UNIVERSITY OF SAN DIEGO CALIFORNIA

Nonconceptual Metacognition

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Philosophy and Cognitive Science

by

Nathaniel Greely

Committee in charge:

Professor Matthew Fulkerson, Chair
Professor David Barner
Professor Jonathan Cohen
Professor Rick Grush
Professor Eric Schwitzgebel

2023

The Dissertation of Nathaniel Greely is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego
2023

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

Chapter 6, in part, contains material that has been published. The dissertation author is the sole author of this paper.

VITA

1996 BA, Communications, University of Southern California

2014 MA, Philosophy, California State University Los Angeles

2023 PhD, Philosophy and Cognitive Science, University of California San Diego

PUBLICATIONS

"Higher-order Theories of Consciousness are Empirically False," Journal of Consciousness Studies, Vol. 27, pp. 30-54, 2020.

"Epistemic Feelings, Metacognition, and the Lima Problem," Synthese, Vol. 199, pp. 6803-6825, 2021.

FIELDS OF STUDY

Major Field: Philosophy and Cognitive Science

ABSTRACT OF THE DISSERTATION

Nonconceptual Metacognition

by

Nathaniel Greely

Doctor of Philosophy in Philosophy and Cognitive Science

University of California San Diego 2023

Professor Matthew Fulkerson, Chair

A metacognitive mental state or process takes another mental state or process as its intentional object. Reality testing is the ability to distinguish one's own perceptual mental states from imagination or episodic memory. I argue that reality testing is a metacognitive ability and that creatures that lack mental state concepts can perform reality testing. This entails that reality testing is a nonconceptual metacognitive ability. I offer a novel account reality of testing in terms of sensorimotor contingencies. This explains how reality testing is accomplished without the use of mental state concepts.

Introduction

I want a burger right now. I also believe that my desire for a burger should be resisted – burgers are high in cholesterol, produce greenhouse emissions and animal suffering, and cost more than eating in. This belief is a metacognitive state. Call it $Belief_M$:

$Belief_M$: I should resist my desire for a burger.

Like any other mental state, $Belief_M$ is about something. This property of representing or "being about" some state of affairs is a central component of mentality. Some argue it is the mark of the mental.

What is interesting about $Belief_M$ is that its intentional object is another mental state – my desire for a burger. Thought about thought is known as *metacognition*. Metacognition is taken by most philosophers and psychologists to be a sophisticated type of mental state. This is because it is largely assumed that in order to entertain metacognitive states we must possess mental state concepts, and that possessing mental state concepts involves a more or less complete theory of mind. A smaller but substantial group of philosophers and psychologists claim that in order to engage in metacognition about one's own mental states, we must first be able to think about other people's mental states, an ability termed *mindreading*. This latter view, taken to the extreme, is the "opacity" of mind thesis. On this view we lack access to the bulk of our own mental states and must learn about them using many of the same processes that we use to learn about the mental states of others, by observing our own behavior and making inferences. Other theorists claim that metacognitive states are the result of mental operations applied to first order

states. If eating burgers is wrong, then I can infer that any desires to eat burgers should be resisted. No introspection required – I look to the world to make inferences about my own mind.

While such views are popular in contemporary psychology and philosophy of mind, they come with significant caveats. We do, and indeed we must, have direct access to our own sensory states. I don't need to observe my own behavior to determine whether I am in pain, or whether I am smelling roses. And I can't know the contents of my own imagination by "looking to the world." These caveats, however, have huge and largely ignored implications for the study of metacognition. Unfortunately, the study of such basic forms of metacognition is underdeveloped, largely consisting of brief asides in accounts of mindreading. In this essay I hope to begin to correct this omission with a study of what I will argue is the most fundamental form of metacognition – *reality testing*.[1]

Reality testing is the ability to distinguish one's own perceptual states from imagination and episodic memory. Humans exercise this ability almost constantly. We entertain auditory experiences like the "inner voice" without confusing them with external speech. We recall images of late relatives without mistaking those images for the dead come to life. We largely live our lives with one foot in each of the two realms of external and internal experience. And while reality testing occasionally fails and we mistake an imagined sound for a perceived one, or vice versa, for neurotypical adults this is the exception rather than the rule.

Reality testing is pervasive and largely automatic. But determining whether it requires the sorts of concepts and inferences assumed by the dominant theories of metacognition will involve empirical investigation. Any creature capable of both perception and imagination or episodic memory must perform reality testing if it is to survive for long. Perception and episodic memory

[1] I did not coin the term 'reality testing'. To "test" reality may connote something conscious and sophisticated, but as we shall see reality testing is largely unconscious and primitive.

have been documented in creatures to which researchers do not typically attribute metacognitive abilities. But if these creatures are capable of distinguishing perceptions from memories, then they are capable of reality testing, and this pushes metacognition further down the phylogenetic scale than had been previously believed. Furthermore, if creatures who lack mental state concepts employ reality testing, then the ability is nonconceptual, in which case there is such a thing as *nonconceptual* metacognition.

This empirical argument that reality testing is nonconceptual can also be supported philosophically. Philosophical accounts of the necessary conditions for conceptual thought implicitly presuppose reality testing in their accounts of the mind/world distinction (or *objectivity*). Although extant accounts of objectivity differ in many respects, all presuppose reality testing. If so, reality testing is a form of nonconceptual metacognition that is fundamental to human cognition in its conceptual form – to states like belief and desire. In a sense, there is no cognition proper without metacognition.

Extant accounts of reality testing typically serve as supplements to more general accounts of metacognition. As such, they are often underdeveloped and collapse under scrutiny. Many are couched at the propositional level, and thus cannot account for the nonconceptual nature of reality testing, and even those that allow for nonconceptual interpretations have serious problems. I will offer an account of reality testing that does justice to the complexity of this metacognitive ability while keeping it at the nonconceptual level. This can be accomplished if we conceive of reality testing as a sensorimotor ability. Sensorimotor accounts of mental content are not new, but they have not been employed in accounts of metacognition. Reality testing is one example of a broader ability humans have of distinguishing the various attitudes we take toward the same content. Here I will focus on the attitudes of perceiving, imagining, and

remembering, but in doing so I hope to establish a framework for understanding the way we sort our mental content into attitudes that could be extended by future research. The structure of the essay is as follows.

In chapter 1 I will introduce the phenomenon of reality testing and argue that it is a metacognitive ability. There is little agreement in the literature as to what constitutes a rigorous definition of metacognition, so I will begin by surveying the range of phenomena that, in practice, serves as the object of metacognition research. I will then construct a definition that is at least extensionally adequate – an improvement on extant definitions in the literature. The phenomena studied as metacognition can be captured by considering two dimensions – the object of the metacognitive state and the format of the metacognitive states. The result is four general types of metacognition that ought to be covered by an adequate definition of the term. These include "mindreading" – mental states about the mental states of others, introspection – mental states about our own mental states, propositional metacognition, and metacognitive feelings. The resulting definition of metacognition will be rather broad, including any form of higher-order representation – i.e., representation of a representation. I will then argue that reality testing must be accomplished using higher-order representations. Finally, I will survey some extant, and more restrictive, definitions of metacognition that have been proposed in the literature. I will argue that even on these definitions, reality testing qualifies as metacognitive.

In chapters 2 and 3 I will argue that reality testing is nonconceptual. This argument will take two forms. In chapter 2 I will use empirical evidence from comparative psychology, cognitive ethology, and developmental psychology to argue that some animals that lack mental state concepts do engage in reality testing. Reality testing has been studied explicitly in rodents. I will explain the relevant studies and argue that they are properly interpreted as providing

evidence of reality testing in rats and mice. I will then argue that rats and mice do not possess mental state concepts. This entails that these creatures' reality testing ability is nonconceptual. I will then expand my argument to corvids and human infants. While there are no explicit studies of reality testing in these creatures, I will show that in each case they possess both perceptual abilities and imagination or episodic memory while lacking the relevant concepts. The existence of both faculties without radical confusion between them implies a reality testing ability.

