That Seems Right: Reasoning, Inference, and the Feeling of Correctness

Jeremy Wolos

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

ABSTRACT

That Seems Right: Reasoning, Inference, and the Feeling of Correctness

Jeremy Wolos

In my dissertation, I advance and defend a broad account of reasoning, including both the nature of inference and the structure of our reasoning systems. With respect to inference, I argue that we have good reason to consider a unified account of the cognitive transitions through which we attempt to figure things out. This view turns out to be highly inflationary relative to previous philosophical accounts of inference, which, I argue, fail to accommodate many instances of everyday reasoning. I argue that a cognitive transition's status as an inference, in this broad sense, depends on the subject's taking the conclusion of the inference— a new, revised, or supposed belief— to follow from a trustworthy internal process. Furthermore, taking such a belief to follow from a trustrworthy process consists in its accompaniment by the *feeling of correctness* to the subject, which I call the assent affect. With respect to the structure of our reasoning systems, I defend a dual process model of reasoning by addressing certain alleged deficiencies with such accounts. I argue that the assent affect— or more precisely its absence— is a strong candidate to serve as the triggering condition of our type 2 reasoning processes. That is, a subject's more effortful reasoning processes engage with a problem when the output of a type 1 intuition is not accompanied by the assent affect. This account, I argue, fits well with both empirical and theoretical claims about the interaction of dual reasoning processes. In this dissertation, I use the assent affect to solve puzzles about both the nature of inferences and the structure of our reasoning systems. Puzzles in rationality become easier to solve when our intellectual feelings are not excluded from the picture.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This dissertation would not have been completed without the support of my advisor, John Morrison. Beyond his academic guidance, which has been considerable, he has served as a mentor and role model since before I accepted my offer of admission to Columbia University. He has been an unwavering source of encouragement, advice, and polite insistence on dissertation progress. I could not be more grateful for his help.

I would also like to thank fellow committee members Christopher Peacocke and Eric Mandelbaum, both of whom have contributed immensely to the intellectual content of my dissertation. Their questions, objections, and recommendations for further reading improved the quality and rigor of this thesis tremendously.

Finally, I would like to thank my family. The burden of completing this dissertation was lightened immensely by the emotional encouragement of my mother, Stephanie Wolos and the practical advice of my father, Dr. Gregory Wolos, my sister, Dr. Cassandra Pattanayak, and my brother-in-law, Dr. Vikram Pattanayak, all successful dissertators themselves. My family helped me keep my eyes on the finish line of this project.

Chapter 1 – Inferences and the assent affect

1. Introduction

Thinking consists of thoughts and the transitions between them. The metaphysics of mind ought to be concerned not only with what types of thoughts we have, but also with how we get from one to the next. This chapter focuses on the class of cognitive transitions that are involved in reasoning. What can we say about the cognitive transitions through which we attempt to figure things out?

I call my target phenomenon in this chapter "inference," but my construal is highly inflationary relative to previous accounts. A prominent philosophical tradition on inference began with Frege (1881/1979) and has resurfaced recently in the work of Paul Boghossian (2014), Crispin Wright (2014), and John Broome (2013, 2014). This tradition holds inferences to a formal standard. On such a view, an inference is a transition from content to content, based on paradigmatically deductive rule-following. Accounts along these lines are, to varying degrees, oriented towards explaining certain special epistemological properties of inference: how are inferred beliefs justified by the beliefs that precede them?

When we reason, we often reach new beliefs in ways that are not captured by this type of traditional, formal account of inference. This is not an accident: such accounts generally do not aspire to explain the transitions that serve much of our everyday reasoning, which may not meet a formal standard. We can make this clear with a few examples:

(a) A logic student correctly applies deductive rules like modus ponens on an exam.

(b) A father reasons that, due to a spike in shark attacks in South Africa, he should not allow his children to swim during their Cape Cod vacation, a fallacy of weak induction.

(c) While absent-mindedly playing a game on her smartphone, a woman solves a puzzle by relying on a primitive matching bias rather than rule-based reasoning.

Accounts of inference in the Frege-Boghossian tradition are designed to explain phenomena like (a). I will argue that they are worse at explaining both bad and inductive inferences, like (b). And such accounts do not regard (c) as an inference, even while reasoning like (c) may be functionally very similar to more "proper" inferences.

I am interested in all three of these. A number of cognitive transitions result in beliefs and serve our reasoning activity, and I intend to give a unified account of these transitions. There are several reasons to be interested in a unified account of the cognitive transitions that factor into reasoning— to be interested in an account of inference broadly construed.

First, the commonsense category of inference includes each of the cognitive transitions above. Anytime a subject regards herself as having engaged in good reasoning, the event ought to be considered an inference by a commonsense standard. The unified account I offer in this chapter can be regarded as an account of commonsense inference. Boghossian agrees that a subject's regard for her own reasoning is important to a transition's status as an inference, but in Section 2 I will argue that his account is not well-equipped to extend to the full range of transitions I am considering.

Second, a unified account is useful because inferences that meet a formal standard and those that do not have quite a lot in common. They often serve the same reasoning functions, and it may be difficult to distinguish between a rule-based inference that meets a formal standard and a cognitive transition that has traditionally been considered sub-inferential. By offering an account of the general category to which all inferences (broadly construed) belong, we can more easily understand different types of inference and the relations between them. We will also be better able to individuate particular inferences, accounting for both a "macro" inference a subject might perform when reaching a grand conclusion, as well as the constituent inferences that served that broader goal.

Finally, a broad, unified account of inference will be useful in addressing other types of questions about cognition. The later chapters of this dissertation are devoted to dual process models of reasoning. Very generally, the dual process hypothesis is that there are two distinct types of cognitive activity under the umbrella of "reasoning": one is fast, automatic, effortless; the other is slow, deliberate, effortful. Despite the prominence of dual process models, I believe that the literature is missing a coherent account of the category of cognitive events over which the dual process hypothesis quantifies: *what* comes in two varieties? I intend the unified account of inference I offer in this chapter to fill this void. My construal of inference, by design, admits any instance of Type 1 or Type 2 reasoning. The account of inference in this chapter will fix the target phenomena for the dual process model of reasoning I develop in the following chapters, and will be essential as I respond to certain recent critiques of such models.

I argue that an inference (broadly construed) involves a subject's taking a new or revised belief[1] to be the output of a trustworthy internal process. This represents a departure from the Frege-Boghossian tradition on inference, which places requirements on the subject's attitude towards the premises that lead to an inferred conclusion. In order to infer, Boghossian says a subject must take her premises to support her conclusion. I argue that such an account, in which the subject must have access to the causal history of inferred belief, is not extendable to the broader category of inference I wish to consider.

Once our orientation of "taking" is shifted towards conclusion-beliefs rather than premise-beliefs, it is necessary to consider what it means to take a belief to be the output of a trustworthy internal process. I argue that the best candidate to serve this taking function in all inferences— both good and bad— is the feeling of correctness that accompanies inferred beliefs. I argue that this feeling is the phenomenology that accompanies the intuition that a conclusion-belief followed from a trustworthy internal process. I refer to the feeling as the "assent affect," and argue that its presence is sufficient for a cognitive transition to count as an inference.

In Section 2, I will consider certain general features of inferences. I will argue that an inference is a causal process resulting in the endorsement of a belief. Boghossian's recent account of inference holds that this endorsement consists of a subject "taking" her premises to support her conclusion. I will agree with intuition behind Boghossian's "taking condition" on inference, but I will argue that it cannot be extended to include a number of commonsense inferences. In Section 2.2, I will argue that subjects may be

[1] There is a subset of inferences which to not yield beliefs: hypothetical inferences. I will discuss these cases in Section 3, where I argue that they are a subcategory of intellectual inferences. These are exceptional rather than typical cases, on my view.

ignorant of or confused about the premises on which an inference was in fact based, and I will dismiss the view that inference requires a subject to take her premises to be a particular way. I will also argue, relatedly but distinctly, that subjects are often blind to the rules linking their premises with their apparently inferred conclusion. I conclude that taking cannot involve the premises of an inferred belief to be any particular way at all.

Having dismissed this construal of taking, one is left with two options, which I consider in Section 3. One can deny that inferences involve taking at all, or one can argue that inference involves a form of taking that does not involve regard for either an inferred belief's premises or the rules tying premises to conclusions. I deny the first option and accept the second. A version of the first-option has recently been endorsed by Jake Quilty-Dunn and Eric Mandelbaum. I will argue that accounts along these lines struggle to give appropriate individuation conditions for inference.

In Section 4, I will catalogue the diverse types of inference that ought to be included in a unified account, and I will argue that the view I lay out is best positioned to cover them all.

In Section 5, I will explain what it is for a subject to take a conclusion to follow from a trustworthy internal process. Here, I will develop my account of the assent affect, the phenomenology that accompanies all commonsense inference. I will argue that the mere presence of this feeling and the intuition it reflects is sufficient for a cognitive transition to count as an inference.

Before I move on to discuss the general features of inference, I would like to briefly discuss the historical basis of the relevance of phenomenology to accounts of reasoning.

Wittgenstein, a pivotal figure in the rule-following tradition about inference that I hope to subvert, occasionally describes the process of inference in phenomenological terms. In his *Investigations* §219, he describes what it is like to infer based on a rule that has been internalized:

> "All the steps are really already taken" means: I no longer have any choice. The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space. – But if something of this sort really were the case, how would it help?
> No; my description only made sense if it was to be understood symbolically. – I should have said: This is how it strikes me. When I obey a rule, I do not choose. I obey the rule *blindly*.

Crispin Wright notes about this passage that "blindness" here refers to the phenomenology of immediacy— to follow a rule one has internalized may feel as if one is not following a rule at all (2007, 490). There is something it is like, in other words, to obey an inference rule on a Wittgensteinian picture.

I do not intend to argue that Wittgenstein would agree with the particular account of inference I am going to offer, or for the claims that follow to hinge upon interpretation of Wittgenstein. This excerpt is meant merely to note the persistent impulse among philosophers to describe the most fundamental trustworthy internal processes in terms of feelings and phenomenology. These excerpts highlight that the issues surrounding inference have historically been difficult to account for without incorporating or mentioning phenomenological considerations, as I do in this chapter.

2. The general features of inference

2.1 "Taking" and other conditions on inference

In this section, I will develop some initial conditions on inference. I will agree with Boghossian about certain critical features, but I will argue that his account cannot be extended to cover the full range of commonsense inferences, as I intend to.

Inferences are transitions that result in conclusions. In general, these conclusions will be beliefs; in cases of supposition, we infer hypothetical conclusions we do not necessarily believe. For ease of explanation, I will proceed in this section as if all inferences result in beliefs. In Section 4, once my view has been laid out, I will explain how cases of supposition—a subcategory of inference— fit into the picture.

Inferences come in different forms: sometimes we might infer a new belief, other times we might infer that an existing belief ought to be adjusted or revised, and still other times we might "check our work" via inference, confirming an existing belief.

One straightforward preliminary is that one must infer *from* some basis. If I announce that I have inferred that it will rain tomorrow, my audience will be curious to hear from what I have inferred the weather of the future. Perhaps the basis of this inference is the projection on a weather app, or perhaps the basis of this inference is a pattern that I have personally observed— maybe my elbows itch at the moment, as they often do the day before a rainstorm. The quality of these potential bases of inferences notwithstanding, one cannot make an entirely baseless inference. In other words, if I were to respond to my audience's question by saying, "I have inferred that it will rain tomorrow from no basis whatsoever," they would correctly deny that I had made an inference at all.

I am very liberal about what might serve as the basis for an inferred belief. Others, like Boghossian, have argued that the basis of an inferred belief must be some prior belief or set of beliefs. My eventual account will be much more inclusive about what might have figured into the causal history of an inferred belief

My goal is to produce a set of necessary conditions for a cognitive transition to count as an inference. These conditions should conjointly be sufficient for a transition's status as inference. So far, I have agreed with Boghossian on two initial conditions:

(a) the transition results in a new, revised, or reaffirmed belief

(b) this belief is caused by some prior states or events

Not all transitions meeting these conditions should count as inferences, however. Consider how the problem of deviant causal chains might apply in cases of inference. Suppose that my cell phone rings, which causes me to believe my phone is ringing, which causes me to reach my hand towards my pocket, which causes an overzealous police officer to shoot me, which causes me to believe I have been shot. While my belief that my phone is ringing is part of the causal explanation for my belief that I have been shot, it is incorrect to say that I inferred from the fact that my phone is ringing that I have been shot. One belief does not count as inferred from another simply because the latter appeared in the causal chain preceding the former.

What more is necessary for an inference? Boghossian observes that, in order for a new or revised belief to count as having been inferred, there must be some constraint on the subject's attitude towards the apparently inferred belief. Specifically, the subject must take this belief to be a conclusion that is supported by some other facts that the subject

believes to be true. Boghossian follows Frege along these lines: "To make a judgment because we are cognisant of other truths as providing a justification for it is known as *inferring*" (Frege 1881/1979, 3). Boghossian correctly notes, however, that Frege's definition includes a type of success grammar, in that it suggests that the premises from which an inference follows need to be true. Instead, Boghossian contends that the thinker must merely presume that these premises are true and take their judgment to be supported by them. This provision importantly allows for the possibility of bad reasoning. Boghossian calls this his "taking condition" on inferences:

> *Taking:* "Inferring necessarily involves the thinker *taking* his premises to support his conclusion and drawing his conclusion *because* of that fact" (4).

Boghossian holds, in other words, that inference is a cognitive transition in which a subject tries to determine what might follow from what she already believes to be true.

But how widespread is the phenomenon that Boghossian's taking condition accommodates? In formal academic settings, it is common for subjects to reason by explicitly considering existing beliefs, and then deciding what conclusion it is appropriate to draw from those beliefs. I will refer to this type of inference as paradigmatic inference. A chemist might infer from a pattern of lab results that some new scientific law is true; that is, she takes the generalized law to be supported inductively by her lab results. A mathematician might infer that a certain theorem is true from a formal proof; she takes the theorem's truth to be deductively guaranteed by the steps in the proof. Note that neither of these examples depends on the conclusion-belief actually being true. Either inference might be bad. The chemist's lab assistant might have systematically mishandled some aspect of the experiment and the mathematician may have misapplied a rule in her

deductive proof. In neither case does the mistake deprive the thinker's transition of its status as an inference, but merely renders it a bad inference. All that matters for the relevant transition to qualify as an inference is that the subject *takes* the conclusion-belief to follow from her premise-beliefs.

I believe, however, that there are many cases of inference in which the premise-beliefs and the rule on which the processing is based are not just implicit, but inaccessible to the subject. Given the existence of cases along these lines, like cases where an inference is based on a probabilistic Bayesian algorithm, it does not seem apt to say that a subject has "taken" the premises or rule to be any way at all. A full accounting of the variety of good and bad inferences encourages an account on which a subject takes her conclusion to be a certain way.

I believe our third condition for a cognitive transition to count as an inference should be a version of the taking condition broader than the one Boghossian offers. I will defend the following condition, which I will call Taking*.

> *Taking*:* A cognitive transition resulting in a conclusion-belief qualifies as an inference if the subject takes the conclusion-belief to follow from a trustworthy internal process.

I will say much more about Taking*, including what it means for a subject to take a conclusion-belief to follow from a trustworthy internal process. For now, I want to note that unlike Boghossian's taking condition as described above, Taking* shifts the subject's focus; rather than taking the premise-beliefs she is considering to support the conclusion-belief as on Boghossian's view, no consideration of premise-beliefs is implied by

Taking*. Instead, the subject must merely consider the conclusion-belief— the newly formed belief— and take it to follow from a trustworthy internal process. I will first argue for this reorientation, and then I will argue that taking a belief to follow from a trustworthy internal process involves the presence of a certain phenomenological datum— the feeling of correctness— which I call the assent affect.

So far, Taking* has been offered mostly as a matter of stipulation, without a full defense (though some may find, as I do, that it has significant intuitive appeal). In Section 2.2, I will argue that subjects are often blind to the premises leading to their inference. I will also argue, relatedly but distinctly, that subjects are often blind to the rules linking their premise-beliefs with their apparently inferred belief. I will conclude that a transition's status as an inference cannot require the subject to take the preceding premises and rules of an apparently inferred belief to be any particular way; in other words, that Taking* is preferable to Taking.

2.2 Ignorance of the premises and rules of an inferred belief

I argued in the previous section that an inference must meet these two conditions:

(a) the transition results in a new, revised, or reaffirmed belief

(b) this belief is caused by some prior states or events

I also argued that merely meeting these conditions is insufficient for a cognitive transition to count as an inference. Boghossian offers Taking as a third condition:

> *Taking:* "Inferring necessarily involves the thinker *taking* his premises to support his conclusion and drawing his conclusion *because* of that fact" (4).

