**: 104810.

Prior, Arthur. 1967. "The Runabout Inference Ticket". In Strawson, ed, *Philosophical Logic*. Oxford: O.U.P.

Proust, J. 2008. "Epistemic Agency and Metacognition: An Externalist View," *Proceedings of the Aristotelian Society*, **108**: 241-68.

———. 2010. "Metacognition," *Philosophy Compass*, **5**: 989-98.

———. 2012. "Metacognition and Mindreading: One or Two Functions?". In Beran, Brandl, Perner and Proust, eds, *Foundations of Metacognition*. Oxford: OUP, 234-51.

———. 2013. *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford University Press.

Prowse Turner, Jamie A., and Valerie A. Thompson. 2009. "The Role of Training, Alternative Models, and Logical Necessity in Determining Confidence in Syllogistic Reasoning," *Thinking & Reasoning*, **15**: 69-100.

Quilty-Dunn, Jake, and Eric Mandelbaum. 2017. "Inferential Transitions," *Australasian Journal of Philosophy*, **96**: 532-47.

Rademaker, R. L., C. H. Tredway, and F. Tong. 2012. "Introspective Judgments Predict the Precision and Likelihood of Successful Maintenance of Visual Working Memory," *J Vis*, **12**: 21.

Rounis, Elisabeth, Brian Maniscalco, John C Rothwell, Richard E Passingham, and Hakwan Lau. 2010. "Theta-Burst Transcranial Magnetic Stimulation to the Prefrontal Cortex Impairs Metacognitive Visual Awareness," *Cognitive neuroscience*, **1**: 165-75.

Shea, Nicholas. 2013. "Naturalising Representational Content," *Philosophy Compass*, **8**: 496-509.

———. 2014. "Reward Prediction Error Signals Are Meta-Representational," *Nous*, **48**: 314-41.

Shea, Nicholas, Annika Boldt, Dan Bang, Nick Yeung, Cecilia Heyes, and Chris D. Frith. 2014. "Supra-Personal Cognitive Control and Metacognition," *Trends in Cognitive Sciences*, **18**: 186-93.

Shea, Nicholas. 2018. *Representation in Cognitive Science*. Oxford: Oxford University Press.

Shea, Nicholas, and Chris D Frith. 2019. "The Global Workspace Needs Metacognition," *Trends in Cognitive Sciences*, **23**: 560-71.

Shea, Nicholas. 2023a. "Moving Beyond Content-Specific Computation in Artificial Neural Networks," *Mind & Language*, **38**: 156–77.

———. 2023b. "Millikan's Consistency Testers and the Cultural Evolution of Concepts," *Evolutionary Linguistic Theory*, **5**: 79-101.

Shynkaruk, Jody M, and Valerie A Thompson. 2006. "Confidence and Accuracy in Deductive Reasoning," *Memory & Cognition*, **34**: 619-32.

Siegel, Susanna. 2019. "Inference without Reckoning". In Jackson and Jackson, eds, *Reasoning: New Essays on Theoretical and Practical Thinking*. Oxford / New York: OUP.

Smithies, Declan. 2016. "Reflection On: On Reflection," *Analysis*, **76**: 55-69.

Sosa, Ernest. 1985. "Knowledge and Intellectual Virtue," *The Monist*, **68**: 226-45.

Thompson, Valerie A., Jamie A. Prowse Turner, and Gordon Pennycook. 2011. "Intuition, Reason, and Metacognition," *Cognitive Psychology*, **63**: 107-40.

Thompson, Valerie A., Jonathan St B. T. Evans, and Jamie I. D. Campbell. 2013. "Matching Bias on the Selection Task: It's Fast and Feels Good," *Thinking & Reasoning*, **19**: 431-52.

Thompson, Valerie A., and Stephen C. Johnson. 2014. "Conflict, Metacognition, and Analytic Thinking," *Thinking & Reasoning*, **20**: 215-44.

Valaris, Markos. 2017. "What Reasoning Might Be," *Synthese*, **194**: 2007-24.