In chapter 3 I will argue that reality testing is nonconceptual using a primarily a priori, philosophical approach. I will survey a variety of philosophical and psychological accounts of the nonconceptual/conceptual content distinction, including work by P.F. Strawson, Gareth Evans, Adrian Cussins, J. L. Bermudez, John Campbell, Michael Dummett, John O' Keefe, Rick Grush, and Tyler Burge. These accounts largely agree that there are certain necessary conditions for the possession of conceptual content, primary among them the ability to conceive of mind-independent objects (Burge being one exception). But in all cases these authors hold that the ability to entertain conceptual content requires faculties of perception and imagination or episodic memory. This, I will argue, implicitly requires a reality testing ability. I will show that in each case these authors' accounts of the necessary conditions for conceptual content implicitly rely on a nonconceptual ability for reality testing. This entails that reality testing is not only nonconceptual but is a necessary condition for conceptual content, and thus a necessary condition for prototypical cognitive states like belief. Thus, in a sense, metacognition is prior to cognition proper.

In chapter 4 I will examine extant accounts of reality testing and argue that they fail. My task is simplified by having already argued that reality testing is nonconceptual, as this rules out accounts that require the use of mental state concepts. I will instead focus on accounts that allow

for nonconceptual reality testing. I will examine and refute Humean "vividness" accounts, on which imagination and episodic memory are distinguished from perception by their "degraded" appearance. Pitt's account of cognitive phenomenology implies that any mental attitude is accompanied by a distinctive, individuating "feel," which might be extended to account for reality testing. The account is similar in some ways to Bermudez's proposal that a "feeling of reality" distinguishes perception from memory. I will argue that these accounts of reality testing beg the question – whatever faculty reliably produces the appropriate phenomenology must itself perform reality testing. Nichols and Stich propose a functional account of reality testing that largely addresses propositional forms of the ability. But their account might be adapted to a purely architectural account of reality testing which dissolves the problem by keeping perception and other states in distinct functional channels. I will argue that such an account is inconsistent with the empirical evidence. Goldman offers an account of reality testing which makes use of the neural properties of perceptual and imaginary states, and I will argue that this account fails. Finally, I will consider McGinn's claim that what distinguishes imagination from perception is that the former is responsive to the will. This account also fails, as there are obvious exceptions. Having argued that no extant account of reality testing is sufficient to explain the ability, I will then offer my own account of reality testing.

In chapter 5 I will develop an account of reality testing that succeeds where other accounts fail. My account holds that reality testing distinguishes perceptual content from mental imagery on the basis of the sensorimotor contingencies associated with each. By "sensorimotor contingency" I mean the sort of sensory experience that can be expected given a particular movement. I take "sensory experience" to include imagined or remembered imagery.[2] An

[2] I will also take 'motor' to potentially include mental actions.

afterimage, for example, follows my gaze when I move my head.[3] Perceived objects typically do not. I will begin the chapter by surveying prominent sensorimotor accounts of perception in order to elucidate the concept of the sensorimotor contingency and show that exploiting these does not require conceptual content. I will focus on Cussins' account of the sensorimotor "feature space" in particular as a model for the two "worlds" of perception and imagination/memory into which we constantly sort our sensations. I will then consider two sensorimotor accounts of imagination. While these accounts in some ways presage some of my points about the nature of reality testing, the extant accounts by Ulrich Neisser and Nigel Thomas, I will argue, are ultimately implausible. I will then elaborate my own sensorimotor account of reality testing one modality at a time. I will begin with audition, enumerating several well-established cues the auditory system uses for spatial location. When these cues are ambiguous between inner and outer experiences, failures of reality testing may occur. Such ambiguity is so frequent, however, that context plays an important role. I will use Bregman's framework of higher-order schemata in audition, which maps onto Cussins' notion of the feature space, as the basis of contextual sensitivity in auditory reality testing. I will then proceed in a similar manner with vision, olfaction, and smell. Each modality has its own sensorimotor contingencies for sorting experience into perceptual or imaginary feature space. My account will be necessarily preliminary but sufficient to provide a framework for further elaboration. In each case I will consider the predictions of my account and show that they are consistent with the evidence where other accounts fail. Finally, I will consider cases in which reality testing could be said to fail regularly – dreams, imaginary content in perception, and predictive processing. In each case I will argue that my account can make sense of the data.

[3] The afterimage is a handy example, whether or not in the end it qualifies as imagination.

In chapter 6 I will consider the case of epistemic feelings, which have been posited as a form of nonconceptual metacognition, most prominently by Joelle Proust. Epistemic feelings are commonly studied in memory research, and they include feelings of knowing and tip-of-the-tongue experiences. On her account of epistemic feelings, which goes by the name 'evaluativism' and is largely accepted by a group of theorists including Jerome Dokic, Asher Koriat, and Santiago Arango-Munoz, nonconceptual metacognition does exist. Evaluativism holds that epistemic feelings are nonconceptual affordances which respond not to first-order mental content (e.g., the memory one is trying to recall) but to cues and heuristics like reaction time which correlate with the presence of the relevant first-order content. On this account epistemic feelings tell us whether a memory affords recall or whether it is irretrievable. However, the evaluativist conception of nonconceptual metacognition is, I will argue, unmotivated, unsupported, ill-conceived, and generally impoverished. It is unmotivated because it is posited to solve supposed problems in the study of metacognition that can be solved by other accounts. It is unsupported because the empirical evidence marshalled in defense of the view is open to other, more plausible explanations. It is ill-conceived because the cues and heuristics to which the epistemic feelings are meant to respond just are, for the most part, first-order content. The evaluativist conception of nonconceptual metacognition is impoverished because it contains less information than conceptual content, turning the traditional conceptual/nonconceptual distinction on its head. I will then briefly sketch out how my account of nonconceptual metacognition might be extended to make sense of epistemic feelings.

My overall point is that little of what we consider cognition would be possible without metacognition. Higher-order representation is a fundamental mental ability, and we will never understand the mind without understanding the mind's ability to reflect on itself.

Chapter 1. Reality testing is metacognitive.

1.1 What is reality testing?

Consider the following two scenarios. Suppose that they occur on subsequent nights:

Night 1. You are lying in bed, trying to fall asleep. A song pops into your head, unbidden. It's quite vivid and keeps you from sleeping. But you don't get up and ask your roommate to turn down her stereo. You know the song is in your imagination. You never question that for a moment.

Night 2. You are lying in bed, trying to fall asleep. You hear a song. It is your roommate listening to her music loudly downstairs. You get up and ask her to turn it down and she complies. After you return to bed, you hear faint music – or so you think. Now you don't hear it. Now you do. Maybe you are imagining it. It would be embarrassing to get up again and start a confrontation if you are imagining it. Eventually you fall asleep.

The question you faced on night 2 was a metacognitive one. You were inquiring about your own mental states - was your experience perceptual or imaginary? The metacognitive nature of the problem was, in this case, conscious and propositional. You asked yourself whether you were perceiving or imagining music. You briefly believed that you were hearing it, subsequently doubted that you were hearing it, believed that you were imagining it, doubted that, and eventually gave up. On night 1, in contrast, you made the determination effortlessly, without

entertaining conscious beliefs or doubts on the subject. In fact, the great majority of similar determinations happen in this way. Night 2 is the aberration. How is this possible? You have, it seems, a remarkable ability to maintain two distinct realms of experience – that of perception and that of imagination or episodic memory. Vivid memories of deceased relatives arise without even momentary shock at their resurrection. You imagine various types of food you might like to eat, but the imagined delectability of the imagined pizza never tempts you to try and take a bite. This ability is called *reality testing* (Arlow, 1969; Hohwy & Rosenberg, 2005; M. K. Johnson & Raye, 1981; Kim & Koh, 2016; M. A. McDannald et al., 2011; M. McDannald & Schoenbaum, 2009).

It is uncontroversial that the explicit form of reality testing that occurred on night 2 is metacognitive. Beliefs, doubts, and other propositional mental states which take other mental states as their objects are metacognitive. When you believe that you are imagining music, you engage in metacognition. But what about night 1? You apparently determined that the experience was imaginary, for your behavior seemed to indicate it. You made no attempt to get up and turn off the music. But if there was a process that made that determination it was not conscious. Was that form of reality testing propositional? Are conscious and unconscious reality testing the same type of process? I will argue that the ubiquitous, unconscious type of reality testing exercised on night 1 is also a metacognitive ability, but of a different sort.