In this section, I will deny Taking by arguing that there are cases in which subjects are ignorant of or blind to the premises on which an inferred belief is based. In fact, I will cast my argument even more broadly: I will argue that there are cases in which a subject is not only to the premise-beliefs, but also to the rule linking the premise-beliefs to the inferred belief. As I will demonstrate, it is often very difficult to clearly differentiate between what counts as a premise and what counts as a rule. Because of this ambiguity, I will consider both premise-blindness, which straightforwardly contradicts Taking, and rule-blindness, the presence of which encourages my broader claim that the relevant sense of taking cannot involve orientation towards an inferred belief's premises and rules.

It is important to note here that some of my disagreement with Boghossian emerges from the fact that he is interested in a more limited set of inferences than I am. He would likely deny that some of the inferences I consider are within the purview of his account. Some of the cases I consider, however, like fallacies of weak induction, do seem to me as if they should count in even Boghossian's more restricted category. So, while I do think my present remarks demonstrate some shortcomings of Boghossian's view, my present aim is not to parse the precise boundaries of the category of inference as he draws them. Instead, I will consider only the broad category of inference I have defined as my target, and I will argue that Boghossian's account is insufficient to cover it.

First, I would like to lay out what I mean by premise/rule ambiguity more directly. When formalizing certain types of arguments, the rule or rules in virtue of which the premises support the conclusion are sometimes included in the sets of premises. This is usually not the case in formal deductive arguments:

1. $A \rightarrow B$

2. A

---

3. B via Modus Ponens (1, 2)

In deductive arguments like this one, it would be strange to include the "modus ponens" rule as a premise. It is no problem on a view like Boghossian's for a subject to be ignorant of modus ponens even if she can still make this inference from these premises. As long as she takes these premises to support the conclusion, she can satisfy Taking.

The case is somewhat different when it comes to inductive arguments. Here, it is more natural to include a rule-like premise. Consider the following popular broad-strokes reconstruction of an argument central to Hume's problem of induction:

1. The sun has risen every day.
2. The future is likely to resemble the past.

---

3. The sun will rise tomorrow.

Here, 2 is included as a premise, when it could also be construed as an inductive rule in virtue of which premise 1 supports the inferred conclusion 3. If 2 is included as a premise, then Boghossian's account would require that the subject takes 1 and 2 to support 3, and draws her conclusion 3 because of that fact. Consider the psychology of someone inferring that the sun will rise tomorrow. It seems far more natural to say that she takes this conclusion to be well founded in a trustworthy internal process— i.e., that Taking* is met— than it does to say that all such subjects take premise 2 to be true, even in a minimal sense of Taking that does not require conscious reflection on that premise.

This intuitive appeal of Taking* relative to Taking is bolstered by considering examples of fallacies of weak induction. Fallacious inferences still ought to meet the sufficient conditions for an inference. For example, consider a subject, Swimmer, who takes his son to the beach on Cape Cod, where he overhears a story about a shark attack in South Africa. His immediate response might be, "Perhaps it isn't smart to go swimming in the ocean this summer." This appears to be an inference, in that he has endorsed a conclusion-belief that he takes to follow from certain premise-beliefs. We might formalize his inference:

1. There was recently a shark attack in South Africa.

---

2. It is not smart to swim in the ocean this summer.

Note that in this case I have avoided including a rule-like premise, as in the Humean example above.

It is unintuitive to say that Swimmer has taken the rule his inference relied upon— a weak analogy or hasty generalization rule of along the lines of "shark attacks in one place indicate that shark attacks anywhere are likely"— to be any way at all. Swimmer is rule-blind, in that he probably would not have endorsed this conclusion if he considered the rule: he would not believe that it is not smart to swim in the ocean this summer.

Of course, rule-blindness is compatible with Taking: someone could take a premise to support a conclusion without taking any particular rule to link them. Taking requires only that a subject takes certain premises to support the inferred conclusion. While Swimmer may not be unaware of this premise, I find it uncomfortable to say that

Swimmer endorses his conclusion because he takes it to follow from this premise and this premise alone. If we were to ask Swimmer why he believes it is unsafe to swim, would he list the single premise above and stop there? It seems far more likely that subjects would, when asked to explain the premises that led them to a (bad) inferred conclusion, attempt to rationalize their conclusion post hoc with a stronger, more complete set of premises than the veridical set of premises on which their original inference was based. That is, as soon as a subject's attention is directed to the causal history of a poorly informed belief, they do not take their original, veridical premises to be sufficient for inferring the conclusion.

The most natural thing to say about subjects who make bad inferences is that they do so because they are not regarding the causal history— both premises and rules— of their conclusion-belief very carefully.

In order to make this point even more clear, consider the example of safe consensual sex between siblings, popularized by Jonathan Haidt (2001). In this test, subjects are presented with a brief story:

> Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love? (Haidt 814).

Haidt notes that a large majority of subjects respond that it is not acceptable for these siblings to have had sex, and then subsequently begin searching for reasons. Haidt notes that they often cite concerns about inbreeding and emotional damage, even though the

story explicitly rules out these problems. Eventually, subjects often say something along the lines of, “I can’t explain it; it’s just wrong” (ibid).

Subjects in this case are making an inference: they are reaching a conclusion-belief, namely their judgment that this sex act is wrong. An account that denies that this type of case is an inference must also exclude much other everyday reasoning, good or bad, that is structured similarly. The subjects in this case seem unable, however, to articulate the premises in virtue of which they inferred this conclusion. Their suggestions about inbreeding and emotional damage could not be the veridical premises that led to the inference, because dismissing these premises does not lessen their inclination to accept the conclusion. Taking is not able to cover this type of case, as it claims that an inference necessarily involves a subject taking premises to support a conclusion, and endorsing the conclusion because of that fact. In this case, it seems that the subjects are not endorsing the conclusion in virtue of their taking any premises to be any way whatsoever. Taking* seems much better here; subjects are endorsing this conclusion because they are taking the conclusion to be the output of a trustworthy internal process, even though they are unable to articulate that trustworthy internal process at all. If a subject is not aware of the premises that led to an inference, it is difficult to say that she took these premises to be any particular way.

There are other types of reasoning in which subjects are blind to the veridical premises in virtue of which they judge a certain conclusion to be true. Consider the well-known Wason selection task. In the Wason selection test, subjects fail at a deductive reasoning task at rates as high as 96% (Oaksford & Chater). In one example of this task, subjects are shown four cards on a table, each of which they are told has a color on one

side and a number on the other: solid red, solid brown, the number 3, and the number 8. Subjects are then asked which card or cards they must turn over to test the truth of the proposition, “If a card shows an even number on one side, then it must be red on the opposite side.” This proposition can only be invalidated by a card that has an even number on one side but is not red on the other, so the 8 and the brown card should be turned. Most subjects respond that the 8 and the red card should be turned.

The interpretation of the results in these cases is controversial. Evans hypothesizes that this common false response is produced by an associative matching bias, an example of what is often called a Type 1 process: “red” and “even” are named explicitly in the prompt. Oaksford and Chater argue that the results might be better interpreted as adhering to probabilistic processes. Oaksford and Chater argue that subjects’ standard, incorrect responses to deductive prompts like the Wason selection task might actually be a byproduct of generally reliable everyday reasoning processes, though they argue that reasoning in these cases does not proceed according to learned deductive rules about the material conditional operator, as standard interpretations of the Wason results hold. In everyday reasoning, Oaksford and Chater argue, subjects assess not the strict logical rules of conditionals, but rather whether there is a dependency relation that may admit exceptions. For example, subjects may endorse the conditional “If something is a bird, then it flies,” and they may understand that accepting “Tweety is a bird” means that they must also accept “Tweety flies.” But, if it turns out that Tweety is a flightless bird, this means that the conditional they accepted was defeasible in this everyday rather than formal context (Oaksford and Chater 349). Subjects interpret this prompt as a

dependency relation rather than a strict logical conditional: it is a type of rule that may bear exceptions.

In the version of the Wason task discussed above, subjects are asked to evaluate the conditional, "If a card shows an even number on one side, then it must be red on the opposite side." Let us call "a card shows an even number on one side" p, and "it must be red on the opposite side" q. By Oaksford and Chater's hypothesis, subjects begin the task assuming that there is an equal (.5) probability that p and q are dependent or independent of each other; subjects are fully uncertain about whether p and q are dependent or independent. They then select the cards that will provide the greatest reduction in this uncertainty, according to Bayes' theorem. If following Bayes' theorem correctly, and if subjects make what is known as the "rarity assumption"— namely that the probabilities of both p and q are low— then turning over the card with the even number and the red card will in fact be a rational strategy (Oaksford and Chater 351).

There is significant debate in the literature over whether Oaksford and Chater's "expected information gain" account is the best explanation of the Wason task. Evans and Over (1996), for example, have argued that Oaksford and Chater's model is psychologically implausible, and furthermore that the "rarity assumption" would only be rational in certain formulations of the Wason task, such as when the apparent dependence is between colors and certain letters (rather than odd vs. even numbers, where the probability is even).

It is beyond my ambition here to resolve the debate about the type of processing on which standard responses to the Wason task rely. Regardless of whether subjects reach

their conclusions via associative or Bayesian processing, it seems clear that they are thoroughly blind to the details of the causal history of that conclusion.

One may be tempted to deny that common responses to the Wason task are inferential. Such an objection would say that only the subjects who reasoned through the problem carefully have made inferences. My goal in this paper is to accommodate all types of reasoning into my account; I want to give an account of the category to which both the "Type 1" associative/Bayesian process and the "Type 2" effortful process belong. I believe that there is quite a lot in common in terms of the phenomenology of any response to the Wason task; both the correct and incorrect subjects take themselves to have reasoned through the problem correctly.

Either way the results of the Wason task are interpreted, the subjects seem to be blind to the details of the causal history of their conclusion. Taking, though focused on a narrower range of cognitive transitions that I am at present, requires that subjects take the premises of their inference to support its conclusion. This seems incompatible with cases like this; it seems subjects in the Wason task are premise-blind.

The proponent of Taking, or a similar causal-history account targeting a broader category of inference, could argue that much of what is contestable in the debate between the associative and the Bayesian explanations is about processing rather than premises. That is, she could argue that subjects are blind to the nature of the rule-following that their inferences follow rather than blind to the premises themselves. But this response will only go so far. If the Bayesian explanation is the veridical explanation of poor performances on the Wason task, then subjects rely upon the rarity assumption, which is critical for the argument to go through, and which is difficult to construe as a rule rather

than a premise. It is very unnatural to say that a subject endorses her conclusion *because* she takes the rarity assumption to be true. There are likely to be analogous implicit premise-beliefs in other cases of probabilistic inference.

Regardless of the correct explanation of the reasoning that leads to the incorrect inference on the Wason task, the most natural thing to say is that the subjects seem to have not thought very hard about the conclusion they reached. Daniel Kahneman discusses the phenomenon of "cognitive ease" at length: subjects prefer to answer prompts in a way that will prevent them from having to engage in the type of mental strain that might be necessary to reason through a task like Wason (2011, 62). In the final section of this chapter, I will discuss cognitive ease in greater detail, and it will figure centrally into my claim that an inference occurs when a subject takes a conclusion-belief to follow from a trustworthy internal process, as signaled by the feeling of correctness, or the assent affect.

In this subsection, I considered cases where subjects seem blind to the details of the causal history of an inferred belief, and more specifically blind to the premises on the basis of which a new inferred belief is endorsed. The examples of everyday (especially weak) induction, Haidt's incestuous siblings, and the Wason selection task make clear that a causal history construal like Boghossian's Taking cannot be extended to cover the unified category of inference I am considering. If a subject is not aware of the premises that led to a conclusion, it is not clear how she could have endorsed that conclusion *because* she takes those premises to support it.

Once we dismiss the causal history construal of Taking, we are left with two options. One option, which I have already argued is a natural and intuitive interpretation

of the data, is to accept Taking*, on which an inference occurs when a subject takes her conclusion to follow from a trustworthy internal process without regard for the details of that process. In the following section, I will consider and reject a second option: that an account of inference can be delivered without any construal of taking.

3. The rule-following option

3.1 The bare inferential transition account

So far, I have demonstrated that inference does not involve a subject's taking the premises or rules on which an inferred belief is based to be any particular way. I have, along the way, presented arguments that Taking*, which does not require a subject's regard for the details of an inferred belief's causal history, is a preferable condition for inference over Taking. Before I conclude that Taking* is the correct requirement on inference, however, I must first consider the possibility that inference does not involve taking at all. On this type of view, some other factor must be responsible for elevating certain cognitive transitions to the status of inference. The most prominent views along these lines have held that a cognitive transition counts as an inference in virtue of its proceeding via a particular type of rule. I will discuss the recent account offered by Jake Quilty-Dunn and Eric Mandelbaum in particular.

Before discussing their account in particular, I want to briefly discuss in general the role that rules and rule-following play in inference. In the paradigm reasoning cases of the chemist and the mathematician, the subjects make a cognitive transition from certain premises to a conclusion they take to follow from those premises. For John Broome (2012) the appeal of a rule-following view of inference emerges from

epistemological considerations: the rule is what justifies the inference, and following a rule explains how the inference was made. Boghossian (2014) offers a speculative rule-following view based on similar epistemological considerations, but he acknowledges considerable difficulty in developing a view along these lines, and explicitly declines to offer more than a speculative rule-based construal of Taking (17).

There is no shortage of discussion of rule-following in the post-Wittgenstein literature, and it is well known that profound philosophical challenges arise from the consideration of the psychological role that rules play. What type of facts are rules such that they can exist independently and objectively from any particular application? Given their generality, how can rules be sufficiently specific to any instance— how do we know that we are not applying Kripke's *quus* rule when we mean to perform addition? How do we grasp these objective yet relevant rules and wield them so readily in our psychological practice? These are important questions about which I wish to remain largely silent. My present concerns are neither normative nor epistemological— I will not address the curious property of premises by which they might confer justification upon an inferred conclusion. I am building a descriptive account: what makes certain cognitive transitions unique in that they qualify as inferences? Jake Quilty-Dunn and Eric Mandelbaum (2015) have recently advanced a descriptive, naturalistic account of inference, on which a rule-following requirement for inference can render irrelevant a taking condition like Boghossian's or mine.

Quilty-Dunn and Mandelbaum (QDM, henceforth) build an account on which inferences are unconscious transitions that are necessarily based on rules built into our cognitive architecture. They call the minimal unit of inference a basic inferential

transition, or BIT. They claim that a cognitive transition qualifies as a BIT iff (i) the states involved are discursive rather than iconic, (ii) the relationship between the states is described by a logical rule built into our cognitive architecture, and (iii) the inference is processed based on this rule rather than some other factor (11-12). The notion of "built into the architecture" does quite a lot of work here. QDM define that notion as:

> "A rule is built into the architecture of a representational system iff the system is constructed in such a way that, if a mental representation is tokened that satisfies the antecedent of the rule, then, *ceteris paribus*, the system will generate a token representation that satisfies the consequent of that rule." (10)

On QDM's view, a subject infers from "If it is raining, I should bring an umbrella" and "it is raining," to "I should bring an umbrella" because the logical structure of her discursive representations of the premises tokens a modus ponens rule that is built into her cognitive architecture. Similarly, she can make a conjuctive inference along the lines of "the dog is big" and "the dog is smelly" to "the dog is big and smelly" because a rule like *If X is A and X is B, then X is AB* is built into her architecture. The subject's mental representations of the premises token the rule's antecedent, thus unconsciously triggering endorsement of the rule's consequent.

QDM distinguish BITs from rich inferential transitions, or RITs, in that in the latter type of inference, the subject is disposed to endorse the cognitive transition itself (21). That is, when reasoning with RITs, the subject explicitly takes her premise to support her conclusion— she meets Taking in Boghossian's sense. More specifically, QDM hold that a subject explicitly takes her premise A to support her conclusion B when she is disposed to form a thought along the lines of "A <u>therefore</u> B," where "<u>therefore</u>" is a concept requiring some appreciation of logic (*ibid*). In the case of a mere BIT rather than a RIT, a subject can only be said to meet Taking in a minimal, implicit sense: she

takes her premise to support her conclusion insofar as the former caused the latter in virtue of a built-in rule. For QDM, "RITs are distinguished from BITs not in how the conclusion thought is produced, but simply in that the thinker is additionally disposed to endorse the inference" (22). That is, there both BITs and RITs are processed via the same built-in-rule, but only in the case of an RIT does the subject explicitly take the conclusion to be supported by the premises.

QDM argue that Boghossian's Taking condition serves two purposes: (a) to offer a satisfactory causal explanation of the transition from a premise-belief to a conclusion-belief (in my terminology) and (b) to explain the transfer of justification from premise to conclusion (22). Like QDM, I am not concerned with (b). With respect to (a), QDM believe that their account is a satisfactory causal explanation of the transitions involved in inferences—both BITs and RITs— and that as a result, a taking condition is unnecessary. We infer B from A *because* an unconscious logical rule built into our cognitive architecture is triggered by the token A.

QDM's account seems to face some difficulty when it comes to the individuation of inferences. This is especially vivid when we consider inferences that intuitively rely upon rules that are not plausibly built into our cognitive architecture. Questions like this are common on standardized tests:

If x ♣ y means 2x+5y, what is 3 ♣ 8?

In solving this question, we go from a mental representation involving "3 ♣ 8" to a mental representation that "3 ♣ 8 = 46" is the right answer. Intuitively, this is an inference based on the ♣ rule. Of course, the ♣ rule is not built into our cognitive architecture. QDM would have to argue that we solve this question via a chain of BITs,

each requiring a suitably abstract conditional rule so as to accommodate new rules. Some of the BITs involved might require a rule like *If a new operator is introduced, replace the symbol with more familiar operators in the new operator's definition.*

If you asked a test-taker, Cindy, how many inferences she made when solving this problem, she would say "one." We might call the one inference she means the "macro-inference"; in the macro-inference, she inferred "3 ♣ 8 = 46" from the prompt and the definition of ♣ . On QDM's view, she has made several BITs in order to get to this inference, but the macro-inference does not count as a BIT.

Is it an RIT? On QDM's view, a RIT differs from a BIT only in the subject's disposition to endorse the transition. The macro-inference in this case differs from its constituent BITs in more than the presence of this disposition: it is not a single BIT at all, since the ♣ rule is not built into her cognitive architecture, but rather an agglomeration of BITs. If the macro-inference is not a BIT or an RIT, it seems that Cindy must be incorrect in calling it an inference.

I do not object to the idea that "macro-inferences" like the one Cindy endorses in this case rely on significant unconscious processing, nor do I object to the idea that this unconscious, underlying processing involves logical rules built into her cognitive architecture. I furthermore grant that some of this unconscious processing may involve inferences. I do think it is an uncomfortable consequence for QDM, however, that Cindy would be *incorrect* on their view to say that she inferred "3 ♣ 8 = 46" (the conclusion) from the prompt (the premise) and the definition of ♣ (the rule).

Note also that this problem applies more broadly than cases of novel rule definitions. An analogous problem would arise in any case of an inference based on a rule

that combined multiple built-in rules. It is unlikely, for example, that DeMorgan's Law is built into our architecture. When I transition to the representation that (~P ∧~Q) from the representation that ~(P∨Q), this probably relies on many underlying BITs. And yet the only inference I am intuitively inclined to describe as such in this case does not rely on a built-in rule, and is thus not as easily accommodated under QDM's view as BITs, which seem to belong to a category I am less inclined to call inference.

Furthermore, it is not obvious to me that I am incapable of making a DeMorgan's inference that veridically relied upon the non-built-in rule. If I am presented with a step in a proof where DeMorgan's Law is applicable, it seems I might be able to entertain the premise-belief in my working memory and solve it via cognitive brute force in a way that operates on DeMorgan's Law. While there surely would be underlying processes at play that would rely upon built-in logical rules, there seems to be a regard in which I can make an inference whose processing was veridically based on DeMorgan's Law, and not merely on built-in constituent rules. Something similar could be said about Cindy's capacity to make a brute-force inference with the ♣ rule.

It may be the case that the QDM account does not aspire to explain transitions like these. They acknowledge at the start of their paper that the psychological approach to which they subscribe to focuses attention on the "involuntary, unconscious, and perhaps normatively degenerate aspects of inferential transitions" (2). I readily agree that unconscious rational processes are important for inference, and that subjects may often be mistaken about the details of these processes. I argued at length in the previous section that subjects are often blind to the details of the causal history of an inferred conclusion. It strikes me as problematic, however, that the QDM account of inference seems to

neglect some the transitions that seem to fit most intuitively into the category in favor of logical transitions that are more controversially inferential.

Based on this discussion, some further virtues of a taking condition in general and of Taking* in particular begin to emerge. An account of inference that incorporates some form of a taking condition provides for a natural explanation for the individuation of inferences. An inference achieves its status as such because (in part) of our taking it to be one. Any causal transition we take to be an inference is an inference. This also makes clear the importance of considering a broad, commonsense category of inference.

Additionally, the example of Cindy highlighted again the fact that subjects are often blind to the premises and rules preceding our inferences. She took her belief that "3 ♣ 8 = 46" to follow from a trustworthy internal process. The veridical nature of the processing that got her to that conclusion, however, is a complex issue. Taking* explains how Cindy could be able to make this macro-inference without any reflection on how she got there. This gives us reason to prefer Taking* over Boghossian's Taking.

In Section 2 of this chapter, I argued that despite certain virtues of a taking account of inference, this taking could not plausibly involve regard for the details of the causal history of an inferred belief. In Section 3, I have discussed Quilty-Dunn and Mandelbaum's recent account of inference, on which there is no taking condition on inference, but rather an inference achieves its status as such based on a particular type of rule-following. I argued that, like Boghossian's Taking, the bare inferential transition account of inference is worse off than Taking* in explaining the target category of cognitive transitions.

4. Cataloguing the types of inferences

I have fixed my explanandum in this paper as commonsense inferences, very broadly construed; any cognitive transition resulting in a belief where a subject takes this new, revised, confirmed, or supposed belief to follow from a trustworthy internal process. Along the way, I have referred to several different types of inference that fall under the umbrella of this unified category. I have not yet laid out a full accounting of the transitions I intend to include. In the catalogue that follows, I will make clear that Taking* is well positioned to cover this full range.

It is also important to note that the types of rule-following I discuss in this section may in fact not turn out to be discrete kinds. Some of these are likely to be combinable, some may be alternative explanations of the processing in a single type of rule, and some may be divisible into two or more subcategories. The categories of rules offered in this section are merely one way of demonstrating the diversity of cognitive transitions that must be accommodated by a full account of inference.[2]

Deductive inferences

First, let us consider reasoning that appears to be deductive. I say "appears to be," because some may take "deductive reasoning" to necessarily involve the application of logical rules. That is, "deductive reasoning" might, on the one hand, refer to reasoning whose actual processing consists of the application of rules to infer certain logically guaranteed results. A broader construal of deductive reasoning, on the other hand, would

[2] Consideration of the full diversity of types of processing that underlie inferences will also be essential to the second chapter of this dissertation, where I address whether a dual process model of reasoning is well-equipped to handle this diversity.

apply to any inference— regardless of the nature of the actual processing underlying it— that is appropriately evaluated according to a deductive standard. For example, consider the prompt, "Given that A, and that If A then B, what follows?" On the second, broader construal just mentioned, the reasoning required to infer an answer to this prompt should count as deductive regardless of the nature of the processing that actually underlies the subject's inference. It is in this second construal of deductive reasoning that I am interested. This is an important distinction, as subjects do not reason through all deductive prompts via reliance upon deductive rules. For each of the types of deductive reasoning I will discuss, it is important to remember that "reasoning" ought not to have a success grammar built in; each of these categories admits both good and bad reasoning.

Explicit reliance on deductive rules in processing ("paradigmatic deductive inferences")

Earlier in this chapter, I referred to the mathematician as a "paradigmatic" case of deductive inference. The mathematician engages in explicit deductive inference in each step of her proof: she considers what she has proven so far, considers the rules of mathematical inference she has learned, and endorses a new fact that is deductively supported by the combination of the previous steps and the application of the rule. Students taking an exam in an introductory formal logic class may also rely explicitly on deductive rules in each step of a proof. They also, however, might misremember a rule: say that they confuse the deductive rules they have learned and execute one step in their proof based on their mistaken belief that "affirming the consequent" is a valid rule of deductive inference. While they are explicitly relying on a deductive rule, they are clearly making a bad inference.

Explicit reliance on deductive rules meets both (a) and (b), and in fact it does so paradigmatically: a subject considers certain premise-beliefs and forms a conclusion-belief (a) that is caused by her deliberation on her premise-beliefs (b). The subject explicitly relying on deductive rules meets Taking* in that she takes her conclusion to follow from a trustworthy internal process: namely the explicit consideration of certain premise-beliefs and certain deductive rules. She also meets the higher standard of Boghossian's Taking, in that she takes her premise-beliefs to support her new conclusion-belief.

One important note about explicit reliance on deductive rules is that it is quite costly in terms of the burden it places on our working memory resources. It is difficult for us to sustain this type of paradigmatically deductive reasoning for long periods of time. This fact is central to Kahneman's endorsement of a "default-interventionist" dual process model, on which subjects prefer to remain in a default state of "cognitive ease," only firing up more demanding reasoning processes when the defaults are in some way inadequate (Kahneman 2011).

I so far have intended to make no claims about the commonality or rarity of explicit reliance on deductive rules relative to other types of inference, deductive or otherwise. I have simply argued that subjects do on some occasions engage in reasoning that proceeds by relying on explicit deductive rules.

Implicit reliance on deductive rules in processing

Subjects may sometimes appeal to deductive rules without explicitly considering them. Presented with the prompt, "Given that A, and that If A then B, what follows?" it

seems that I (or anyone with the requisite amount of experience with modus ponens) is able to answer the prompt without consciously reflecting on the rule itself. This is perhaps even more evident in the case of basic arithmetic. If asked to add 83 and 38, a subject may respond quickly and correctly. The processing underlying this inference might involve the following of learned deductive rules, but the response can be given without conscious reflection on these rules. One can also rely implicitly on deductive rules while making an incorrect inference. If a subject responds to the above addition problem as 111 rather than 121, the most natural explanation seems to be that they have applied a rule incorrectly— perhaps not carefully enough— rather than that they were relying on a means of processing other than rule-following.

Implicit reliance on deductive rules clearly meets (a) and (b), in that a conclusion-belief is created as the result of a causal process. Implicit reliance and explicit reliance on deductive rules should not differ in this respect. Implicit reliance on deductive rules meets Taking* in that the subject takes the new conclusion-belief— say, the answer to an arithmetic problem— to follow from a trustworthy internal process. We can also grant Boghossian that his Taking condition to accommodate cases of implicit reasoning, though some potential difficulties may arise here, as discussed in section 3.

These comments should be relatively uncontroversial; I have not committed myself to any claim or generalization about the frequency with which subjects make inferences based on implicit application of deductive rules. Nor have I committed myself to the existence any categories of reasoning that are *always* addressed via this type of inference. I merely wish to suggest that this is one way in which we do sometimes make inferences. Instances of implicit reliance on deductive rules ought to count as inferences.

Associative deductive reasoning

In general, a cognitive process is said to be associative when it operates based on similarity, correlation, and statistical regularity between features of the environment in a subjects' experience. Associations are neither propositionally structured nor sensitive to logical intervention. Associative processes are contrasted with logical or rule-based processes, whose operation involves the manipulation of abstract logical rules or symbols that are propositionally structured. One very important distinction for present purposes between associative and rule-based inferences is that the former are relatively immune to self-revision in any particular reasoning task, while the latter are more easily self-revised. Because associations are based on statistical regularities, they are relatively rigid; they can only be modulated over the long term via counter-conditioning.

The prevalence of associative reasoning has recently become a matter of controversy in the philosophy and psychology literature on reasoning. Some advocates of dual process models, most notably Steven Sloman, contend that "Type 1" processes are essentially and necessarily associative (1996, 2002, elsewhere). They point to examples like performance on the Wason selection task in order to bolster their claims that any reasoning process that is not essentially rule-based is associative. Eric Mandelbaum, on the other hand, has argued that many ostensibly associative inferences are in fact responsive to logical intervention, and are therefore not associative (2014). This argument will be taken up at greater length in Chapter 2.

In Section 2, I discussed the associative interpretation of common incorrect responses to the Wason selection task. There is no need to revisit that discussion now. I

merely wish to note that an associative deductive inference does in fact count as an inference, broadly construed, and that the nature of the rules that figure into this type of inference are categorically different than the types of rules that figure into inferences that rely on learned deductive rules, for example.

Probabilistic deductive reasoning

The possibility of probabilistic deductive reasoning was also discussed at great length in Section 2, in my discussion of the Oaksford and Chater interpretation of the common incorrect responses to the Wason selection task. Once again, we must only note here that such reasoning does in fact count as inference, and that the type of rule on which such an inference would rely is different from either associative reasoning or reliance on learned deductive rules.

Inductive reasoning

Explicit inductive reasoning

I will now turn my attention to inductive reasoning. Once again, I must make clear that I am interested in reasoning that appears to be appropriately held to an inductive standard. In Section 2, I described the case of a chemist who follows specific learned, inductive guidelines to generalize a new law based on repeated lab results. This type of inference is clearly inductive, and it is paradigmatic in the sense that the subject explicitly calls upon inductive rules for the generalization of chemical laws that she has learned. Something along the lines of “only after an experiment has been repeated x

number of times with y degree of similarity can a result be generalized into a scientific law."

Just as with explicit deductive reasoning, it is likely that explicit inductive reasoning is largely restricted to academic contexts. There are of course exceptions: subjects may use learned deductive rules explicitly if calculating a percentage in a non-academic setting using basic algebra, or a subject might apply a learned inductive prediction guideline when playing fantasy football. Regardless of how common or rare it is for subjects to apply inductive rules explicitly, it seems that they do make inferences along these lines on occasion. These inferences meet (a) and (b) in that a new belief proceeds from a causal process. These inferences also meet both Taking and Taking*, for similar reasons described in the section on explicit deductive reasoning.

Everyday inductive reasoning

I discussed some examples of everyday induction in section 3. Swimmer was such a case, and any number of examples of reasoning based on fallacies of weak induction would serve equally well. In any instance of everyday induction, a subject relies on an inductive rule— generalization, analogy, etc.— but these generally will not be rules they are able to articulate, as the fallacious cases make clear. As I have said several times, this rule-blindness may not be a problem for Boghossian, given a charitable reading of Taking, but it does seem to encourage the alternative view that a subject endorses an inferred conclusion because she takes that conclusion-belief to be a certain way, and not the causal history of that conclusion-belief. Everyday induction may rely on associations, it may rely on probabilistic reasoning, or it may rely on processing different from either

of these. Everyday inductive inferences are inferences, but they are categorically different from inferences relying on rules that were explicitly learned.

Hypothetical reasoning

As I mentioned at the start of Section 2, there are certain exceptional cases in which inferences do not result in beliefs. When a subject reasons via supposition, she entertains certain premises, which she also might not believe, and considers what further conclusion she must also believe if she *were* to believe those premises. This type of hypothetical reasoning could be deductive or inductive, but it will necessarily be explicit. I view it as a subcategory of paradigmatic inference like the types that Boghossian considers— it is demanding on our working memory. Unlike many types of inference, it seems more likely that hypothetical inference requires some logical sophistication. A subject likely must have at least some concepts about arguments and logic in order to make a hypothetical inference. In Section 5, I will argue that these exceptional cases are naturally included under my picture of inference, as they yield the assent affect despite the fact that the subject will not believe the conclusion.

5. Taking*

On my account, a subject makes an inference if and only if she meets these 3 criteria:

(a) the transition results in a new, revised, or reaffirmed belief

(b) this belief is caused by some prior states or events

(c) *Taking*:* A cognitive transition resulting in a conclusion-belief qualifies as an inference if the subject takes the conclusion-belief to follow from a trustworthy internal process.

In this section, I will flesh out Taking* in much greater detail. I will explain what it is for a subject to take a conclusion belief to follow from (or be the output of) a trustworthy internal process. I will argue that the relevant form of Taking* is intuitional—it involves a metacognitive intuition, and that this intuition carries the feeling of correctness, or the assent affect.

Boghossian considers a few possibilities for how his Taking condition might be construed. Although his Taking is different than my Taking*, in that his version requires a subject to take her premises to support her conclusion, the construals he considers are still instructive in developing how we should understand Taking* and its intentional contents.

The first possibility that Boghossian considers is that the Taking* condition might involve a meta-cognitive *belief* about the relationship between premise-beliefs and conclusion-belief (6). On this first such construal he considers, the "full-fledged normative doxastic construal," the Taking condition is met when a subject has a belief along the lines "My judging (1) and (2) supports my judging (3) (ibid). Boghossian rejects this construal for two reasons: first, it is the presumed truth of 1 that motivates the subjects metacognitive judgment (Taking), not the fact that she judges (1) to be true. Second, and more germane to my concerns in this chapter, Boghossian worries that a normative metacognitive belief along these sides is not plausible for most ordinary

thinkers. It is implausible that children, or even many non-philosopher adults, have metacognitive beliefs about the relations between their premises and conclusions (7).