The term ‘reality testing’ originates in the psychoanalytic literature (Arlow, 1969), but has made its way into philosophy (Hohwy & Rosenberg, 2005), cognitive psychology (M. K. Johnson & Raye, 1981), clinical psychology (Dagnall et al., 2018), and cognitive ethology (Kim & Koh, 2016; M. A. McDannald et al., 2011; M. McDannald & Schoenbaum, 2009). The term has been used in slightly different ways. Arlow defines it as “the ability to distinguish between

perceptions and ideas" (p. 28) but understands 'ideas' in a broadly empiricist sense as "repetitions" (p. 31) of perceptions. The ability does not entail any strong epistemological conclusions, but rather "emphasis is placed upon the differentiation between representations of what is external – of the object world – from representations of what is internal – of the self or of mental life" (p. 28). Reality testing on this construal refers to the way we sort our own representations into two classes - perception and mental imagery. This sorting may not always be entirely accurate. If, for example, the states that we typically class as perception are partly derived from memory or imagination, as is posited by predictive processing accounts of perception, then reality testing is systematically inaccurate. In its early psychoanalytic use, the term implies something like this. For the psychoanalyst, our everyday perceptions are constantly intertwined with and distorted by mental imagery or "fantasy." The task of the psychoanalyst is to help the patient develop her reality testing abilities and disentangle the two. The hallucinations that occur in schizophrenia are an extreme version of this phenomenon, and the notion of reality testing has been employed in more recent philosophical accounts of such delusions (Hohwy & Rosenberg, 2005). Reality testing remains a subscale of the Inventory of Personality Organization, a more general clinical assessment (Dagnall et al., 2018; Lenzenweger et al., 2001). As a clinical assessment the concept of reality testing has broadened Arlow's 'ideas' to include false beliefs as one of the internal factors that might distort our relationship to reality. My interest in this essay is closer to the earlier, empiricist form of reality testing, as I will be concerned with the reality testing ability of creatures that lack propositional mental states like belief. Recent studies of reality testing in rodents adopt this interpretation as well (Kim & Koh, 2016; M. A. McDannald et al., 2011; M. McDannald & Schoenbaum, 2009). Whether reality

testing for beliefs is a development of this same primitive ability or a distinct sort of ability is an interesting question that I will not be able to address here, but I hope to do so in future research.

I have lumped imagination and episodic memory together in a manner that some may find disconcerting, and others may find pleasing. The two are undeniably linked. When I call up a mental image, that image is often called from memory. If the image is novel – an eight-legged cat – the parts from which I fashion the image are called from memory. More controversially, some claim that the act of remembering *just is* an act of imagining, and that memory is reconstructed, not recalled (Michaelian, 2016). I will not take a stand on the relationship between imagination and episodic memory. For the purposes of this essay, then, when I refer to imagination or episodic memory, I truly mean to indicate a disjunction. Nor will I take any stand on whether and how a subject distinguishes her episodic memories from imagination. I maintain only that we frequently and systematically distinguish imagined or remembered content from perceived content. That is, insofar as episodic memory has imagistic content, we distinguish those images from perceptual images.

Reality testing has also been invoked in contrast to two related notions in the literature on memory – reality monitoring and source monitoring (M. K. Johnson & Raye, 1981). Reality monitoring is the ability to determine whether a memory originated in perception or imagination, and source monitoring is a more general ability to determine the origin of one's memories. The feeling of presence is a nearby concept (Matthen, 2010). It is the feeling that an object is nearby and has been invoked as one way that perception differs from imagination. On that construal, however, the feeling of presence represents a criterion that might be employed *in* reality testing, but it is not reality testing itself. Accounts of reality testing are also distinct from accounts of the metaphysical difference between perception and imagination, or of the metaphysical difference

in their contents (e.g., Nanay, 2015). An accurate metaphysics of perception and imagination will place constraints on reality testing. Insofar as reality testing is accurate, this must be because it tracks real properties of our mental states. But the fact that humans also fail at reality testing in systematic ways (for example when dreaming) suggests that reality testing does not track essential properties of perception and imagination.

But even if reality testing fails for large domains of experience, reality testing is still accurate in an important sense. Even if the states we class as perceptual consist to some greater or lesser degree of imagined or remembered content, there are other states that are wholly remembered or imagined, and these we typically do not mistake for perception. Neurotypical adults walk through life juggling two distinct realms of experience. We experience inner speech that we do not mistake for outer speech, songs run through our minds, but we do not attempt to turn down the stereo, we visualize what we want for lunch, but we do not attempt a bite. Our minds are truly awash in images that we only rarely mistake for perceptions. Whatever the metaphysics of perception, imagination, and memory, we manage to make subjective distinctions among these states that are largely adaptive.

I have had to borrow the term 'reality testing' from psychoanalytic and psychological literature because it is rarely treated by philosophers with the sort of emphasis that it should be. Typically, in the philosophical literature reality testing is studied as one case of a more general phenomenon – that of taking an attitude to one's own mental content. Accounts have been given of how we distinguish our beliefs from our desires, hopes, fears, and rest, and perception and imagination are often thrown in to boot. I will consider some of these accounts in chapter 4, but the conclusion of this essay will be that reality testing is a fundamental mental ability that precedes the others ontologically, developmentally, and ontogenetically.

I have offered one example of reality testing – the ability to distinguish an imagined song from a perceived song. I also described that situation as maintaining two distinct realms of *experience*, which suggests that the mental content involved is conscious. But this is not always the case. It is broadly believed, and there is much evidence to suggest, that there is unconscious perception. Subjects who report no conscious experience of a stimulus in the visual field can nonetheless make accurate judgments about the stimulus. Sometimes they deny any confidence in the accuracy of their judgment, insisting that it is a guess, but in other cases they express confidence in the judgment (e.g., Charles et al., 2013; Jachs et al., 2015). This suggests that we can not only perceive unconsciously, but that we can distinguish the content from imagination. This may suggest that reality testing extends to unconscious perception as well. I do wish to leave this possibility open.

In this chapter, I will argue that reality testing is a metacognitive ability. Some may find this claim obvious, but others may resist it. I will present several definitions of metacognition, some from the literature and one of my own, and argue that in each case reality testing qualifies as metacognitive. While it may seem obvious to some that reality testing is a metacognitive ability, it is important that I offer support for this claim, for it in turn supports claims in latter chapters that are not so intuitive.

1.2. What is metacognition?

Metacognition research in contemporary psychology began in studies of memory, typically dated to Hart's (1965) dissertation on feelings of knowing. The term 'metacognition' was coined by Flavell (1975) and the discipline came in to its own with Nelson and Narens' (1990) model of

metacognition and Nelson's (1992) edited volume on the subject. Philosophical study of metacognitive phenomena predates Hart, of course, typically as a strand of epistemological research. Descartes' cogito is metacognitive. Most recent philosophical studies of metacognition, however, occur in the territory established by psychology (one important exception being transparency theory, more of which later). There is no single, canonical definition of metacognition, and the attempts offered in the seminal texts are inadequate. After a brief survey of these attempts and their inadequacies, I will construct a definition of metacognition that is at least extensionally adequate. I will survey the range of phenomena currently studied under the name 'metacognition' along two main axes – the intentional object and the representational format of the metacognitive state – and construct a definition that covers all these cases. I will then argue that reality testing satisfies this definition. I will then consider Nelson and Narens' (1990) account of metacognition and argue that reality testing satisfies most of its conditions, and that where it fails the problem lies with Nelson and Narens' account itself. Reality testing is a self-directed form of metacognition – the higher- and lower-order mental states both belong to the same subject. Thus, it might be considered a form of introspection. Schwitzgebel (2014) offers a set of criteria that partially define introspection. If reality testing qualifies as a form of introspection, then it automatically qualifies as metacognition. I will survey these criteria and argue that reality testing meets all the criteria but one, and that this criterion would be question-begging in the current context.

In the early work on the subject, metacognition is defined in various ways, sometimes inconsistently, and always inadequately. Nelson (1992) calls metacognition "cognition about one's own cognition." This definition rules out mindreading – the ability to form accurate beliefs about the mental states of others - as a form of metacognition. While the term 'metacognition' is

typically restricted to self-directed cases, ‘mindreading’ being used for other-directed cases, this is not entailed by many of the other definitions offered. Metcalfe and Shimamura (1994) define metacognition as “knowing about knowing.” This does not restrict metacognition to self-directed cases but rules out many other cases if we take “knowing” to mean the same thing epistemologists mean by it. Metacognitive states might be false, unjustified, or simply not truth-apt (e.g., desiring that I desire to eat veggies). Flavell (1979) uses “cognition about cognitive phenomena” which is perhaps the most general, depending on how one defines ‘cognition’. Psychologists and cognitive scientists sometimes use ‘cognition’ as synonymous with ‘mental’. Philosophers often take ‘cognitive states’ to be propositional states like belief, leaving out sensation, mental imagery, and other types of mental states one might either target in metacognition, or which might be the vehicle of metacognition. I might, for example, believe that I smell cheese or see that you want a hug.

These early definitions typically fail to capture the full extent of metacognition research. They tend to do so along one of two dimensions – illegitimately restricting either the object of the metacognitive state or restricting the representational format of the higher- or lower-order state involved. A proper definition ought to capture all the phenomena studied as metacognition (or as near to all as is reasonable) and only those cases (again, within reason). I will provide examples of metacognitive research on each end of these two dimensions, providing a logical space that ought to be captured by any extensionally adequate definition of metacognition. I will then construct a definition of ‘metacognition’ that covers this space and argue that reality testing satisfies that definition. There may be other important dimensions of metacognition that are left out by this survey. In sections 1.5 and 1.6 I will consider some additional conditions on

metacognition proposed by Nelson and Narens (1990) and on introspection by Schwitzgebel (2014).

1.2.1 Propositional metacognition

Definitions of metacognition as "thought about thought" or "knowledge about knowledge" are not wholly inaccurate. Much metacognition research focuses on "propositional" mental states like belief, desire, thought, or knowledge. Propositional mental states are comprised of some propositional content and an attitude taken toward that content. Such states include believing that Johnny Marr is the greatest living guitar player or desiring that I will meet Johnny Marr. The attitudes are belief and desire, respectively, and the content is 'Johnny Marr is the greatest living guitar player' and 'I will meet Johnny Marr'. When the propositional content in question features a mental state, then the overall mental state is metacognitive. When I believe that I am imagining a song, the propositional content is 'I am imagining a song', where "imagining" is a mental state, and the attitude is belief. Other propositional attitudes include desire, thought, hope, fear, and many more. Propositional attitudes typically pick out types of mental states, so belief about a belief will be a form of metacognition.

Evidence from research in child development is taken to demonstrate the existence of propositional metacognition. It is often claimed that metacognitive abilities require a theory of mind - a set of folk-psychological generalizations couched in a propositional format. There is a marked contrast between children who lack such a theory and those who do. Studies suggest that older children better understand which memorization techniques are effective for them and that this underlies their greater memorization abilities (Flavell et al., 1970). Differences have also

been demonstrated between children who possess a theory of belief and those who do not (Gopnik, 1993). One demonstration that young children lack the requisite theory of mind to engage successfully in metacognition involves the false belief task. Children are shown a scenario in which a character sees an object of value. The object is then moved when the character is not present but in view of the child. Younger children who lack a concept of belief, and thus of false belief, make false predictions about where the character will look for the object. Importantly, this ability is symmetric between self and others. Children who lack a theory of mind cannot attribute false beliefs to themselves. Even perception is sometimes the subject of theoretical metacognition. Studies have shown, surprisingly, that many subjects believe that vision is accomplished by rays emitted from the eyes (Winer & Cottrell, 2004).

The benefit of the use of the propositional format in a theory of mind is that it explains our ability to make inferences and predictions about mental states. Given a theory of mind that explains actions in terms of belief-desire pairs, among other things, I can predict and explain others' actions if I can form beliefs about their beliefs and desires. If I believe that you desire to eat my candy bar, and I also believe that you believe that theft is an acceptable way to obtain a candy bar, I can infer that you will try to steal my candy bar and I can take steps to safeguard it. Researchers like Carruthers (2011) hold that a subject's reasoning about her own mental states occurs in a similar fashion. I must retroactively infer my own beliefs by observing my actions and applying a theory of mind.

There are also purely philosophical accounts of metacognition which make use of the propositional format. Transparency theories of metacognition, associated with Evans (1982) and Byrne (2018), hold that we form beliefs about our own beliefs by applying an "ascent routine" to first-order propositional content. If I want to know whether I believe that there is a cat on the

mat, I first determine whether there is a cat on the mat. I then apply the rule 'If p is the case, then I believe p' and conclude that I believe that there is a cat on the mat. Here too the inference is made possible by the propositional format.

1.2.2 Metacognitive feelings

Many philosophers argue that some mental content is not propositional. The color content of perception, for example, cannot always be expressed propositionally. Certain skills, like knowing when to swing at a pitch in baseball, are probably not stored and recalled as sets of propositions. Primitive, non-propositional forms of representation might also explain intelligent behavior in creatures who lack language. These claims are controversial, and I will have much to say about non-propositional mental content in chapters 2 and 3. Here I simply want to point out that metacognition researchers often study arguably non-propositional forms of metacognition.

'Non-propositional metacognition' is not established terminology, so its extension is largely up for grabs. At its most permissive, non-propositional metacognition would include any metacognitive state in which either the higher- or lower-order mental state is non-propositional. The latter variety certainly exists if non-propositional mental states exist. One can always form a metacognitive belief about 'that', where 'that' is some mental state that cannot be expressed propositionally – say the color content of a visual state. The former variety, however, is an area of active interest because of the controversy around the representational powers of non-propositional mental states. How could a non-propositional mental state take an attitude toward some mental content? This essay is largely concerned with developing a model for this type of metacognition, but there is already an area of metacognition research working in the same vein.

Phenomena like tip-of-the-tongue states, feelings of knowing, and feelings of confidence have been objects of metacognition research from its inception (J. T. Hart, 1965; T. O. Nelson & Narens, 1990). These states are known as 'epistemic feelings'. They are called 'epistemic' because they are involved in the recall and management of states like memory and belief – states that admit of truth or falsity. Not all epistemic feelings are metacognitive. DeSousa (2009) describes fear as an epistemic feeling directed at objectively dangerous states of the world. But many epistemic feelings are commonly conceived as metacognitive states (though there are dissenters, e.g., Carruthers, 2017). A tip-of-the-tongue experience tells me that a memory is there to be retrieved. A feeling of confidence tells me that a recent judgment was correct. These states have been shown to be accurate at a level greater than chance, which supports the metacognitive interpretation. Epistemic feelings provide genuine information about the content of your mind.

Epistemic feelings are called 'feelings' because they present as such. A tip-of-the-tongue experience is different than a simple judgment that I know some actor's name. It is a palpable, "phenomenal" experience of almost being able to say the name. Phenomenal states are mental states, like the color content of perception or the painfulness of pain, that have a qualitative "feel". The exact sense in which epistemic feelings are "feelings" is a matter of dispute, however. Some hold that epistemic feelings are emotional states (e.g., de Sousa, 2009). But if one adopts a cognitivist account of emotions, then emotions are just propositional states. Some explicitly claim that epistemic feelings are phenomenal states (e.g., Arango-Muñoz, 2011), which are often (though not always) taken to be different in kind from propositional states. Others claim that epistemic feelings can be unconscious, and thus probably not phenomenal, but explicitly claim that they are "nonconceptual" (Proust, 2013). Much more will be said on nonconceptual mental content, but for now we can say simply that propositional content is

composed of concepts, so for mental content to be nonconceptual is for it to be non-propositional.