Boghossian moves on to consider a different doxastic construal of Taking, on which the subject merely forms a "meta-propositional" belief, e.g. the belief that (3) follows from (1) and (2) (ibid). This type of metacognitive belief would avoid the first problem that the normative meta-belief discussed above faced, in that a subject is not required to judge a conclusion true in virtue of judgments (rather than truth). The meta-propositional belief will, however, succumb to the second concern above, i.e. that it requires subjects to have more sophisticated concepts of rationality than seem plausible in all human thinkers.

After rejecting these doxastic construals of taking, and others whose consideration is less relevant to our present consideration of Taking*, Boghossian considers an intuitional construal (8). I believe that an intuitional construal is the best explanation of Taking*, with intuition understood as "intellectual seeming." While Boghossian is agnostic about whether it carries a distinctive phenomenology, I will argue that this intuitional state that underlies Taking* is accompanied by the assent affect— the feeling of correctness. Boghossian argues that an intuitional construal is better able than a doxastic construal to survive some of his concerns, in part because "an intuition, like a perception, is not subject to epistemic assessment— it is beyond justification" (9). He ultimately worries that an intuition will not be able to non-circularly *explain* how a subject is justified in performing an inference. As discussed in Section 3 of this chapter, however, my ambitions are descriptive and naturalistic, rather than normative and epistemological. Like the Quilty-Dunn & Mandelbaum account, I am interested in what

property of inferences distinguishes them from other types of cognitive transitions, and not in the conferral of justification from premise to conclusion. Boghossian's worry about an intuitional basis of a taking condition does not apply to my ambitions with Taking*.

Where *X* is the content of a conclusion-belief, I hold that Taking* involves an intuition like this:

[I intuit that] X *is true because it follows from a trustworthy internal process.*

When a subject infers *X*, she has an intuition with this intentional content. As discussed at length in the foregoing sections of this chapter, she need not regard her premises or the details of the causal history of her inferred conclusion belief at all.

In designating the attitude underlying Taking* as an intuition rather than a belief, I wish to make the requirements on thinkers as spare as I can. As the examples in the previous section of this chapter made clear, subjects very often make commonsense inferences without any awareness of the details of the inferred conclusion's causal history. If Taking* involves an intuition of this type, then these considerations about premise- and rule-blindness are fully accommodated. I also hold that this intuition carries a distinct phenomenology— the feeling of correctness, or the assent affect— which I will discuss at greater length shortly.

First, let us consider the intentional content of this intuition in a bit more detail. Note that, on my view, subjects need not have any of the sophisticated meta-cognitive concepts that worried Boghossian as he evaluated doxastic underpinnings of his Taking condition. I believe that the requisite concepts that figure in the intentional content of the

intuition—that conclusion-belief *X* follows from a trustworthy internal process— are available in a wide variety of reasoners, including children.

The intentional content of the intuition requires that subjects understand themselves to have at least one "internal process" that produces ideas. I wish to characterize this conception in the most general way, without committing myself to any particular account of self-knowledge or theory of mind. Rather, on my view, any being that is capable of self-ascribing an idea (as opposed to, say, a perception) understands that some mental events exist thanks to internal mental processes. If a subject can muster the thought, "this idea came from *me*," they have the relevant concept of an internal process. They also understand the relevant sense of "follows from"— it seems to them that the conclusion-belief (which they need not conceive of us a belief in robust terms) came about because of something they did. The conclusion is *their* conclusion, or so it seems to them, even in the absence of what the inferential effort, i.e. the details of the causal history of the conclusion-belief, consisted of.

The requirement that subjects take the belief to follow from an *internal* process does some particular work here. In contemporary epistemology, some accounts hold that perceptual experiences hold immediate justification in virtue of carrying a particular type of phenomenal force. This view is known as dogmatism, or perceptual dogmatism (Pryor, Huemer). First, I hold that there is a distinctive phenomenology associated with inference, and with the Taking* intuition. I am doubtful that the phenomenology in virtue of which perceptual experiences may be justified is indistinguishable from the assent affect. Moreover, on my view, the Taking* intuition involves the subject ascribing a conclusion-belief to an *internal* process. It is critical to a belief's status as inferred that

the subject takes it to follow from an internal process. Since subjects do not take perceptual experiences to follow from an internal process, there is no risk of perceptual experiences being construed as inferred by the dogmatist's lights.

Additionally, subjects must intuit that the internal process from which the conclusion followed is *trustworthy*. Intuiting that a process is trustworthy means that it *seems* trustworthy to them; they are tempted to accept the truth of the belief. Ernest Sosa has characterized intuition as a temptation or disposition to believe (1998). It must seem to the subject that they are disposed to accept the truth of the conclusion because of the nature of the internal process from which the conclusion followed. Very importantly, they need not have any regard for the particular premises or the rules that figured into the causal history of that process, nor even any particular idea about what facts about the internal process made it trustworthy. They merely must intuit that they assent to the truth of the conclusion because the internal process was a process that seems to produce true conclusions.

6. The assent affect

As discussed previously in his chapter, my account holds that this intuition has a distinctive phenomenology, namely the feeling of correctness, which I am calling the assent affect. I think that this phenomenology is important: subjects are likely to say that a conclusion "felt right" as they are to say that it "seemed right"— they are not inclined to make a commonsense distinction between the phenomenology and the intuition. The assent affect is an intellectual feeling that accompanies a conclusion-belief; it is the feeling that the conclusion-belief has followed from a trustworthy internal process. I view

the intuitional state and the assent affect as fully coextensive: the intellectual seeming as outlined in the previous section consists in the manifestation of the feeling of correctness that accompanies a conclusion-belief. When a conclusion-belief is accompanied by the assent affect, the subject is in the correct intuitional state.

Defining inference in terms of seeming and feeling accommodates the widespread existence of bad inferences and inferences where the subject is correct but is nonetheless ignorant of or blind to the details causal history of a conclusion-belief. All that matters, on this account, is that a conclusion-belief is accompanied by this seeming or feeling; it is in virtue of this feeling of rightness— this intuition and feeling that the belief has followed from a trustworthy internal process— that a subject endorses the conclusion-belief. In the previous section, we examined cases of fallacious inductive inferences. In these cases, subjects formed an irrational conclusion-belief based on certain premise-beliefs, and the only type of explanation they had for the endorsement of their conclusion-belief was the intuition or feeling that this conclusion-belief is a good conclusion.

In cases like those of associative or probabilistic reasoning, subjects are not aware of the reasons in virtue of which they inferred their conclusion, even if their inference is good. An account of inference on which their "taking" is meant to do any explanatory work in terms of the actual justificatory reasons for endorsing their conclusion is useless in these cases. The fact of their taking the conclusion to follow is irrelevant to the type of processing on which their inference relied. This shift is necessary to accommodate the full diversity of good and bad inferences.

I do not take myself to be committed to the conclusion that all types of inference will enjoy identical phenomenology. Especially in cases of explicit reasoning, different phenomenology might accompany deductive versus inductive inferences. I merely hold that phenomenology accompanying all inferences have something minimal in common: the assent affect, or the feeling of correctness.

We are now able to say something about the types of being to which this account of inference can be extended. If a being is capable of achieving a conclusion-belief whose propositional content is targeted by the Taking* intuitional state, which involves the phenomenology of correctness, then that being is capable of inference. It seems like that human children are fully capable of making a wide range of inferences on this view, which I take to be a virtue. The view is extendable to non-human animals to the extent that those animals are capable of manifesting the Taking* intuitional state and the assent affect. I do not wish to take any hard stance on this issue, but I argued in the previous section that any being capable of simple belief self-ascription is likely capable of manifesting the Taking* intuition, and so I believe that those species which are more successful in self-knowledge tests are likely capable of inference as well.

I have argued that the Taking* intuition involves a particular phenomenology—what it is for a conclusion belief to *seem to have followed form a trustworthy internal process* is for it to *feel* a certain way. But what if a thinker were capable of a similarly structured intuition with similar intentional content, but it were not capable of manifesting the relevant phenomenology? I am willing to bite the bullet that this being is incapable of inference. Consider Watson, the IBM question-answering computer that beat the two most successful champions in the history of the television show Jeopardy!

Watson operates on a probabilistic, Bayesian search algorithm (Yuan). Based on the words in the clue entered into Watson's database, Watson searches its vast storage of literary and reference materials for frequently associated ideas, and comes up with several possible solutions, to which Watson attaches percentile degrees of confidence. If Watson's confidence in one response meets a certain threshold, Watson will "buzz" in and offer its answer. Charitably, Watson's confidence-based judgments could be described as a type of intuition— it seems to Watson that its conclusion is based on trustworthy internal processes. But Watson does not, of course, experience the phenomenology of correctness. It strikes me as no great loss to say that Watson reaches its answer by means of conjecture, not inference, unlike the human who experiences the assent affect.

One might alternatively object that my account allows for radical misfires of the assent affect to qualify as inferences. Consider, for example, someone who suffers brain damage such that she produces the Taking* intuition and the assent affect every time she smells popcorn. In each instance of popcorn odor, she will undergo a causal cognitive transition, the output of which will be a belief, whose contents is targeted by the Taking* intuition, which is accompanied by the assent affect. I would say that this subject is afflicted in such a way that she makes an enormous number of terrible inferences. I do not think this subject is so far off from someone who has not suffered any acute brain damage, and yet still makes bad inferences. Imagine someone who infers, with regularity, the negation of the correct conclusion in every case. I am inclined to say that both of these cases are extreme versions of bad thinkers we have all encountered. The cause or

consistency of someone's bad inferences is not reason to doubt their thoughts' status as inferences.

Hypothetical reasoning also involves atypical firing of the assent affect. In cases of supposition, subjects make inferences without actually believing the conclusion. As discussed in the previous section, subjects who entertain what might follow from premises they do not actually believe can still infer from these premises. This type of supposition counts as an inference even though the subject will not indorse the hypothetical conclusion, by my lights, because the hypothetical conclusion is still accompanied by the assent affect.

The assent affect figures critically into the later chapters of this dissertation, and so I ought to flesh it out a bit more. Ronald de Sousa has written extensively on the role that emotions play in rationality. De Sousa argues that the role emotions play in reasoning should not be downplayed. He takes very seriously what Daniel Kahneman has called "the pleasure of cognitive ease" (Kahneman 2011; De Sousa 2013). Kahneman describes studies in which subjects tend to answer prompts incorrectly in order to avoid using more cognitive effort than might be necessary. For example, in one study, subjects who were exposed to the phrase "the body temperature of a chicken" were much more likely to assent to the proposition "the body temperature of a chicken is 144 degrees" or any other arbitrary number at a later stage of the experiment (Kahneman 2011, 62). The "familiarity bias" of the certain phrases was enough to cause subjects to assent to propositions that should have been obviously false. Kahneman argues that remaining in a

state of "cognitive ease"— not calling upon more effortful reasoning resources— involves a type of pleasure, in which subjects prefer to remain.

Studies such as this, De Sousa holds, demonstrate that "affective phenomena [are] essentially involved in the pursuit of epistemic aims" (de Sousa 16). He focuses on a range of states that he refers to as "epistemic feelings," and identifies four in particular:

> "*Wonder* motivates inquiry, but presupposes no specific prior belief, and need not target any existing supposition. […]*Doubt* also motivates inquiry but bears on hypotheses already entertained. […] *Certainty* bears on specific beliefs; it is, in a sense, antithetical to inquiry, in that it freezes any further quest for evidence or argument. On the other hand, it frees us for action by stamping certain facts or values as appropriate ones to be acting upon. The *feeling of rightness* seems to belong in the same general category. […]The *Feeling of Knowing* bears on specific propositions, but is unable to specify them: it is a kind of indication that it is worth the time and effort to keep trying to recall something that is in fact 'somewhere in my head.'" (de Sousa 16-17, emphasis his)

De Sousa believes that the types of cases Kahneman mentions make clear that epistemic feelings like these are required for reasoning, but he does not make any claim about how they fit into dual processes, and seems to preserve his neutrality with respect to those further questions. He claims that these "specific epistemic feelings […] have very specific roles, either in stamping a kind of seal of approval on the steps of an argument or the conclusion of an inference, or, on the contrary in spurring further inquiry" (De Sousa 17).

It is worth noting briefly here that De Sousa's conception of the relationship between doubt and cognitive ease was preceded by C.S. Peirce's similar account. In the "Fixation of Belief" (1877), Peirce wrote, "Doubt is an uneasy and dissatisfied state from which we struggle to free ourselves and pass into the state of belief; while the latter is a calm and satisfactory state which we do not wish to avoid, or to change to a belief in anything else." (Section III). This Peircian view of doubt has much in common with my

view, on which Type 2 intervention is triggered by intellectual emotions like doubt or the absence of cognitive ease.

I believe that De Sousa is on the right track with his cataloging of epistemic feelings, but he does seem to undersell the importance of the "feeling of rightness," which I am calling the assent affect. The assent affect is an epistemic feeling— an attitude which causes the subject to endorse certain beliefs. In other words, in the causal chain of inference, certain premise-beliefs trigger a certain type of processing— perhaps rule-based, associative, or probabilistic; perhaps consciously available to the subject or unavailable. The processing causes a new belief to appear. This new belief may or may not be targeted by the assent affect, the metacognitive epistemic feeling; if the new belief is targeted (or accompanied) by the assent affect, the subject endorses the belief and the cognitive transition counts as an inference.

In many cases— especially in the vast majority of everyday inferences in which the subject is not reasoning in a deliberate, effortful way with explicit regard for learned inference rules— the assent affect seems to play an essential role in allowing us to remain in a state of cognitive ease, along the lines that Kahneman describes. The assent affect seems to be the vessel through which a number of biases and heuristics affect our human reasoning; it is a mechanism of cognitive efficiency that is required for inference.

Chapter 2 – The structure and triggering conditions of Type 2 reasoning

1. Introduction

Over the past few decades, dual process models of cognition have become prominent in philosophy and psychology. Such theories contend that human reasoning may occur in one of two distinct ways: Type 1 reasoning processes are fast, associative, automatic, and effortless; Type 2 reasoning processes are slow, rule-based, deliberate, and effortful. The dual process distinction has been defended using empirical follows from a number of psychological domains, and it has enjoyed significant endorsement from philosophers. The proposed dichotomy within reasoning has given philosophers a new tool with which to navigate traditional questions, including those related to moral reasoning (Greene, Saunders) and belief ascription (Apperly & Butterfill, Goldman). Theoretical problems become easier to solve when not all reasoning must be accommodated under a single umbrella.

Unfortunately, the eagerness with which dual process models have been applied has outpaced the clarity of the distinction itself. Despite the popularity and prominence of dual process models in recent psychology and philosophy, significant and substantive questions remain about what the distinction is really claiming and whether it holds up to scrutiny. It is my goal in this chapter and the following to present a novel dual process account that accommodates psychological evidence and philosophical considerations. A preliminary step (very often neglected in the foregoing literature) to a coherent dual process account is to be very clear about the category of cognitive events over which we are quantifying when we claim that reasoning can be dichotomized. My goal in the first chapter of this thesis was to establish such parameters. I presented inference, taken to be

belief-producing cognitive transitions, as the relevant category, and I argued that an inference occurs when a causal cognitive transition follows in a belief accompanied by the assent affect, or a feeling of correctness. The assent affect will figure centrally into the dual process account I offer in this chapter.

I will argue that there are three criteria a dual process model must meet, and I will discuss two of these in the present chapter. The most natural question to ask about dual process models is about which qualitative features are fundamental to each type of processing rather than merely typical. I will hold that a theory is sufficiently explanatory only if it explains something about the unique nature of each type of processing; how does each process handle information differently? What is the nature of the reasoning performed by these two processes, such that they are not merely duplicating functionality? I will refer to this as the *fundamentality* criterion.

The final, third chapter of my thesis is devoted to the fundamentality criterion. Before I can address fundamentality, however, I must first paint a clear picture of what these two processes look like when they are in operation, and why anyone might think we have two reasoning processes in the first place.

I begin by exposing a variety of psychological cases that motivate the dual process distinction. Based on these cases, I argue that a dual process model should offer a coherent explanation of the *structure* of the relationship between the two types of process: do they run in parallel, sequentially, or in some other way? Closely related to structure, a dual process account should explain specific *triggering conditions* for Type 2 processes to engage with a certain task.

I offer a dual process account that meets these three criteria. In this chapter, I address the structure and triggering criteria; I address fundamenality in the following. I agree with the recent work of Kahneman, Evans, and Stanovich in arguing for a default-interventionist processing structure. According to a default-interventionist schema, subjects typically rely on Type 1 default processes— in a state that Kahneman calls "cognitive ease"— unless the greater accuracy of Type 2 processing is demanded. Previous default-interventionist accounts are silent, however, on the crucial question of triggering: in virtue of what is Type 2 intervention triggered or not triggered? I argue that the assent affect— or more correctly its absence— is an excellent candidate to play this triggering role. I argue that these proto-emotions play the role of "rational traffic cops"; they determine when our fast, sloppy Type 1 is sufficient and when our careful, effortful Type 2 must intervene, thereby preserving a parsimonious and efficient reasoning schema. I spell out a full account of the metaphysics of the assent affect and the functional role these proto-emotions play.