We have seen that some accounts of metacognition require that metacognitive states exist in a propositional format, either to construct a theory of mind or to apply ascent routines. It is curious, then, how epistemic feelings manage to provide information about lower-order mental states and accomplish their metacognitive function. One view known as *evaluativism* holds that epistemic feelings help us monitor and evaluate the function of our own mental processes through simple heuristics that tend to correlate with successful recall, memorization, perception, etc. Proust compares epistemic feelings to Gibsonian affordances (more on these in Chapter 3) and characterizes their content as simple indications of "good" or "bad" prospects for "*A*-ing," where *A* is some mental process like remembering.[4] Other evaluativists include Santiago Arango-Munoz (2011), Jerome Dokic (2012), and Asher Koriat (2000), though they sometimes vary in their precise accounts of the representational format involved.

I am not an evaluativist, but I will argue for the existence of nonconceptual metacognition. My account of reality testing will develop a different way of explaining how non-propositional mental states perform metacognitive tasks. In chapter 6 I will take issue with the evaluativist account of epistemic feelings and will argue that my account of reality testing offers a better framework for understanding them. But in the present context, it is clear that non-propositional forms of metacognition are objects of metacognitive research and ought to be included in an extensionally adequate definition of 'metacognition'.

[4] Evaluativists sometimes employ idiosyncratic terminology, reserving the term 'metacognition' for epistemic feelings and referring to propositional metacognition as 'metarepresentation'. This is not only confusing, as nonconceptual mental states also represent, but it is not reflective of the more general state of the field, which includes propositional forms under the heading of 'metacognition'.

Other accounts of metacognition scattered through the literature might be interpreted as positing non-propositional metacognition as well. Inner-sense views of self-knowledge date back at least to Locke and hold that we learn about our own mental states through a quasi-perceptual mechanism. Insofar as the content of perception is sometimes characterized as nonconceptual, inner-sense can be interpreted as a form of non-propositional metacognition. When I am aware that I am imagining a pink elephant, the color content of my inner perceptual state may well be nonconceptual. Gallagher (2004) suggests that there is nonconceptual mindreading, though he does not develop the idea.

There are two forms of non-propositional metacognition that are active areas of study. One consists of mental states whose intentional objects are non-propositional states. This form is only as controversial as the existence of non-propositional mental states in general. The other form involves a non-propositional mental state taking another mental state as its intentional object. While this is more controversial, it is an active, and indeed foundational, area of metacognitive research. An extensionally adequate definition of metacognition ought to make room for both forms.

1.2.3 Mindreading

Mindreading is metacognition directed at mental states other than one's own. Prominent accounts include theory-theory (Gopnik, 1993) and simulation theory (Goldman, 2006; Gordon, 1986). Theory-theory, as we have seen, is a propositional account of metacognition. The inferences facilitated by that propositional format allow us to form beliefs about the content of other people's minds. Thus, if a friend ignores me at a party, I can use my theory of mind, along

with the fact that I still owe her money, to deduce that she is angry at me. Broken promises tend to cause anger and anger tends to cause snubbing behavior.

Simulation theory holds that I come to understand the mental states of others by running an offline simulation in my own mind. By putting myself in my friend's shoes, I might become angry myself, and conclude that she feels angry. While there may be inference involved at some point in the process, much of the knowledge is gained by directly experiencing the simulated state. This experience may be partly feeling-based, and thus non-propositional.

1.2.4. Introspection

We have already encountered accounts of self-directed metacognition in our discussion of representational format. This form of metacognition is often called 'introspection' and theories about its nature predate the use of the term 'metacognition' by hundreds if not thousands of years. But it certainly a subject of contemporary metacognition research as well.

Theory-theorists apply their account of mindreading to the reading of one's own mind as well, claiming that we come to know our own mental states by inferring them from our observed behavior along with a theory of mind. I have also discussed transparency theory, on which we gain knowledge about our own mental states by applying ascent routines to first-order propositions – if p is the case, then I believe that p. As a philosophical theory it less commonly referred to as an account of 'metacognition', but it is undeniably a propositional account of how we form beliefs about our own mental states. Inner-sense accounts of introspection posit a quasi-perceptual mechanism of access to the content of one's own mind.

We now have two axes that define a space of metacognitive theorizing – one for representational format and one for the person whose mental states are represented by the metacognitive state. Theory-theoretical accounts of mindreading occupy the propositional, other-directed corner of the space, while the application of theory-theory to introspection occupies the propositional, self-directed corner along with transparency theory. Epistemic feelings occupy the non-propositional, self-directed corner of the space while simulation theory and Gallagher's account of mindreading occupy the non-propositional, other-directed corner. Any adequate definition of metacognition ought to include these prominent accounts, and thus ought to cover the entire space thus defined.

	Propositional	Non-Propositional
Self-Directed	Transparency theory Theory-Theory	Epistemic Feelings
Other-Directed	Theory-Theory	Simulation Theory Nonconceptual Mindreading

Figure 1.1 Areas of Metacognition Research

1.3. Defining metacognition

I have characterized the breadth of metacognition research along two axes – the representational format and the owner of the target state. While there may be other ways of characterizing metacognition research, a definition that captures the whole of this space ought to capture a great

deal of what is studied as metacognition. It is, of course, provisional, but it should be more extensionally adequate than "thought about thought," "knowledge about knowledge," or "cognition about one's own cognitions."

> **Metacognition**$_{\text{DEF}}$ is any mental state, mental process, or mental property that takes another mental state, mental process, or mental property as its intentional object.

This definition captures the 'meta' in metacognition, in that it is a higher-order phenomenon, as well as the 'cognition', in that each order is mental. That the relation between the two states is a mental one is captured by characterizing the lower-order state as the intentional object of the higher-order state. I make no restrictions on the format of those states beyond their status as mental. Likewise, I make no restriction as to the subject of either the higher- or lower-order mental state, process, or property (or indeed whether there is a subject). I won't try to define 'mental', as that's a task for another essay. I will say that, on my conception, mental states, processes, or properties must involve representations. This may rule out some, but certainly not all, neural states, processes, and properties. And obviously not all representations qualify as mental, for example this sentence.

But not anything goes. By characterizing these states[5] as intentional I assume a higher bar than mere causality. Mental states cause and are caused by other mental states, but this by itself does not make them higher-order states any more than a cue ball in billiards is a higher-order ball. What is required is that the higher-order state be "about" the lower-order state in the intentional sense of 'about'. There is no consensus on the proper analysis of intentionality, but

[5] From here on out, take 'mental state' to mean 'mental state, process, or property' except where I make an explicit distinction among the three.

there is broad agreement on which states are intentional. Belief, desire, and other propositional attitudes are intentional states. Perception is also an intentional state. There is more controversy over whether phenomenal states, feelings, or nonconceptual mental states are intentional. There is some controversy over whether nonconceptual mental states even exist. While metacognitive research includes, in practice, the study of states that many would identify as phenomenal or nonconceptual, this will not convince everyone that such states exist and have intentional content. I will present more reason to believe in the existence and mental status of nonconceptual content in chapters 2 and 3. But we should not rule it out a priori, so my provisional definition of metacognition will allow for the possibility. In the next section I will argue that, given my provisional, ecumenical definition of metacognition, reality testing qualifies as metacognitive.

1.4. Reality testing is metacognitive

In its propositional form, reality testing is clearly metacognitive. Recall the two episodes described at the beginning of this chapter. When, on night 2, you believe that you are perceiving the song, subsequently doubt that you are, and so on, you entertain propositional attitudes whose objects are mental states. This satisfies our definition of metacognition because belief and doubt are mental states or processes and the intentional object – a perceptual state - is also a mental state or process. Even if you were mistaken and the state was in fact imagined, the reality testing process was still metacognitive, as it involved beliefs (false ones) about mental states.

The reality testing that occurred on night 1 is a more difficult case. You determined that the song was imagined without conscious deliberation. Whatever process made that determination, if there was such a process, was unconscious. You simply noticed the imagined

song, became annoyed, and engaged in appropriate behaviors to fall asleep despite it. Your behavior reveals that successful reality testing occurred because, rather than getting up and attempting to eliminate some external source of the song, you attempted to end it by distracting yourself mentally. This unconscious, automatic discrimination of mental state types is by far the most common form of reality testing. If we had to engage in conscious deliberation over each episode of imagination or perception, we would hardly get anything else done. In what follows I will survey the ways it could have been accomplished and will argue that all the plausible options qualify as forms of metacognition.