I will not address fundamentality until the next chapter. There, I will argue that Type 2 inferences involve the manipulation of self-revisable rules, while Type 1 inferences do not involve such rules. For the sake of clarity, it is worth noting now that I follow an emerging orthodoxy in holding that Type 1 processing corresponds to a set of autonomous, domain-specific processing modules, while Type 2 reasoning corresponds to a single, flexible, domain-general processing system (Stanovich 2004, 2011; Evans and Stanovich 2013). For that reason, Type 2 inferences are united by a fundamental explanation (i.e., self-revisable rules), and Type 1 inferences are defined in opposition. These issues will be discussed at length in Chapter 3; Chapter 2 is devoted to laying out a

plausible picture of how dual processes might cooperate: how are they *structured*, and how is Type 2 *triggered*?

Before moving on, I will briefly provide some philosophical context for these issues.

Philosophical interest of dual process models

The dual process distinction holds philosophical interest both for intrinsic reasons and because of its applicability to a variety of philosophical problems. Historically, philosophers dichotomized mental activity long before empirical evidence for the distinction began to emerge: William James claimed we are capable of two types of thinking: "empirical thinking" is "only reproductive," consisting of elements and abstractions from past experience and appearing to the subject as "trains of images suggested one by another" (James 1890/1950, 325). James distinguishes empirical thinking from "genuine thinking," which is "productive," in that it can handle new types of experiences. Norman Malcolm more recently drew a distinction between "thinking" and "having thoughts," where "having the thought that p" refers to a subject's explicitly formulating and entertaining the proposition p, and "thinking that p" is a more primitive state that does not require the subject to formulate or entertain the proposition p (1973).

Though neither James nor Malcom's distinctions map neatly onto the modern dual process hypothesis, their accounts are informative in at least one way. Both James and Malcolm define each of their types of thought in terms of the nature of its processing. That is, their accounts tell us something about how each type of thought handles

information differently than its counterpart. Their accounts meet what I am calling the fundamentality criterion.

Many contemporary philosophical dual process accounts have focused on domain-specific reasoning. That is, they have taken the generic dual process distinction and applied it to philosophical questions within moral reasoning (Greene, Haidt, Saunders) and belief ascription, both third- and first-personal (Apperly & Butterfill, Goldman). I believe that the dual process model has been applied to these domain-specific philosophical problems even while substantial questions remain about how such a model might work. It is my intention in this chapter and the following to provide a philosophically coherent generic dual process model, which could be applied to these and potentially other domains.

2. Motivating cases

Most generic dual process models share a common set of motivating cases, namely, cases in which subjects seem to simultaneously possess two conflicting or contradictory beliefs. One such case is the Wason selection task, in the domain of deductive reasoning (Evans 1977 and many subsequent sources). I discussed the Wason case in the previous chapter, and I will briefly reiterate that here. In one version of this test, subjects were shown four cards on a table, each of which they were told had a color on one side and a number on the other: solid red, solid brown, the number 3, and the number 8. Subjects were then asked which card or cards they must turn over to test the truth of the proposition, “If a card shows an even number on one side, then it must be red on the opposite side.” This proposition can only be invalidated by a card that has an even

number on one side but is not red on the other, so the 8 and the brown card should be turned. Fewer than ten percent of subjects succeed in this task; most subjects responded that the 8 and the red card should be turned. Evans hypothesizes that this common false response is produced by a primitive matching bias, a Type 1 process in modern terminology: “red” and “even” are named explicitly in the prompt. When the correct answer is explained to them in terms of the rules of logic, i.e. a Type 2 process, many subjects can see that it is correct, but continue to feel the “pull” of their original, incorrect belief. In these cases, subjects simultaneously believe that the correct answer is true and that the incorrect answer is true. The outputs of their Type 1 and Type 2 processes, in other words, contradict each other.

Another motivating case, from the domain of social reasoning, emerges from implicit association tests, which claim to demonstrate contradictions between a subject’s stated and implicit beliefs. One prominent example evaluates racial stereotypes: subjects are first asked to evaluate a set of statements about their preferences for certain races over others— most subjects “disagree” with the racist propositions. In the test itself, subjects are shown a series that includes words and faces and are asked to categorize each. First, subjects are asked to categorize the words or faces as, e.g. “European-American [faces] or Bad [words]” versus “African-American or Good.” Then, subjects are shown the same series of words or faces, except the categories change to “European-American or Good” versus “African-American or Bad.” Many European-American subjects demonstrate much slower reaction times in the first portion of the test, suggesting that they have strong associations between positive words and their own race and between negative words and a different race. These results are often interpreted as demonstrating the

simultaneous presence of contradictory beliefs: Type 2 processes output explicit beliefs that races are equal in subjects' eyes, while Type 1 processes output implicit beliefs corresponding to subjects' racial preferences.

Apparent conflicts between the two reasoning processes can emerge even in cases dealing with explicit argument examination. Sloman, for example, presented subjects with the proposition, "All birds have an ulnar artery." He asked subjects to rate several conclusions based on how convincing they found these inferences. Sloman found that subjects (even those with statistical training) rated the conclusion, "Therefore, all robins have an ulnar artery" as far more convincing (9.6 out of 10) than the conclusion, "Therefore, all penguins have an ulnar artery" (6.4 out of 10) (Sloman 1996, 12). Obviously, any subject with a grasp on the relevant concepts should rate these two conclusions as equally plausible. Sloman interprets these results as revealing dual processes: many subjects will have a Type 1 judgment that penguins are less likely to have a general feature of birds than robins are, but this judgment will be overruled by a Type 2 judgment about the logic of the prompts in subjects who engage in Type 2 processing.

The cases of the Wason selection task, the implicit racial bias tests, and Sloman's logic tests are a few of the cases most commonly cited as evidence for dual reasoning processes. The presence of contradictory judgments in these examples motivates dual process accounts in that each judgment is thought to be the output of one or the other system. Due to the potential for conflict between them, the two processes are commonly framed as competing with each other for control over behavior.

3. The structure criterion

The adjudication of the competition discussed in the previous section has generally been taken to be a significant explanatory burden for dual process models. While cases in which beliefs compete with or contradict each other are interesting and informative, my account emerges partly from the observation that these examples are a limited subset of reasoning cases. In many standard cases— perhaps even most of our mental lives— the two types of reasoning do not (as a matter of fact) output different or contradictory beliefs when presented with the same problem or information. There are cases in which the use of only one process allows us to successfully reason through a problem, and there are cases when the two processes are both employed and one "confirms" the judgment of the other. The dual process literature, I believe, has come to focus its attention on one limited set of cases in which the central phenomenon is employed.

I believe that a dual process model should be concerned with the full range of reasoning cases. I can see five types of reasoning cases that would require explanation on a dual process model. It is possible that one or more of these cases is not plausible, but a successful account ought to explain why not.

i. Type 1 process engages. Type 2 process does not engage.

ii. Type 1 process engages. Type 2 process engages. The judgments output by each process are aligned.

iii. Type 1 process engages. Type 2 process engages. The judgments output by each are not aligned; Type 2 judgment controls behavior.

iv. Type 1 process engages. Type 2 process engages. The judgments output by each are not aligned; Type 1 judgment controls behavior.

v. Type 2 process engages. Type 1 process does not engage.

Accounts are broken down into three general categories in terms of their structure: preemptive, parallel-competitive, and default-interventionist. In this section, I will argue that preemptive models cannot accommodate cases ii., iii., and iv., in which both processes simultaneously engage, and I will argue that parallel-competitive models cannot accommodate cases i. or v., in which only one process engages. I will argue that only the default-interventionist paradigm is capable of handling cases where either one or both processes engage, although existing default-interventionist accounts have trouble with cases iv. and v. In a later section, I will present a default-interventionist view capable of handling each of these five cases.

Preemptive conflict resolution models

Some dual process accounts, such as Klaczynski (2000), hold that there is early separation of processing between the two systems. That is, certain superficial aspects of the stimulus or problem at hand trigger either a Type 1 or Type 2 process. For example, when evaluating deductive arguments, one might use Type 1 reasoning to evaluate arguments whose conclusion is superficially "believable"— compatible with the subject's set of existing beliefs— and only trigger Type 2 reasoning to evaluate arguments whose conclusions one finds surprising. That is, one is more likely to invoke effortful Type 2

processes to scrutinize arguments for conclusions one disagrees with. This type of account has been referred to as the "preemptive conflict resolution" model.

As seen in the previous section, the dual process hypothesis tends to be motivated by cases in which a subject entertains conflicting attitudes. It is unclear how a model according to which conflicts are resolved preemptively could account for cases like the Wason selection task where a subject simultaneously entertains conflicting beliefs.

Even if this type of model is relaxed to allow that two different types of superficial stimuli could simultaneously trigger the two different types of processes, the theory still owes us an account for how the conflicting outputs of these two processes would be adjudicated. The preemptive conflict resolution model is inadequate to explain case ii., as well as cases iii. and iv., which are the competition cases that motivate the dual process hypothesis.

Parallel-competitive conflict resolution models

The problem with the preemptive conflict resolution model is that it does not allow for both processes to ever be simultaneously applied. "Parallel-competitive" conflict resolution models hold that both processes simultaneously engage in *every* instance of reasoning. These models hold that the types of reasoning process engage in parallel, proposing responses to reasoning tasks. If the responses of the two processes are different, this conflict must be resolved after both processes have run. One very prominent account along these lines comes from Sloman (1996, 2002). Sloman holds that the distinguishing factor between the two types of reasoning is that Type 1 tends to involve "associative" reasoning, while Type 2 tends to involve "rule-based" reasoning. It

is worth briefly mentioning here that Sloman's associative vs. rule-based characterization of the dual process distinction meets the fundamentality criterion in that it characterizes the nature of each type of process. I will discuss this distinction at greater length in the third chapter of this dissertation, which is devoted to the fundamentality criterion.

For present discussion of the structure of various dual process models, it is important only that Sloman's associative system corresponds to Type 1 reasoning and that his rule-based system corresponds to Type 2 reasoning.[3] In his dual process account, Sloman makes very clear that he supports a parallel processing model:

> Both systems seem to try, at least most of the time, to generate a response. The rule-based system can suppress the response of the associative system in the sense that it can overrule it. However, the associative system always has its opinion heard and, because of its speed and efficiency, often precedes and neutralizes the rule-based response. (2002, 391)

This model is conventionally parallel-competitive in that both processes reason through a problem independently first, and then the dispute must be adjudicated. In other words, every time one is presented with a reasoning problem, both processes must engage.

Consider the Wason selection task, in which most subjects reach an incorrect answer very quickly via Type 1 reasoning. It does not seem likely that Type 2 reasoning has begun to simultaneously engage in cases like this. Sloman, or someone with a model like his, would be forced to argue that a deliberate Type 2 process was already engaged with the task, but was aborted as soon as the Type 1 judgment was output. After subjects' attention is redirected towards the logical features of the prompt, the parallel-competitive

[3] Sloman's use of "system" rather than "type" or "process" is meant to signal his further claim that the dual process hypothesis obtains at a deeper level of cognitive architecture (although these terms are not used consistently throughout the literature). That is, Sloman holds that each type of processing refers to natural kinds within cognitive activity. I stick with the weaker "dual process" nomenclature, as I intend my account to be in dialogue with a broad range of previous work covering an array of claims about connection to deeper cognitive architecture.

model is forced to say that they are now reengaging Type 2 processes, rather than engaging them for the first time.

In many simple reasoning tasks, it seems most plausible that Type 2 does not engage at all. Imagine playing a simple matching game on your smartphone on your subway commute home: your thoughts are elsewhere, but you are able to perform Type 1 tasks more or less effectively. That is to say, case i. above describes a real phenomenon. If Type 2 processes are not engaged in this case, why are they not? The parallel-competitive paradigm offers no natural explanation for any cases in which Type 2 does not engage in a reasoning task. It does not help to allow that there are some cases in which Type 2 does not engage; we are then left without an explanation for what does or does not trigger this engagement.

It may not be obvious at first pass whether cases like v., where there is no Type 1 default and Type 2 judgments response immediately to a prompt, actually exist. Consider, though, reading a question on a math or logic test. In these cases, I generally do not have a "default" intuition, but rather I immediately input the prompt into my Type 2 processes and begin hypothesizing the applications of various rules of math or logic. The parallel-competitive model is inept at explaining this type of case as well; we have no explanation for why Type 1 would not engage in these cases.

While the preemptive paradigm does not easily accommodate cases where both processes engage, the parallel-competitive paradigm does not easily accommodate cases in which only one process engages.

Default-interventionist conflict resolution models

The third category of dual process models holds that most of our reasoning processes occur via Type 1 processes running as a default in the background, with Type 2 processes occasionally intervening. This is called the default-interventionist paradigm.

According to Evans (2009), the defining feature of Type 2 processing is that it requires the resources of working memory. The recent account endorsed by Evans & Stanovich (2013) maintains this claim and extends the view by claiming that Type 2 reasoning is demanding on working memory because it necessarily involves "cognitive decoupling." This refers to the formation of a mental copy of a representation, in order to allow for hypothetical manipulation. Trying to solve a logic proof, for example, one may imagine several possible next steps, but must do so by forming a copy of the original mental representation of the proof's premises and goal, so that the original representation is not lost due to the hypothetical manipulation. Correspondingly, the defining feature of Type 1 processes for Evans and Stanovich is that, "They do not require 'controlled attention,' which is another way of saying that they make minimal demands on working memory resources" (2013, 236).

Both Evans (2009) and Evans & Stanovich (2013) agree that Type 2 processing is costly in terms of cognitive resources. Only in cases where Type 1 default processes are unable to reach a satisfying judgment do Type 2 processes step in to settle the issue. The default-interventionist paradigm is parsimonious: if one believes that we have evolved to use our cognitive resources economically, this type of model should hold substantial appeal. When the relatively modest demands of Type 1 processing are sufficient, Type 2 is not involved. When Type 1 processes are insufficient to reach a satisfying judgment, our Type 2 processes are recruited to help get the job done.

The basic default interventionist paradigm seems to be able to explain the reasoning cases that the previous two paradigms could not. The preemptive conflict resolution model could not explain any cases in which both processes engage; the default-interventionist holds that these are cases in which Type 2 intervenes on Type 1 defaults. The parallel-competitive model could not explain cases in which *only* Type 1 was engaged; the default-interventionist holds that these are cases in which Type 2 simply does not intervene. Because it can explain many cases in which either one or both processes are engaged, and because of the force of its arguments with respect to cognitive economy, default-interventionism seems the strongest available paradigm.

Existing default–interventionist models, however, cannot naturally explain cases like iv. or v. We have already discussed case v., where only Type 2 engages, such as when taking a math or logic exam. Cases like iv., where Type 1 outputs "beat out" Type 2 outputs for control of behavior, also seem to exist; in certain types of reasoning, we might describe such cases as weakness of will. In particular, when engaged in decision-making, we might feel the immediate pull of some certain decision that is unhealthy or unwise. We may reflect on the choice, via Type 2 processing, and think of several reasons to choose a healthier option, and yet the Type 1 impulse wins out anyway. Standard forms of default-interventionism cannot easily accommodate these cases: if Type 2 has intervened, how can the default win out anyway? That outcome is not precluded by the structure of the model, but there is no explanation for how it could occur either. The form of default-interventionism I offer will be able to accommodate both cases iv. and v.

Another problem with existing default-interventionist accounts is that they have no adequate explanation for what specifically triggers Type 2 intervention— this is one of the criteria I established for dual process models. I will presently explain the importance of including a triggering mechanism in a default-interventionist model, and demonstrate that Evans and Stanovich fail to do so. I will offer a positive account of default-interventionist triggering in a later section.

The importance of the triggering criterion

Examining views like Kahneman's, or Evans' and Stanovich's, we might ask what, specifically, triggers the intervention of Type 2 processes— when is it inappropriate to substitute simpler attributes? When is the judgment output by Type 1 insufficiently satisfying? Evans and Stanovich claim that Type 2 processing is invoked "when a decision matters[…] for example, when we are evaluating important risks," giving the example of a threat to one's children (237). But what determines whether a decision is important? As best I can tell, Evans and Stanovich do not provide an answer to this question. First of all, a threat to one's children seems like a strange example of a case when Type 2 processes are likely to be invoked. If a mother sees a snarling, unleashed dog near the playground where her children are playing, for example, one would think that she is likely to act quickly upon a Type 1 judgment that there is a threat, perhaps by ushering her children towards the car or reprimanding the dog's owner.

Putting this example aside, let us consider the Wason selection task, a more typical dual process case. Suppose a subject is given the prompt and immediately reaches a Type 1 judgment that he ought to flip two particular cards. Some factor then triggers his

Type 2 reasoning: he rereads the prompt, reflects on his understanding of the material conditional, and responds to the question. Importantly, the input into his Type 2 process would *not* be the output of his Type 1 judgment; one would not begin explicitly deliberating on the task by thinking about one's previous incorrect answer. Rather, Type 2 reasoning shares an input with Type 1 reasoning: the prompt itself. Furthermore, there was nothing in the content of the Type 1 judgment that would trigger Type 2 to kick in—the Type 1 judgment was simply an endorsement of a false proposition, something like, "The correct response is that 8 and brown should be flipped." There is nothing within that proposition that signals its potential falsity; its potential falsity only comes into view in light of some external consideration.

One line would be to argue that Type 1 judgments are accompanied by feelings of confidence, and the extent of that confidence determines whether Type 2 processes are recruited. But what about cases where it is the importance of the task that triggers Type 2? In these cases, one would have to say that running parallel to whichever Type 1 process is devoted to a particular problem is a second Type 1 process devoted to evaluating that problem's importance.[4] Though either confidence or a belief in the task's importance is a potentially plausible trigger of Type 2 processing, either of these would be additions to the default-interventionist account. It appears that some additional feature would be required if a default-interventionist account wishes to explain what might cause Type 2 processes to intervene on default Type 1 judgments.

In Evans (2009), he claims that "working memory [is] largely recruited by preconscious systems," giving the example of a distracted motorist jarred into

[4] I will argue that some of these cases may be like v., where Type 1 has not engaged at all on a particular reasoning task.

attentiveness by the car in front of him braking suddenly (48). This example, however, is not a case of dual process reasoning— if there is any reasoning involved here, it seems like a very basic Type 1 spatial reasoning process; no deliberation or hypothetical thought was involved. In this earlier work, Evans describes his account as default-interventionist, but also proposes that we may also have Type 3 processes, which Type 1 would input into directly. Type 3 processes would then determine whether or not Type 2 processes needed to be recruited, and, if the Type 2 outputs ultimately disagreed with the Type 1 outputs, Type 3 processes would resolve that conflict.

The Type 3 proposal is not worth paying much attention to. Evans himself acknowledges that he is only able to define it functionally (2009, 49). In other words, it is clear to him that some further mechanism or detail must be added to the default-interventionist schema in order to explain the recruitment of Type 2 processes, but this unexplained triggering condition is the entire basis for proposing Type 3. As such, Type 3 is either a very unparsimonious solution to the problem facing default-interventionism, or it is an empty placeholder for an eventual explanation he hopes will eventually validate his account. Evans himself seems to have abandoned Type 3 by his 2013 paper with Stanovich. This leaves default-interventionism without any explanation for what specifically triggers Type 2 processes to intervene on Type 1 processes.

4. The triggering criterion

In the previous section, I argued that we have many reasons to favor a default-interventionist paradigm along the lines of the accounts developed both by Kahneman and by Evans and Stanovich. No existing default-interventionist account, however, offers

an explanation for the triggering of Type 2 reasoning. That is, in virtue of what do Type 2 processes intervene on Type 1 defaults? I find this silence on the part of foregoing accounts very troubling: the parsimony enabled by this type of intervention is the crucial virtue of this type of dual process model, and it is unclear what explanatory value such a model could provide if we are left without any idea of what might trigger Type 2 to intervene.

I have so far mentioned a few possibilities about what the triggering conditions for Type 2 intervention *cannot* be. It cannot be the case that Type 2 reasoning constantly monitors Type 1 reasoning, as this would violate the considerations of cognitive economy that motivated the default-interventionist schema to begin with. If Type 2 reasoning always ran in the background, its resources could not be said to be conserved by this paradigm. If there were some kind of "off-line" monitoring on which Type 2 processes could monitor Type 1 processes, it is not clear how these monitoring processes could qualify as Type 2 processes at all— they would be neither effortful nor deliberate.

It is also not possible that something internal to a belief or judgment output by Type 1 could trigger Type 2 intervention. As discussed earlier in this chapter, many of the motivating cases for dual process models rely upon the observation that these two types of reasoning can output identical judgments— dual process models could be construed as answers to the question, "why would this duplication of reasoning function exist?" If Type 1 and Type 2 are capable of outputting the same judgment, then it does not seem that something internal to the Type 1 output could be the trigger of Type 2 intervention.

In the previous chapter, I argued that all that is required for a new belief to count as inferred is for the subject to take that new belief to follow from a rational process. Furthermore, I argued that a subject's taking a new belief to follow from a rational process requires only the presence of a certain phenomenological datum: the feeling of correctness, which I called the assent affect. A subject makes an inference if and only if a causal cognitive transition produces a belief accompanied by the assent affect.

In this section, I will argue that, if a Type 1 judgment is not accompanied by the assent affect, Type 2 intervention is triggered. This absence of the assent affect may carry its own phenomenology— there may be a set of uncertainty-based intellectual emotions or epistemic feelings that present to the subject when a judgment is reached. In order to develop this claim that certain intellectual emotions (or their absence) are the triggers of Type 2 intervention, I must first discuss the role of emotions in reasoning in greater detail.

In a recent, unpublished paper called "Reasoning and Emotion, in light of the Dual Processing Model of Cognition," Ronald de Sousa argues that emotions play important roles in a variety of reasoning tasks, and that these roles span the two tracks of reasoning proposed by dual process models. De Sousa defends this claim with a series of cases borrowed from Kahneman's 2011 *Thinking Fast and Slow*. Kahneman argues that humans are cognitively "lazy," in the sense that they tend to offer biased and obviously false answers to certain types of tasks rather than think through them carefully. In one study, Princeton undergraduates were given the prompt along the lines of: "*In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the*

lake?" (Kahneman 2011, 66). Half of the subjects were given the prompt in a clear, easily readable font, and the others were given the prompt in a grey, low-quality photocopy that was somewhat difficult to read. Subjects who read the prompt in the clear font made an error 90% of the time, while subjects who had to strain to read the prompt made an error only 35% of the time (ibid).

Kahneman argues that results like this suggest that humans tend to prefer states of "cognitive ease" where only Type 1 processes are deployed. He claims, "These findings add to the growing evidence that good mood, intuition, creativity, gullibility, and increased reliance on System 1 form a cluster." I believe that these results are well-explained by the assent affect, a very general feeling of confidence in one's judgment. If one's judgments are disposed to be accompanied by the assent affect more readily, they will be more inclined to rely upon their Type 1 reasoning resources, without Type 2 intervention.[5]

Epistemic feelings along the lines De Sousa lays out are precisely the sort of thing that might trigger Type 2 intervention. If a subject has a Type 1 judgment that is accompanied by the assent affect, or the feeling of correctness, Type 2 processing will not intervene.[6] If, however, the Type 1 judgment is not accompanied by the assent affect, Type 2 processing will intervene. There is a further question, of course, about how to describe the absence of the assent affect. I believe that the mere absence of the assent affect carries phenomenology: there is something it is like to not be satisfied with a judgment. This absence— or the array of feelings that might constitute this absence— is

[5] Kahneman's theorizing along these lines draws heavily upon Norbert Schwartz's work on metacognition (Schwartz et al. 2007).
[6] In the final Chapter 3, I will discuss cases of unconscious rule-following. In order to fully explain such cases, I must first grapple with the fundamentality condition for dual process models.

all that is required to trigger Type 2 intervention. It also seems, however, that the absence of the assent affect alongside a Type 1 judgment might carry further phenomenology: confusion, doubt, etc. These epistemic feelings could also serve as triggers for Type 2 intervention. These further feelings, however, are not universal: what unites them is that in each case the subject makes a judgment that is not accompanied by the assent affect.

When a subject reasons her way to a new belief, but that belief is not accompanied by the assent affect, she has not made an inference. That is, when a subject's fails to reach a response to a math problem using Type 1 processing, she has not yet made an inference. The absence of the assent affect will trigger Type 2 intervention; those Type 2 processes may yield a judgment that is accompanied by the assent affect—a judgment the subject is satisfied with. Only at this point has the subject made an inference. If the Type 2 processes still do not produce a judgment that is accompanied by the assent affect, Type 2 processes are once again triggered. Type 2 processes will reapply themselves until the subject reaches a judgment accompanied by the assent affect, or until the subject gives up. In this way, the assent affect and its absence act as rational "traffic cops," directing Type 1 and Type 2 reasoning processes to intervene or not, as is required by the situation.

At this point, I must say more about the nature of the assent affect. In his 2013 paper "The Nature of Cognitive Phenomenology," Declan Smithies considers the phenomenal properties of cognition in considerable depth. He focuses on (a) the relationship between the phenomenal properties of cognition and the intentional properties of cognition, and (b) the relationship between the phenomenology of cognition

and the phenomenology of sensory experience. The second question is entirely outside the scope of my present project.

With respect to the first question, Smithies identifies two general views. The first view, intentionalism, claims that "the phenomenal properties of cognition are necessarily connected with, and even identical with, its intentional properties" (Smithies 746). The alternative view, which I will call anti-intentionalism, claims that phenomenal properties of cognition are related only contingently with the intentional properties of cognition. Proponents of intentionalism claim that phenomenal properties play a role in the individuation of cognition. It is beyond my present ambition to give an account of how every aspect of cognition is individuated: the individuation conditions for states like belief may vary considerably in nature from the individuation conditions for intentional attitudes or particular types of intentional contents.

I have, however, given the individuation conditions for inferences in terms of one particular phenomenal property of such cognitive events: namely, the assent affect. I am agnostic with respect to intentionalism, broadly construed: the view that phenomenal experience is relevant and necessary to the individuation of any aspect of our cognition. I am arguing for a narrower and more modest version of intentionalism, on which at least the category of inference is individuated with respect to the presence of this particular phenomenal property of cognitive experience.

Two individuals may share an identical set of background premise-beliefs, and they may arrive at an identical new belief based on these background premise-beliefs. The two subjects may differ with respect to whether or not they have made an inference, however, based on whether or not the new belief is accompanied by the assent affect.

Whether or not the new belief is accompanied by the assent affect— that is, whether or not the chain of beliefs can be said to constitute an inference— is a matter of individual difference. Some subjects may simply be more likely to be satisfied with their Type 1 judgments, and so this judgment may constitute an inference for these subjects more often.

This final note corresponds nicely to one of the pairs of features generally associated with Type 1 and Type 2 processing. It is often claimed that individuals are largely similar with respect to their Type 1 processing abilities, with differences in Type 2 processing accounting for differences in cognitive ability. The 2013 Evans & Stanovich account, which grounds Type 2 reasoning in reliance on cognitive decoupling, emerges very directly from Type 2's correlation with individual differences. Now, we can say that (at least some) individual differences reflect differences in subjects' dispositions to feel the assent affect. Subjects whose Type 1 judgments are less frequently accompanied by the assent affect will be less likely to trigger Type 2 intervention, and they will rely upon relatively less accurate Type 1 judgments more often.

It is very likely that the assent affect can vary with respect to degrees of intensity. Some individuals might feel extremely confident that a certain conclusion is right— Descartes basking in the glow of the natural light, for example. The same inference, for another individual, might yield the assent affect to a lower degree— she will be less confident in the output of her judgment, though still confident enough to endorse it. This result dovetails very nicely with the observation that there is significant individual difference with respect to the application of Type 2 reasoning. It is possible that individuals differ in terms of the extent to which they are disposed to feel the assent

affect. It is also possible that individuals differ with respect to the threshold of the assent affect below which Type 2 reasoning is triggered. For the sake of example, Person A might have Type 2 reasoning triggered whenever the assent affect falls below 70%; Person B might have a 50% threshold. (Of course, I do not believe that there must be literal percentages; these are included in order to demonstrate the point.) Person A is more prone to think things over carefully: their Type 2 processes are more readily invoked.

The assent affect allows us to see how certain contexts or situations might affect the triggering conditions for Type 2 reasoning. Situations of great importance arouse heightened sensitivity, nervousness, fear, etc.; that is, an array of intellectually relevant emotions. These emotions might impact the assent affect itself— e.g., perhaps we are more likely to suppress assent when we are very nervous— or these emotions might lower an individual's triggering threshold. If the bar below which Type 2 processing triggered is lowered in situations the individual takes to be important, Type 2 processing will be triggered more readily. Grounding the triggering of Type 2 reasoning in emotions allows us a very natural explanation of why Type 2 is more likely to be invoked in "important" situations: these situations are emotionally charged, and adjust the emotional triggering conditions for Type 2 intervention.

We also now have an explanation for case v., as described in the previous section. These were cases in which a subject relied only on Type 2 processing, with no Type 1 intervention. Imagine a high schooler taking the SAT: he is very nervous, and this nervousness might countervail the assent affect— these will be competing emotional factors. If the subject's assent affect is suppressed before he even begins to read the

question, he will rely on Type 2 reasoning immediately. A more confident test-taker might not have this countervailing nervousness, and so his Type 1 processes might engage on his first pass through the question; only if and when he fails to have the assent affect will his deliberate Type 2 processes engage.

The assent affect comes in degrees to the same extent that confidence in a judgment comes in degrees. This does not suggest, however, that reasoning is best explained along a spectrum rather than a dichotomized dual process model. There is a definite level of "feeling right" below which Type 2 will be triggered. Even if this level changes based on context, there will always be a definite point along these lines.

Finally, it is worth revisiting cases like iv., in which Type 2 intervention "disagrees" with a Type 1 judgment but Type 1 succeeds in control of behavior. Once affective attitudes are introduced to the picture, it is easy to see how the emotional force of a Type 1 judgment over a Type 2 judgment might allow the Type 1 judgment to successfully control behavior. Consider Alfred Mele's example of a parent who refuses to be persuaded that her teenage son is using drugs (2001). Perhaps reasoning through the issue via Type 2 rules seems to point towards the son using drugs, but she persists in ignoring this evidence in favor of her judgment that she just "knows" her son wouldn't use drugs. The Type 1 judgment that her son is not using drugs is emotionally preferable to such a great extent that it overwhelms the usual emotional and subjectively normative weight of the Type 2 judgment. Now that emotions and epistemic feelings have entered the picture as triggers for preferring both Type 1 and Type 2 attitudes, we can explain cases like this that might otherwise escape easy description on a default-interventionist schema.

The appeal of the assent affect as an explanation for triggering

I have argued that the assent affect plays the role of a rational traffic cop: when it appears alongside a Type 1 judgment, that judgment counts as inferred, and Type 2 reasoning is not triggered. When it does not appear, this event has its own particular phenomenology, and Type 2 reasoning is triggered. My argument, thus, has significant theoretical appeal. I identified an egregious explanatory burden left by foregoing default-interventionist accounts, in that none of them has explained what might cause Type 2 reasoning to intervene on Type 1 defaults. The presence of absence of the assent affect is a very strong candidate to fill this explanatory gap regarding how Type 2 intervention is triggered. In the previous chapter, I argued that the assent affect also solved a critical problem in accounting for the category of inference: if subjects are often unaware of the causal history of an inferred conclusion, then only something like the assent affect could explain how it is that a subject "takes" a conclusion to follow from a rational process. I have demonstrated that an account of reasoning and inference that includes the assent affect can solve multiple theoretical problems at once.

Chapter 3 – Self-revisability as the fundamental distinction between dual reasoning processes

1. Introduction

Let us begin by taking stock of what I have covered so far. In the first two chapters, I argued that two apparently unrelated puzzles— one about inference, and one about dual process models of reasoning— can be solved at once by acknowledging the importance of intellectual emotions in reasoning. In particular, the feeling of correctness, which I am calling the assent affect, does significant explanatory work in addressing these puzzles as soon as it is incorporated into descriptive accounts of reasoning and inference.

The first puzzle, discussed in the first chapter, emerged from Paul Boghossian's powerful observation that inference involves a subject's "taking" a cognitive transition to be a certain way. Boghossian says that a subject must take the premises of her inference to support her conclusion. I demonstrated in Chapter 1 that there are many cases in which subjects are blind to the premises of an inferred belief. Thus the puzzle: if subjects may not know where their inferred beliefs come from, what can this "taking" involve? I argued that, for a cognitive transition to count as an inference, a subject must take a belief to be the output of a rational process, where that is defined by the belief's accompaniment by the assent affect. A belief counts as inferred because the subject has the feeling that it is rational. This is necessary to accommodate the wide range of bad inferences that occur.