1.4.1 Propositional and non-propositional reality testing

It is possible that the reality testing on night 1 proceeded much like the propositional version on night 2, but unconsciously. It is widely accepted that propositional attitudes can operate unconsciously, and if so, the process is no less metacognitive for being unconscious. It still involves mental states taking other mental states as their intentional objects.

Another possibility is that the process occurred in a non-propositional format. If the process was nonetheless mental, then it involved representations, but representations of a nonconceptual sort. Concepts are the building blocks of propositional content and are typically thought to go hand in hand with propositional forms of representation. Nonconceptual content is conceived as a different representational format from the propositional attitudes. Examples include the color content of one's perceptions, or one's "knowledge" of how to throw a football in a perfect spiral. If reality testing is a phenomenal state like color content, then it may be a kind of epistemic feeling, like the tip-of-the-tongue state or the feeling of knowing. Feelings of reality

and feelings of presence have been cited as epistemic feelings that inform us about the perceptual nature of our mental states. Epistemic feelings are, on the evaluativist account, metacognitive states. There is controversy on this point, however (See Carruthers, 2017). While epistemic feelings may serve some metacognitive function, they are insufficient to accomplish reality testing on their own. They simply push the problem back. If feelings of reality are sufficiently accurate to be adaptive (they needn't be foolproof - they obviously fail in the case of dreams), then this must be because they are sensitive to some property that distinguishes imagined from perceived states. The process whereby these properties are detected and a determination of the nature of the state is made is the true target of my investigation – that is reality testing. That the output of the process is a phenomenal state rather than, say, a belief is interesting but ultimately not the point. Phenomenal states seem more suited as a format that reports information to the subject and perhaps motivates behavior. We generally conceive of phenomenal states less as information gathering states than as broadcasters of information.

So perhaps the nonconceptual content involved should be better characterized as a kind of skill. When a skilled basketball player recognizes an opportunity to take a shot at the basket, this is sometimes described as perceiving an "affordance" or "feature" in the environment. Affordances and features differ from the sorts of objects and properties represented by concepts in that they are relations to the subject – opportunities for the subject to engage with the environment in some specific way. I will follow Strawson (1959) and call the form of representation involved in perceiving affordances is called 'feature-placing' content. If such skill were applied to one's own mental states in reality testing, then the feature-placing content would represent mental features or affordances. Chapters 2, 3, and especially 5 will explore this form of representation and produce a novel account of how it could be employed in reality testing. For

now, it suffices to say that, if reality testing is accomplished using this form of content, then it does involve mental states or processes that take other mental states or processes as intentional objects. If one feels that they would like to reserve the term "mental state" for propositional states, then call it a process.

One might worry that nonconceptual content can only represent mental states in a de re sense. That is, if nonconceptual states can detect information about one's mental environment, then it is metacognitive in one sense. But one might argue that to be metacognitive in a more robust sense, the intentional object must be represented *as* a mental state, and it is not clear how this could be done without mental state concepts. Compare reality testing to a different kind of sorting task. I might train a pigeon to sort computer parts by color and shape. This would not require that the pigeon have any conception of computers or that the parts be conceived as computer parts by the pigeon. Nonetheless, you could say that the pigeon represents computer parts in a de re sense. But if the pigeon were able to sort the parts on the basis of their contribution to the function of the computer – RAM of various shapes and colors in one bin and processor chips in another – then we might be inclined to attribute a conception of these parts as computer parts to the pigeon. Here we could attribute mental content about computer parts in a de dicto sense, and it would rightly be seen as a more impressive ability.

If we apply this distinction to reality testing, we see that the ability is of the more robust sort. The very fact that perceived and imagined content are sorted accurately (for the most part and within certain bounds) and appropriately to their function in our psychology suggests that they are sorted on the basis of their relevant mental properties, and not some other property (say duration or color). Now it could be the case that there is some non-mental property of these states that reliably correlates with the relevant mental properties to which reality testing is sensitive.

There are many accounts which posit various properties to which reality testing is sensitive. Some of the properties are mental, such as Hume's "vividness," and thus support the metacognitive nature of reality testing (more on this in chapter 4). Other accounts involve propositional content and are thus irrelevant to the hypothesis of nonconceptual reality testing. But some accounts, Goldman's (2006), for example, could challenge the metacognitive nature of nonconceptual reality testing. Goldman posits that we determine the mental attitude taken to a given content on the basis of direct introspection of neural properties. Neural properties may or may not be identical to mental properties, but humans tend to treat them as different concepts. If reality testing is a result of a sensitivity to properties of neurons, then those properties needn't, it would seem, be conceived *as* mental. Of course, if it turned out that reality testing required sensitivity to the functional properties of those neurons, and those happened to be the same functional properties of the relevant mental states the distinction would not be so clear. In any case, there are other problems with Goldman's account, which I will cover in detail in Chapter 4. There is a prima facie case, then, that reality testing is not only a metacognitive ability in a de re sense, but also in a more robust sense in which the lower-order states are conceived as mental states, in a nonconceptual sense of 'conceived'.

1.4.2 First-order and functional/architectural interpretations.

Is it possible that reality testing could be accomplished entirely by first-order mental states? One might propose that the problem of determining whether one is perceiving or imagining a horse could be solved simply by answering the question 'Is there a horse before me'? This is, of course, not how reality testing is defined in the literature. Nor is it going to be sufficient to

distinguish reality testing from other sorts of abilities. One can sensibly ask and answer questions about the presence of horses regardless of what sort of imagery is being entertained. It only becomes a reality testing problem when one experiences an image of a horse. And if the subject of the process is the nature of an experience of a horse image, we have entered the realm of the metacognitive.

I have argued for the metacognitive status of conscious conceptual reality testing, unconscious conceptual reality testing, and nonconceptual reality testing. But why assume that reality testing is a mental process in the first place? One possible explanation that I will consider in detail in Chapter 4 is that reality testing is accomplished by the architectural properties of the brain. If the outputs of imaginary states are simply not connected to motor outputs in the same way as perceptual states, then our behavior vis a vis these states will appear as if a complicated problem was solved when in fact there was little chance that it could go wrong. In this case we might say that reality testing is not a mental process at all, since it is not accomplished by any representational states or processes. A commonly accepted mark of the mental is intentionality. If we assume this mark of the mental, then whether unconscious, nonconceptual reality testing is a mental process ultimately depends on whether it has intentional content. To have intentional content is to be "about" something, and this is typically cashed out as a representational relationship. An important indicator of whether something is a representation is whether it can misrepresent. So, we might be able to determine whether reality testing is a mental process with intentional content by examining cases where it fails and determining whether these are cases of misrepresentation.

Naturalistic accounts of misrepresentation often define it as a failure to accomplish a representational function. But there are many functional failures that are not misrepresentations.

We can distinguish cases of mere failure to perform a function from cases of misrepresentation by examining the failures and determining whether they exhibit systematicity at a psychological level of description. In other words, are we dealing with a software failure or a hardware failure? The view I have described essentially posits that reality testing is solved by the hardware of the brain. But we can make reliable inferences about what kind of failure we are dealing with by observing the behavior of the system. If I set my word processing program to Times New Roman but it consistently outputs Comic Sans when I type words beginning with 'C' we would suspect a software failure. We observe regularities in the malfunctioning behavior that are naturally couched in the language of the software programming. If the laptop randomly shuts down, restarts, or displays random patterns on the screen we might suspect a hardware issue, as the description of the malfunction is less naturally couched in the language of software.

When reality testing fails, it exhibits predictable patterns at a psychological level of description. An experience of one type is mistaken for an experience of another type, but both perception and imagination are psychological predicates. You are confused on night 2 in part because the sound experience is faint. Had the experience been louder, it would have been obvious that it was perceived. This is one reason that theorists like Hume held that the distinction between imagination and perception lies entirely in the relative vividness of the experiences. In chapter 4 I will argue that this cannot be the whole story, but it does indicate that reality testing is a psychological phenomenon. The loudness or "vividness" of your experience is also a psychological property and manipulation of that variable produces changes in other psychological variables. What we do not observe is the neural activity that realizes imagining a song being mistaken for a desire to eat ice cream, or for a motor command for flexing the right arm. Such failures would suggest that the errors are occurring purely at the neural level, as they

offer no predictable pattern at the psychological level. Of course, an important task of this essay is to offer an account of the mental properties to which reality testing is sensitive. An accurate account ought to predict just the sorts of failures we observe. These will be addressed in detail in chapter 5. But for the moment we can conclude that, prima facie, the ways that reality testing fails suggest an explanation at the psychological level. And if this is the case, then reality testing is a mental process.