The second puzzle, discussed in the second chapter, emerges from default-interventionist dual process models. I argued that the default-interventionist paradigm is the best explanation for how dual processes might be structured: we generally rely on

Type 1 defaults, except when these processes are insufficient to address a prompt or stimulus effectively. When the defaults are insufficient, Type 2 processing is triggered. I argued that default-interventionism is vastly preferable to other dual process paradigms. Unfortunately, every foregoing default-interventionist account I am aware of (Evans 2009, Evans & Stanovich 2013, Kahneman 2011) is silent with respect to what specifically triggers Type 2 intervention. Here we find the second puzzle: what causes Type 2 intervention? I argued that the assent affect is well equipped to solve this puzzle as well. When a Type 1 default does not yield the assent affect, our Type 2 processes intervene, and we think about a problem deliberately and carefully. Explaining this triggering via emotions does a tremendous amount of explanatory work. For example, we can now say that our heightened attention during important scenarios, like final exams—our increased propensity to rely on Type 2 reasoning— occurs because the emotional salience of these events countervails the assent affect. That is, the nervousness we feel during an exam suppresses our inclination to accept the Type 1 "easy answer" as true without thinking things over carefully.

In the second chapter, I met two of the three explanatory criteria I established for dual process models of reasoning. The default-interventionist paradigm meets the *structure* criterion, and the assent affect meets the *triggering* criterion. I have not, yet, explained the *fundamentality* criterion. This chapter is devoted to this final criterion. What is the nature of each type of processing? How do they manipulate a stimulus differently such that they are not merely duplicating functionality? My answer to these questions will rely on the account of inference I gave in the first chapter and the considerations about the structure of dual process models I offered in the second.

In building my account of fundamentality, I will draw on the discussion of the diversity of types of inference discussed in the first chapter. I will argue that Type 2 reasoning involves inferences based on rules that are self-revisable within any particular instance of reasoning, and Type 1 reasoning involves inferences that are not based on self-revisable rules. I will spell out the relevant notion of self-revisability in great detail.

Among contemporary accounts, the orthodoxy is that Type 1 processing corresponds to a set of autonomous, domain-specific processing modules, while Type 2 reasoning corresponds to a single domain-general processing system flexible enough to handle several types of reasoning (Stanovich 2004, 2011; Evans & Stanovich 2013). For this reason, I offer self-revisability as a necessary and sufficient condition for an inference to qualify as an instance of Type 2 reasoning, and define Type 1 inferences simply as those that are not self-revisable. If Type 1 reasoning involves autonomous modules, these modules need not rely on the same type of processing; these modules are similar only in that they facilitate inferences that are not self-revisable.

Before discussing my account in more detail in Section 3, I will briefly address a few ways that fundamentality has been addressed in the foregoing dual process literature.

2. The basis of the fundamentality criterion

Dual process accounts generally involve lists of qualitative features associated with each type of reasoning. The specific features included on these lists will vary depending on the author, but only slightly. In the introduction to their 2009 collection "In Two Minds," Evans & Frankish offer the following list of features they take to appear most frequently in other authors' work:

System 1	**System 2**
Evolutionarily old	Evolutionarily recent
Unconscious, preconscious	Conscious
Shared with animals	Uniquely (distinctively) human
Implicit knowledge	Explicit knowledge
Automatic	Controlled
Fast	Slow
Parallel	Sequential
High capacity	Low capacity
Intuitive	Reflective
Contextualized	Abstract
Pragmatic	Logical
Associative	Rule-based
Independent of general intelligence	Linked to general intelligence

(Evans & Frankish, 2009)

Any particular author might disagree with the inclusion of some of these items; I have no objection to the modest claim that these features are typical but non-necessary. If the dual process distinction is a real distinction, however, we must say which of these features is necessary to each type of process. We must be able to differentiate between Type 1 and Type 2 inferences via some consistent basis.

A wide variety of fundamental explanations of the two types of reasoning have been offered in the literature. In the previous chapter, I discussed Sloman's dual process account (1996, 2002), on which System 1 is fundamentally associative and System 2 is fundamentally rule-based. Frankish (2009) attempts to map the dual process distinction onto the personal/subpersonal distinction. Carruthers (2009) claims that System 2 is a "virtual machine" realized out of Type 1 hardware.

These ways of meeting the fundamentality criterion, and most of the others in the literature, are ruled out by the considerations I laid out in the previous chapter. Few of

them are compatible with the very promising default-interventionist paradigm, and furthermore many of them struggle to explain the fully array of psychological data that motivates the dual process distinction.

In the previous chapter, I mentioned the Evans & Stanovich (2013) answer to the fundamentality question when laying out their default-interventionist model. They argue that Type 2 reasoning relies upon the faculty of cognitive decoupling: forming a mental copy of a representation is a means of entertaining hypotheticals. The autonomous set of systems that engage in Type 1 reasoning do not involve cognitive decoupling, on their view. I do not think that this claim is entirely incorrect, but I also do not think that it fully answers the call of the fundamentality criterion. In the following section, I will show how this is the case.

3. The fundamentality criterion and self-revisability

Let me draw a loose analogy. Suppose I tell a recent Martian immigrant to Earth that humans are capable of representing visual space in two different ways. What are they, the Martian asks. Well, I say, one is the primary way that visual information enters our awareness and helps us respond quickly to our immediate environment, and the other is cognitively downstream and helps us navigate the space around us. The first one provides information that the second process uses to form its different sort of spatial representation. Okay, the Martian might respond, but what *are* they? We can define the first by its primacy, I reply, and the second by its manipulability and the demands it places on our working memory.

This scenario, I believe, is roughly akin to defining Type 1 and 2 reasoning in terms of Type 1's autonomy and Type 2's reliance on working memory or decoupling, as Evans & Stanovich do. The Martian's question would be best answered with the information that the first type is egocentric representation, meaning that it represents the locations of objects relative to the subject's position, and the second type is allocentric, which represents the locations of objects relative to other objects.

I do not deny that Type 1 reasoning is autonomous in the relevant sense or that Type 2 reasoning places heavy demands on working memory or that it requires decoupling. In fact, I think that these considerations give us strong reason to believe that something close to a default-interventionist schema— where scarce and cognitively taxing Type 2 resources only intervene when default, "low cost" Type 1 judgments fail— is correct. I do think, however, that defining Types 1 and 2 as Evans & Stanovich do fails to constitute a response to the most fundamental question about dual reasoning processes: how do the two processes manipulate a stimulus differently such that they cooperate? Put differently, what is the nature of the reasoning performed by the two processes such that they are not merely duplicating functionality? Reliance on working memory is a very important datum, of course, but it does not amount to a hypothesis about the intrinsic nature of the two types of reasoning. It is of course possible that Evans & Stanovich do not aspire to offer such a hypothesis— perhaps they think it is sufficient to say that certain qualitative features of the two processes are typical but non-necessary. If we are to understand why it might be an evolutionary advantage or normatively better for Type 2 processes to intervene on Type 1 processes, however, we are owed an explanation for

how the way Type 2 processes manipulate an input is fundamentally different than the way that Type 1 processes manipulate that same input.

In this section, I will argue that Type 2 processing involves the manipulation of self-revisable rules: rules that the subject can change during the act of reasoning. Type 1 processing is, as I have said, most plausibly construed as a set of autonomous processing systems. These systems are united under the Type 1 umbrella because they facilitate inferences that are not self-revisable in the way that Type 2 processes are.

Revisable rules

What does it mean for a rule to be revisable? On my view, the revisability of rules is the definitional or fundamental feature of Type 2 inferences. As we saw in the previous chapter, the default-interventionist paradigm explains cases in which Type 1 processes fail to deliver a satisfying response by saying that these are cases in which Type 2 processes intervene (though previous authors do not say what it is, specifically, that triggers Type 2 intervention). What about when we find the output of a Type 2 process unsatisfying? Say, for example, that a college freshman is taking Introductory Logic. Perhaps her tutor has taught her to follow the general rule:

> r1: If the main operator of the goal is a negation, proof by contradiction is the appropriate strategy.

Later, the student faces the task: "Given the premises 1. A→~C and 2. A, derive the goal ~C." Type 1 processing will not offer any satisfying answer to problems like this that demand deliberate Type 2 reasoning.

So Type 2 processes are recruited (due the absence of the assent affect, as discussed in Chapter 2). First, the student attempts to apply the rule she was taught, and begins the proof by assuming "C," with the intention of deriving a contradiction from that assumption. She then realizes that this strategy will not get her a satisfying answer. This triggers Type 2 processing to reapply itself; she considers what other rules she knows for solving proofs, and realizes that modus ponens can be performed on the two premises. There is a question, of course, about what made Type 2 processing reapply itself. Later, I will argue that the same sort of thing that triggers Type 2 to intervene on Type 1 defaults can also trigger Type 2 to reapply itself, or to fail to be satisfied, when reasoning through a certain puzzle.

In the relevant case, though, the student's eventual application of modus ponens constitutes a violation of the rule she originally knew. Her eventual success with this prompt constitutes revision of the relevant set of rules she is aware of for solving logic proofs. One possible revision could be that she replaced the original rule r1 with:

> r2: If the main operator of the goal is a negation, proof by contradiction is the appropriate strategy *unless* there is a modus ponens available in the premises given.

Alternatively, perhaps she does not replace r1 with r2, but instead keeps r1 and introduces two new rules:

> r3: Perform any modus ponens available in the premises given.
>
> r4: r3 should be followed before r1.

In this possibility, r4 is a metacognitive rule governing the application of rules r1 and r3. Regardless of whether the subject replaces r1 with r2 or keeps r1 and introduces r3 and

r4, what is relevant here is that the rules are revisable. That is, for any Type 2 inference, a subject has the option of editing the rule that led them from the input to the judgment if the judgment is unsatisfying. This is *not* the case for the inferences that I believe are fundamentally linked with Type 1 processing. If a judgment reached in this way is unsatisfying, Type 2 processes must be recruited to reevaluate the input. The Type 1 judgment cannot be revised in the short term; it can merely be overcome by a Type 2 judgment, or counter-conditioned in the long term.

Jake Quilty-Dunn and Eric Mandelbaum (2015) have recently argued that some basic logical rules, like modus ponens, are built into our cognitive architecture, and that this fact explains why we are able to make unconscious inferences involving these rules. I do not doubt that they are right that rules along these lines are in fact built into our cognitive architecture, nor do I doubt that we are able to rely on them unconsciously. These built-into-the-architecture rules will not be revisable, and so relying upon them will necessarily be a Type 1 process.

To see the distinction between Type 1 modus ponens and Type 2 modus ponens more clearly, imagine that I ask an introductory class to execute a proof, except for the sake of this proof modus ponens is not a valid inference rule. Instead, affirming the consequent should be used as a valid inference rule. Each time a modus ponens is available to the students when solving the proof, they might reach a Type 1 judgment following this rule. Their Type 2 processes, however, would overrule this judgment, relying on the affirming the consequent rule they have self-revised into validity for the sake of the example.

In the first chapter, I argued that the category of inference was much more diverse than many authors had previously taken it to be. I argued that defining inferences as essentially and necessarily rule-based was not useful, because of the diversity of types of rules that different inferences relied upon. In order to demonstrate this diversity, I laid out several different types of inference, categorized based on the types of rules they apparently relied on. I acknowledged that some of these categories could potentially be synthesized into broader categories. This is precisely what I will do now: I believe that the types of inference I laid out in chapter one are neatly categorizable as self-revisable and non-self-revisable— as Type 2 and Type 1.

Type 2 reasoning: explicit reliance on deductive or inductive rules

In the first chapter, I considered the paradigmatic cases of the mathematician and the chemist. The mathematician consciously entertains certain premises and a set of deductive rules she has learned in order to deduce the next step in her proof. The chemist considers her lab results and whether or not they are sufficiently convincing to satisfy an inductive generalization. These are paradigmatic inferences in the sense that the subjects explicitly consider what might follow from certain beliefs they already hold. I argued in the first section that these types of inferences seem to be the type of phenomenon that Boghossian and other writers have focused on in building their accounts of inference; while I acknowledge that they are paradigmatic, I think that they are also somewhat less common than many types of everyday inference.

It is clear that the rules in either of these paradigmatic cases are revisable in the relevant sense. That is, anytime a rule is explicitly relied upon in reaching an inference,

that rule is revisable, and that inference is Type 2. The case of the logic student in the prior subsection demonstrates that explicit reliance on deductive rules is self-revisable. Much the same can be said about the case of the chemist: perhaps she determines that a certain type of extrapolation she has relied upon in the past is more fallible than she previously realized; she can reassess an inference made using this extrapolation and revise that inference. Each of these types of inference should be somewhat uncontroversially Type 2.

Type 1 reasoning: associative and probabilistic deductive and inductive inference

In the previous chapter, I considered cases of deductive inference that may rely on associative or probabilistic reasoning. This was best highlighted by alternative explanations of the Wason selection task, also discussed earlier in this chapter as a motivating case for dual process models. In this task, subjects fail to correctly assess the truth conditions of a card-flipping task involving the material conditional. One (perhaps orthodox) explanation of the incorrect response to this task is that subjects are relying on a primitive associative matching bias; subjects respond incorrectly by naming the cards mentioned in the prompt. An alternative explanation, due to Oaksford and Chater, is that the incorrect strategy in the Wason task is in fact a misfire of a generally reliable probabilistic Bayesian algorithm— the reasoning on this explanation is deeply ingrained and subpersonal. For a more detailed discussion of these alternative possibilities, please see section 3.1 of the first chapter.

For present purposes, I do not need to resolve whether the common incorrect response to the Wason task is best explained as an associative or Bayesian inference. I

also do not need to resolve whether all instances of implicit deduction can be explained as associative, as Bayesian, or whether there is some preponderance of one versus the other in such cases. It is only necessary for my purposes to demonstrate that neither associative nor Bayesian inferences would rely on self-revisable rules in the relevant sense. Because they do not rely on this type of rule, they are Type 1 inferences.

Before I consider the revisability of the rule-following present in these deductive cases, I would like to also incorporate everyday inductive reasoning, as discussed in the previous chapter. Everyday inductive reasoning occurs almost constantly: Hume's discussion of billiard balls makes clear that we expect certain causal principles to obtain very frequently in our daily lives. Anytime we make a prediction, a generalization, or draw an analogy, we are relying on inductive inference, regardless of whether or not this inference is good.

As with deductive reasoning, inductive reasoning may rely largely on both Bayesian and associative reasoning. Associative reasoning occurs when a certain effect is observed as following from a certain cause a sufficient number of times that one can be inferred from the other— association is what Hume had in mind when discussing the constant conjunction of cause and effect. Bayesian explanations of inductive inferences, on the other hand, describe a subject's ability to predict or generalize based on a limited set of prior experiences not in terms of the subjects' constant conjunction or correlated phenomena, but rather in terms of subjectively assessed probabilities of certain outcomes, given prior knowledge. In other words, Bayesianism suggests that probabilistic algorithms— either innate or developed at a very young age— allow subjects to adjust the estimated likelihood of certain conclusions as new evidence is gathered. Bayesianism

was discussed at greater length in Chapter 1, in terms of Oaksford and Chater's explanation of apparently poor performance on the Wason selection task.

At this point, it should become fairly clear that the "rules" guiding associative or Bayesian inferences— whether inductive or deductive— are not self-revisable in the relevant sense. First, suppose that Oaksford and Chater are correct, and poor performance on the Wason selection task is in fact explained by subjects' relying on generally reliable Bayesian principles. When the subject responds incorrectly to the Wason prompt, what happens? They are not able to revise the Bayesian principles they relied upon originally; in fact, these Bayesian principles are probably unknown to the overwhelming majority of subjects, and it seems doubtful that even those subjects who understand Bayesianism would be able to alter their deeply ingrained computations. Anecdotally, it seems that many subjects may continue to feel the "pull" of their incorrect answer to the Wason prompt even if they fully comprehend and endorse a correct deductive answer.[7] Bayesian inference rules are not revisable in the relevant sense, and as a result, these inferences are Type 1.

Say, on the other hand, that Evans' associative explanation for incorrect performance on the Wason selection task: perhaps subjects are simply used to "matching" certain words from the prompt when inferring an answer— at least when they find the question challenging and cannot see an answer otherwise. It does not seem that an associative rule governing an inference of this type could be revised. Associations can only be counter-conditioned in the long-term— by having a large number of experiences in which the two phenomena are not correlated. If any inferences, whether inductive or

[7] In the next section, I will explain this "pull" in terms of the assent affect, which will follow directly from my explanation of the triggering condition of Type 2 reasoning.

deductive, rely on associative "rules," they are not revisable in the relevant sense, and they are Type 1 inferences.