I have argued that reality testing is a metacognitive process if we assume a definition of metacognition that captures the breadth of research conducted under that name. This definition is quite general. There are other accounts of metacognition that have more specific conditions on what counts as metacognition, and in the next two sections I will argue that reality testing satisfies these as well.

1.5. Nelson and Naren's model of metacognition

Nelson and Narens' (1990) model of metamemory is a prominent model of metacognition. Metacognition, on this account, performs three main functions – collecting "accumulated autobiographical information about one's own cognitions," "the ongoing monitoring of one's own cognitions," and "the ongoing control of one's own cognitions" (p. 1). These categories are not presented as necessary and sufficient conditions for metacognition, but rather as types of processes that they consider metacognitive.

The autobiographical information Nelson and Narens cite in their first condition is conceived as propositional content. I have pointed out that such forms of metacognition exist,

and have argued that reality testing in its propositional form is metacognition of this sort. The latter two conditions, however, are unique to this model.

The monitoring function of metacognition is described as a directed informational relation in which first-order information from memory feeds into a metacognitive mechanism that contains a model of the first-order processes, makes predictions on the basis of that model, and in its control function, feeds commands back to the first-order level on the basis of those predictions. This account of metacognition is based in control theory and is mirrored in control-theoretic accounts of perception (See Grush, 2004). But there is an important difference that Nelson and Narens do not consider. They cite Conant and Ashby (1970), who point out that a model must be at least as complex as the phenomenon it models. This makes some sense in control-theoretic accounts of perception, where it is posited that the model (or "plant") produces visual experience indistinguishable from that caused by sensory stimulation. This is captured by the dictum that perception is "controlled hallucination" (Clark, 2015). While controversial, such a view of perception is at least coherent. But to apply this sort of account to memory is much less plausible. It would suggest that metamemory contains, or at least can reliably simulate, all the information stored in first-order memory. There are accounts of memory as a constructive process (e.g., Michaelian, 2016), but these accounts mean to *replace* first-order memory as it is typically understood, not double the processing and storage requirements.

Evaluativists have developed an account of the Nelson and Narens' monitoring function that gets around this problem. On that account, metacognitive processes have no access to first-order memory at all, but instead evaluate and predict the success of first-order processes on the basis of heuristics. The heuristics tell the subject how the recall process is going and whether it is worth continuing, but they do not monitor or simulate the memories themselves. In chapter 6 I

will argue against evaluativism in detail. Regardless, the upshot is that the claim that the monitoring function of metacognition requires a model of first-order processes is false.

Nonetheless, one can still accept that metacognitive processes perform a kind of monitoring function in that they receive information from first-order processes. This condition is not entailed by our definition of metacognition. I have only required that a higher-order mental state or process takes another mental state or process as its intentional object. The monitoring function would only be entailed by the intentional relation if one's account of intentionality were informational in a rather strict sense. Most accounts of intentionality allow that our mental states can be about objects that we are not in a direct, occurrent informational relationship with. I can think about Santa Claus, or Peru while in Los Angeles. I can even think about Santa Claus's beliefs and memories.

Nelson and Narens, however, are primarily concerned with self-directed forms of metacognition. Even here metacognition does not require informational input from the target state. I might wish that I believed that I were a great philosopher, in which case there is no informational relationship as the first-order belief does not obtain. Reality testing, however, does satisfy this monitoring condition. Reality testing concerns occurrent perception, imagination, or episodic memory. Insofar as I am entertaining occurrent content, I am receiving the information contained in that content. This might not be the case if reality testing were a constructive process – if in wondering whether the song is imagined I make it the case that I imagine a song. But the fact that reality testing is by and large accurate in its sorting of both perceptual and imaginary states suggests that it is not wholly constructive. This is evidenced by my tendency to reliably eat only perceived and not imagined pizzas. This suggests that reality testing processes respond to

properties of mental states that reliably correlate with their being perceptual or imaginary. In other words, if reality testing did not perform a monitoring function we would not live long.

Nelson and Narens' control function is not entailed by my definition of metacognition either but is also satisfied by reality testing. In the context of metamemory, the controller is responsible for telling memory when to continue or stop searching for the target. It controls the first order process but not the target itself. The reality testing process does not directly control what is perceived or imagined either, but it does have downstream effects. It is, in part, causally responsible for whether I get up to turn off the song or try to distract myself. In Chapter 5 I will argue that reality testing also involves fine-grained sensorimotor skills that could be interpreted as accomplishing a control function over the perceptual and imagination processes themselves. To take one example, visual saccades will result in different visual experience in perception versus imagination, and thus saccades could be used to test whether a given content is imagined or perceived. If reality testing involves the control of the visual system to perform such tests on visual content, then this could be interpreted as exerting a control function over perceptual processes.

Nelson and Narens' account of metacognition introduces conditions on metacognition that are not entailed by my definition of metacognition. Their account is puzzling in some ways and requires a very specific, and I will argue false, evaluativist interpretation to make sense of some parts of it. Nonetheless, reality testing does satisfy the basic monitoring and control functions that the account takes to be essential to metacognition.

1.6. Schwitzgebel on introspection

Metacognition directed at one's own mental states is often termed 'introspection'. Reality testing also concerns one's own mental states. If reality testing is metacognitive, then it is arguably a form of introspection. Schwitzgebel (2014) offers a set of necessary conditions on introspection. He divides these conditions into two sets - three that are shared by all accounts of introspection and three that are more controversial. In this section I will consider each condition in turn and argue that reality testing satisfies all the shared conditions. One of the controversial conditions is not always satisfied by reality testing, but I will argue that this is no threat to its metacognitive status. In the end reality testing may not need to qualify as introspection in order to qualify as metacognitive. Introspection sometimes connotes a deliberate, thoughtful process, whereas reality testing in its unconscious form is largely automatic. But the comparison is useful, as it establishes a strong family resemblance between these metacognitive processes.

First is the *mentality condition* – the object of introspection must be mental. This is also a necessary condition for metacognition on my definition, and I have argued that it is satisfied by reality testing. The second condition Schwitzgebel calls the *first-person condition*. It requires that the introspected state belong to the subject involved in the introspective process. Reality testing is also explicitly a first-person process. When I determine whether I am hearing or imagining a song, it is *my* perception or imagination that are at issue, not someone else's. This condition could be interpreted in a stronger way as requiring that the subject *conceive* of the object state as her own. It is not at all clear whether such a strong version of the first-person condition could be satisfied by nonconceptual reality testing. In a sense, feature-placing content always involves a relation to the self, as it picks out affordances for action by the subject. But in another sense, there is no *concept* of the self being applied. Schwitzgebel does not explicitly make this stronger commitment and to do so would seem to beg the question against

nonconceptual forms of introspection. Of course, the existence of nonconceptual metacognition is controversial and a large part of my aim in this essay is to argue for its existence. If I am successful, then we ought to conclude that reality testing satisfies the first-person condition to the degree that any nonconceptual process can. Third is a temporal proximity condition – the object of introspection should be in the present or very recent past. Propositional reality testing needn't respect this condition. I may wonder whether the elephant I saw yesterday was imagined. But as it is typically studied, reality testing concerns occurrent mental states, and this form is the primary subject of this essay.

These first three conditions are fairly thin. They amount to "being about one's own present cognition." Curiously, a written sentence could satisfy these conditions. We should probably assume on Schwitzgebel's behalf that the introspective process must be a mental one as well, and I have argued that this is the case in reality testing.