Eric Mandelbaum has recently argued that implicit biases— like the matching bias that figures into Evans' explanation of the Wason task— are not associative, but rather are propositionally structured and are responsive to logical rules (2013). In one example he discusses, white subjects were given implicit association tests to measure their racial biases. Subjects were split into two groups: those who demonstrated racial bias towards African-Americans and those who did not. Each of these sub-groups were again divided: half of each group was told that the extent of their prejudice, either high or low, was in accordance with their peers, and half was told that their prejudice was atypical. They were then sent into a room where an African-American individual was waiting, and the closeness of the seat the subject chose to this individual was measured. Low-prejudice individuals who were told they were atypical sat further away than did low-prejudice individuals who were told they were typical in their peer group. High-prejudice individuals who were told they were atypical sat closer than did high-prejudice individuals who were told they were typical in their peers. Mandelbaum interprets the findings as demonstrating that the subjects' implicit bias had been modulated by intervention (2013, 20).

This result should not threaten my self-revisability based account. Mandelbaum's claim based on this study is that certain basic— apparently Type 1— inferences can be modulated by logical intervention. My claim is that Type 1 does not involve the manipulation of rules that are self-revisable. This draws out an important distinction between self-revisability and mutability. A rule may be propositionally structured and

responsive to logical information, but this rule may still not be self-revisable in the relevant sense to constitute the basis of a Type 2 inference. The subjects in the study above are not able to reflect on the rules that have been modulated, which is necessary for the rule to count as self-revisable, on my view. The rule involved in a Type 2 inference need not occur to the subject consciously when it occurs, but it must be available to the subject post hoc. To state matters even more clearly, a rule counts as self-revisable only if a subject is able to identify the rule upon which they were veridically relying after they have made the inference.

Gray area cases and automatization

There are many cases in which subjects rely on revisable rules implicitly. As discussed just above, implicit reliance on self-revisable rules still constitutes Type 2 inference, as long as subjects are able to post hoc invoke and revise the rule that they veridically relied upon. Imagine someone performing a proof in first-order logic with the help of a tutor. The student writes down a new line in her proof, and the tutor prompts her to reexamine: "what rule were you relying on there?" Suppose that the subject had allowed her attention to lapse, and was relying on the deductive rule implicitly. Though her inference was bad, it was still an inference, and she is able to explain the rule she was relying on. Perhaps she will say something like "I was affirming the consequent, but I remember now that that is not a valid inference rule." She might than explain what she ought to have done instead. In this example and many like it, the subject was relying on Type 2 reasoning even when she made her original implicit inference: she was relying on a revisable rule. Once her attention was drawn to her mistake, she explained the rule she

had veridically been relying upon in reaching her inference. That is, she genuinely was relying on affirming the consequent; this is not a case in which she is post hoc rationalizing her bad inference by identifying a new rule that could have justified it. Bad inferences are still inferences, and this is clearly a case of implicit Type 2 inference.

Suppose the logic student realized on her own that she made a mistake. Recalling the discussion of the assent affect in Chapter 2, we can explain what is going on here in some detail. Perhaps the logic student's Type 2 inference failed to be accompanied by the appropriate level of the assent affect. The absence of the assent affect, or perhaps the feeling of that absence— something like "doubt"— will trigger Type 2 processes to reflect on the rule that led to that judgment, either modifying it or altering its priority relative to other rules. Until a Type 2 process yields the assent affect, it can self-revise and reapply cyclically. Like Type 1 processes, when Type 2 processes are satisfied, they are accompanied by the assent affect, which halts the process of self-revision.

This view on the role of doubt appears to be shared by De Sousa and Peirce as well. As discussed in Chapter 1, section 5, both De Sousa and Peirce regard doubt as an instigator of further inquiry. The "further inquiry" to which they refer may, on my view, often or always be constituted by the triggering of Type 2 reasoning.

Many dual process models have difficulty explaining another type of gray area case: those in which a subject becomes so familiar with some certain category of paradigmatically Type 2 reasoning tasks that such inferences begins to lose some of the classic features of Type 2 reasoning. That is, an inference that would be slow, deliberate, explicit, and effortful for most subjects becomes relatively fast, implicit, and effortless

for a subject who has made very many similar inferences before. I will refer to this phenomenon as automatization.

The best discussion of automatization I have found comes from the developmental psychologist Allison Gopnik, in her writing about first- and third-person belief ascription. Gopnik's view, in very broad strokes, is that self-knowledge occurs via indirect inference, rather than direct access. That is, we must infer (many of) our own mental states in roughly the same way that we infer the mental states of others— based on behavioral clues, etc.[8] At first pass, Gopnik's view seems to contradict the intuitive notion of first-personal authority: we generally take ourselves to have privileged access to our own states. Gopnik argues, however, that well-rehearsed inference types can begin to appear psychologically immediate to the subject, even when they are applied in novel contexts.

Gopnik describes a phenomenon that she terms the "illusion of expertise" (Gopnik 1993: 335). She notes that in some cases, experts believe that they are "perceiving" a phenomenon directly, when in fact they are inferring it indirectly based on a robust theoretical framework. This would apply to golfers who account for a good day on the course by saying they were "seeing the greens well" or a doctor who "sees" cancer in a patient before a full battery of tests has confirmed the diagnosis. In each of these cases, the expert is indeed perceiving something— the grain of the turf, the patient's

[8] Gopnik construes her view as claiming that subjects "infer" their own beliefs as opposed to "directly accessing" them, as other views would have it. While it does seem to me that the type of inference on which Gopnik's account focuses would qualify as inference on my view of the category, it is a tricky question whether or not the "direct access" central to traditional accounts of self-knoweldge from which she distinguishes her view would also qualify as inference. In other words, I am not sure that I would describe Gopnik's unique view as unique in virtue of its including inferences. This question, while interesting, is beyond my scope here. I am merely interested in her account of how any inference might become automatized.

physiognomy— but they are actually indirectly inferring the phenomenon— the putt's gradual break to the left, the cancer— based on their background knowledge and experience. I want to leave aside Gopnik's highly controversial conclusion that something like this is involved in self-knowledge entirely; her discussion of automatization, divorced from the self-knowledge debate, is instructive for present purposes.

Consider the example of the golfer in terms of Type 1 and Type 2 reasoning. The golfer learned how to read greens early in her career: either she was taught certain rules or she learned them herself or some combination. Early in her career, when practicing putting, she would test out certain rules guiding her putting. "Because of x and y features of the species, grain, and moisture of the grass, the putt is likely to require z amount of force." She spends significant time and mental effort considering these rules before each putt. These rules are revisable, in that she learns to trust some more than others and refines them as she proceeds. She is making Type 2 inferences. Over time, her pre-putt inferences become faster and less effortful. In Gopnik's terms, she becomes an expert. Her inferences become psychologically immediate: it appears to her that she infers quite a bit about the putt as soon as she looks at the position of her ball on the green.

The golfer's inferences, based on rules about how certain types of putts might roll, have become automatized. Are they still Type 2 inferences, or have they become Type 1 inferences? My view offers a very natural explanation of these types of cases: they are Type 2 inferences if she finds the relevant rules to be revisable, and they are Type 1 inferences if they are not revisable. Perhaps, at first, she rapidly infers that a putt will be relatively fast: she observes that the green is dry and the hole is downhill from her

ball. This time, though, her caddy points out that the golfer's putts on this course have been rolling more slowly than expected. With this confounding factor pointed out, she might examine her inference more carefully. Perhaps she can identify the specific inference rules that veridically led her to believe her putt would roll quickly (rather than post hoc rationalized rules that she uses to justify her original intuition). It is an empirical question whether or not these inference rules will be revisable. She may be able to edit these rules in light of new information, along the same lines of the logic student who edits her proof strategy discussed earlier in this section. She may, on the other hand, find that her original inference continues to have pull even if she rejects it in favor of a difference based on new, better rules: that is, she may not be able to revise the rules governing her inference even if she rejects them. This "having pull" consists of the presence of the assent affect, as discussed in the previous chapter.

Whether or not an expert in any field would be able to revise inference rules that have become automatized strikes me as a case-by-case question. If the expert is able to revise the rules governing automatized inferences, the automatized inference is Type 2. If she is not, the automatized inference is Type 1. While it may become more or less difficult for subjects to identify and revise the rules that were veridically involved in their inferences, I believe there will still be a fact of the matter about whether or not they have identified and revised the exact rule upon which they relied. That is, the dual process distinction will remain a dichotomy, rather than a spectrum, even in these difficult automatization cases.

4. The big picture

Implicit within this thesis has been the underlying view that it is near impossible to account for our reasoning activity without incorporating the intellectual emotions, and that many recent cognitive scientists and philosophers have made their own lives more difficult by not doing so. I have argued, in fact, that it is impossible to explain fully how our reasoning mechanisms function without invoking some discussion of the feelings that are centrally involved.

Emotions get a bad rap when it comes to thinking: “be less emotional” and “be more rational” are frequently taken as synonymous imperatives. While there surely are cases in which emotions cloud a subject’s rational judgment, I hope to have demonstrated that certain intellectual emotions are crucial to the parsimonious operation of our reasoning systems. This dissertation has not offered a full cataloguing of the relevance of intellectual emotions to our reasoning activity. On the contrary, I have attempted to be very modest in my claims about the assent affect, as I intend it to be compatible with many possible connection points in further research. Still, by incorporating a spare version of a single emotional feature of our reasoning, I have been able to fill in gaps in previous accounts of inference and of the structure of our reasoning systems. I suspect that further incorporation of intellectual emotions into the cognitive science of reasoning will be similarly fruitful.

Bibliography

Apperly, I. and Butterfill, S (2009). "Do Humans Have Two Systems to Track Beliefs and Belief-like States?" *Psychological Review* 116.4.

Ariew, Roger, John Cottingham, and Tom Sorell (1998). Descartes' Meditations: Background Source Materials. Cambridge: Cambridge UP.

Boghossian, Paul (1992). "Externalism and Inference*." Philosophical Issues* 2, 11.

Boghossian, Paul (2014). "What Is Inference?" *Philosophical Studies* 169.1, 1-18.

Broome, John (2014). "Comments on Boghossian." *Philosophical Studies* 169.1, 19-25.

Broome, John (2013). "Practical Reasoning and Inference." Thinking about Reasons: Themes from the Philosophy of Jonathan Dancy. Ed. David Bakhurst, Margaret Olivia. Little, and Brad Hooker. Oxford, United Kingdom: Oxford UP, 2013.

Broome, John (2014). *Rationality through reasoning*. Sussex: Wiley Blackwell.

Carruthers, Peter (2009). "Systems and Levels: Dual-system Theories and the Personal-Subpersonal Distinction." *In Two Minds: Dual Processes and beyond*. Ed. Jonathan St. B. T. Evans. Oxford: Oxford UP.

De Sousa, Ronald (1987). *The Rationality of Emotion.* Cambridge, MA: MIT.

De Sousa, Ronald (2013). “Reasoning and Emotion, in the light of the Dual Processing Model of Cognition.” Unpublished. http://homes.chass.utoronto.ca/~sousa/

Evans, Jonathan StBT., and David E. Over (1996). "Rationality in the Selection Task: Epistemic Utility versus Uncertainty Reduction." *Psychological Review* 103.2, 356-63.

Evans, JStBT., Barston, J. L., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11, 295–306.

Evans, JStBT (1977). “Toward a statistical theory of reasoning.” *Quarterly Journal of Experimental Psychology*, 29, 297-306.

Evans JStBT (1989). Bias in human reasoning: Causes and consequences. Erlbaum, Brighton.

Evans, JStBT (2006). “On the resolution of conflict in dual process theories of reasoning.” *Thinking & Reasoning*, 13 (4) 321-339.

Evans, JStBT (2009). “How many dual-process theories do we need? One, two, or many?” *In Two Minds: Dual Processes and beyond.* Ed. JStBT Evans and Keith Frankish. Oxford: Oxford UP.

Evans, JStBT., and Stanovich, K. E. (2013.) Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science* 8(3), 223–241.

Fodor, J. A., & Pylyshyn, Z. W. (1988). “Connectionism and cognitive architecture: A critical analysis.” *Cognition*, 28, 3-71.

Frankish, Keith, (2009). "Systems and Levels: Dual-system Theories and the Personal-Subpersonal Distinction." *In Two Minds: Dual Processes and beyond*. Ed. Jonathan St. B. T. Evans. Oxford: Oxford UP.

Frege, Gottlob, and Michael Beaney (1997). *The Frege Reader*. Oxford, UK: Blackwell.

Frege, Gottlob (1979). *Posthumous Writings*. Trans. Hans Hermes, Friedrich Kambartel, and Friedrich Kaulbach. Chicago: University of Chicago.

Gendler, Tamar Szabó (2010). "Alief and Belief." Intuition, Imagination, and Philosophical Methodology, 255-81.

Goldman, A.I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*, Oxford: Oxford University Press.

Gopnik, Alison (1993). "How We Know Our Minds: The Illusion of First-person Knowledge of Intentionality." *Behavioral and Brain Sciences* 16.01.

Greene, J. (2008). "The Secret Joke of Kant's Soul", in *Moral Psychology: The Neuroscience of Morality*, ed. Walter Sinnott-Armstrong. Cambridge, MA: MIT Press. Vol. 3, pp. 35-79.

Haidt, J (2001). "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, 814-834.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review* 74(1), 88–95.

Harman, G. (1986). *Change in view*. Cambridge, MA: MIT Press.

Hume, David, and Tom L. Beauchamp (1998). An Enquiry concerning the Principles of Morals. Oxford: Oxford UP.

James, W. (1950). *The principles of psychology*. New York: Dover. (Original published 1890)

Kahneman, Daniel (2011). *Thinking, Fast and Slow*. D. Kahneman, P. Slovic, and A. Tversky. New York; Toronto: Farrar, Straus and Giroux; Doubleday Canada.

Kahneman, D., & Frederick, S. (2001). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 49–81). Cambridge, MA: Cambridge University Press.

Klaczynski, P. A. (2000). "Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent

cognition.” *Child Development*, 71, 1347 – 1366.

Kripke, Saul A (1982). Wittgenstein on Rules and Private Language: An Elementary Exposition. Cambridge, MA: Harvard UP.

Malcolm, N (1973). “Thoughtless brutes.” *Proceedings and Addresses of the American Philosophical Association*, 1972-73, 46, 5-20.

Mandelbaum, E (2014). “Attitude, Inference, Association: On the Propositional Structure of Implicit Bias. Behavioral and Brain Sciences, forthcoming.

Mandelbaum, Eric, and Jake Quilty-Dunn (2015). "Believing without Reason." *The Harvard Review of Philosophy* 22, 42-52.

Mandelbaum, Eric, and Jake Quilty-Dunn. "Inferential Transitions." Unpublished.

Mele, A (2001). *Self-deception Unmasked.* Princeton, NJ: Princeton UP.

Oaksford, M and Chater, N (1994). “A rational analysis of the selection task as optimal data selection.” *Psychological Review*, 101. 608-631.

Ormerod, R. J. (2009). "Rational Inference: Deductive, Inductive and Probabilistic Thinking." *Journal of the Operational Research Society*, 1207-1223.

Peirce, C. S. (1877). “The Fixation of Belief.” *Popular Science Monthly*, 12. 1-15.

Saunders, Leland (2009). “Reason and intuition in the moral life: A dual-process account of moral justification.” *In Two Minds: Dual Processes and beyond*. Ed. JStBT Evans and Keith Frankish. Oxford: Oxford UP.

Samuels, Richard (2009). "The magincal number two, plus or minus: dual process theory as a theory of cognitive kinds." *In Two Minds: Dual Processes and beyond*. Ed. Jonathan St. B. T. Evans. Oxford: Oxford UP.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin,* 119, 3–22.

Smithies, Declan (2013). "The Nature of Cognitive Phenomenology." Philosophy Compass 8.8 (2013): 744-54.

Smithies, Declan (2013). "The Significance of Cognitive Phenomenology." Philosophy Compass 8.8, 731-43.

Stanovich, KE (2009). “Distinguishing the reflective, algorithmic, and autonomous minds” *In Two Minds: Dual Processes and beyond*. Ed. JStBT Evans and Keith Frankish. Oxford: Oxford UP.

Stanovich, KE (1999). Who is rational? Studies of individual differences in reasoning. Lawrence Elrbaum Associates, Mahway, NJ.

Stanovich, KE (1990). “Concepts in developmental theories of reading skill: Cognitive resources, automaticity and modularity.” *Developmental Review*, 10, 72–100.

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

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman. *Heuristics and Biases: The Psychology of Intuitive Judgment.* Cambridge: Cambridge University Press.

Wittgenstein, Ludwig (2006). The Wittgenstein Reader. Ed. Anthony Kenny. Malden, MA: Blackwell Pub.

Wright, Crispin (2007). "Rule-Following without Reasons: Wittgenstein's Quietism and the Constitutive Question." Wittgenstein and Reason , 123-44.

Wright, Crispin (2012). "Comment on Paul Boghossian, “What Is Inference”."

Philosophical Studies 169.1, 27-37.