Schwitzgebel's other three conditions are all denied by one or another account of introspection. It should not, then, be a requirement that reality testing satisfy all three, but nonetheless it does satisfy two. Condition four is the *directness condition*, according to which knowledge or beliefs gained through a process of inference do not count as introspection. This condition is denied by theory-theorists, who claim that metacognition is a theory-driven, inferential process in both self- and other-directed metacognition. It also rules out transparency theory, which holds that metacognition is the result of the application of an "ascent routine" to first-order states. Propositional reality testing could fail to satisfy this condition. One might imagine a version of night 2 where you eventually conclude that you are imagining the song through a series of inferences – my roommate has left, no one else is in the house, and so on. But on night 1 you unconsciously and effortlessly determine that the song is imagined and not heard.

It is possible that this is accomplished by unconscious inference. You may subconsciously sort such content on the basis of generalizations like 'If an experience has property x, then it is imagined', where x is some distinguishing characteristic of imagined content. Maybe this sometimes occurs in humans. The primary topic of this essay, however, is nonconceptual (i.e., non-propositional) reality testing, which exists in rats, corvids, infants, and is almost certainly preserved in adult humans. (I will argue for this claim in chapters 2 and 3.) Inference requires propositional content (more on this in chapter 3 as well). So nonconceptual reality testing will ultimately satisfy the directness condition.

Condition 5, the *detection condition*, claims that introspection must be a response to an ontologically independent state. That is, in introspection we detect a mental state, we don't create it. There are accounts of introspection that deny this, as many metacognitive states contain a first-order state as a constituent. In believing that I believe the sky is blue, I entertain the content that I believe the sky is blue. One might think, then, that the metacognitive state necessitates the existence of the lower-order state. If so, then perhaps introspective processes create the mental states they report rather than detecting them. This seems less plausible in the case of reality testing. Phenomenologically, it seems that the song experience precedes the deliberation about whether it is perceived or imagined, at least on night 2. And if I am correct that the reality testing that occurs on night 1 is nonconceptual, then it does not involve higher-order propositional states that contain, and thus necessitate, lower-order content. Reality testing also concerns perception, which is typically defined as involving some causal relation with external objects. I cannot make it the case that I *perceive* a cat simply by believing that it is so, though I might make it the case that I *imagine* one. So, reality testing satisfies the detection condition.

The only criterion that reality testing, as I have conceived it, sometimes fails is the *effort condition*. Schwitzgebel proposes that introspection is a process that requires special reflection distinct from the sort of everyday mental monitoring that might exist in a functioning psychology. But I am interested precisely in the everyday, unconscious sorting of experience into perception and imagination. If there is a distinction to be made between self-directed metacognition and introspection, the effort condition captures it. This is, in part, why I call reality testing a form of metacognition rather than a form of introspection. But some well-known accounts of introspection deny the effort condition – for example Lycan and Armstrong's higher-order perception theory of consciousness. And by conceding that there is everyday, unconscious *monitoring* (even if it doesn't count as introspection proper) Schwitzgebel implicitly concedes that even if reality testing does not satisfy this condition on introspection, it is a sort of metacognition.

1.7. Conclusion

In this chapter I have introduced the phenomenon of reality testing and argued that it is metacognitive. In order to make that argument I have had to offer a definition of metacognition, as those provided by the literature that are obviously defective. I have defined metacognition as any mental state, mental process, or mental property that takes another mental state, mental process, or mental property as its intentional object. This definition was deliberately broad, constructed to capture all the phenomena actively studied under the name 'metacognition'. These phenomena include mental states directed one's own mental states and those of others. They also include propositional and non-propositional mental states. I then argued that reality testing

satisfies this definition. For good measure, I considered some conditions on metacognition offered by Nelson and Narens' influential account and argued that reality testing meets these as well. Finally, I considered a philosophical account of a similar ability to reality testing – introspection – and argued that reality testing meets all but one of these conditions. I conclude that even if reality testing is not introspection, it easily qualifies as metacognition.

Chapter 2. Reality testing is Nonconceptual

In chapter 1 I argued that reality testing is a metacognitive ability. In this section I will argue for another claim: some creatures possess reality testing abilities but lack the mental state concepts required to perform propositional reality testing. Together these claims entail that a heretofore unrecognized form of non-propositional (or nonconceptual) metacognition exists. In section 2.1 I will focus on rats and mice, as there is a field of empirical research devoted to establishing and explaining reality testing abilities in rodents.

But the claim is more interesting if the phenomenon is more general, and so in later sections I will argue that many creatures, including humans, engage in nonconceptual reality testing. In these cases, the argument requires more steps. First, any creature that possesses distinct faculties of perception and memory or imagination must also possess the ability to tell them apart. If a corvid, for example, can both perceive a food store and entertain an episodic memory of that food store, it must also understand which of the two types of experience affords eating. Eating behavior should be reserved for times when the food store is perceived, not when it is merely remembered or imagined. Failure of this ability would be devastating to the creature's ability to cope with its environment, as is evidenced in the maladaptive behavior

caused by schizophrenia. If a creature like a crow, a human infant, or a rodent is capable of both perception and imagination or episodic memory and manages to survive and thrive, we can infer that it is capable of reality testing. Second, I will argue that such creatures lack the mental state concepts required to perform propositional reality testing. We can then infer that nonconceptual reality testing exists.

Before I present this argument, however, I must explain what I mean by 'conceptual' and 'nonconceptual'. In section 2.2 I will elaborate an account of conceptual content based on Gareth Evans' "generality constraint." This account of the conceptual/nonconceptual distinction is influential, but not uncontroversial. I will offer some reasons for its appeal, but this essay is not a detailed defense of the generality constraint. In some sense the choice is arbitrary. The generality constraint constitutes a threshold for a certain cognitive ability, and I am arguing that creatures which lack that ability are nonetheless capable of reality testing. But the choice of the generality constraint is not entirely arbitrary. As I will show, the possession of mental content that satisfies the generality constraint is necessary to perform certain types of inferences which are in turn necessary for a theory of mind.

In section 2.3 I will return to the main argument to offer evidence that rats, corvids, bees, and human infants engage in perception and episodic memory. In section 2.4 I will argue that rats, corvids, and human infants engage in imagination. Then in section 2.5 I will argue that rats, corvids, and bees lack the requisite concepts to perform propositional reality testing. Rats are of particular interest because, unlike apes and corvids, there is little reason to think that rats possess a theory of mind. Some have argued that corvids and human infants possess a concept of perception. I will argue that, even if this is true, that concept alone is inadequate to accomplish

reality testing. In section 2.6 I will conclude that my argument establishes the existence of a novel form of nonconceptual metacognition.

2.1. Reality testing in rodents

My goal in this chapter is to establish that nonconceptual reality testing exists. A single example of reality testing in a creature that lacks the requisite concepts, then, would be sufficient to prove that claim. The clearest such example comes from the literature on representation-mediated taste aversion in rodents. This literature, which spans three decades, explicitly addresses the existence of reality testing in rats and mice. Behavior consistent with failures of reality testing are induced in rats and mice using methods similar to those which produce reality testing failure in humans. The behavioral evidence is backed up by studies employing brain lesions, gene-knockouts, drugs, and examination of neural activity. In this section I will critically review that literature and conclude that despite some ambiguities, there is overwhelming evidence that rats and mice engage in reality testing.

Holland (1990) begins with a theoretical question about the nature of classical, "Pavlovian" conditioning. In classical conditioning a conditioned stimulus (CS) (e.g., a tone), which produces no natural behavioral reaction in the subject, is paired with an unconditioned stimulus (US) (e.g., food), that does create such a response - the unconditioned response (UR). The classic example is of Pavlov's dogs, who, after repeated pairing of a bell with food, began to salivate when they heard the bell. After repeated pairing with the US, the CS alone will produce the UR, and this is called the conditioned response (CR). Holland's question is at what point the CS inserts itself into the causal chain that normally produces the UR. Does the CS directly

stimulate motor responses (the “S-R” interpretation), or does it stimulate a representation typically produced by the US that then causes the behavior (the “S-S” interpretation)? And if it is the latter, how early in that unconditioned causal chain does the CS interject this representation? Holland posits that the intervention is quite early, and that the CS causes a hallucinatory sensory experience of properties of the US. This has been reported by humans in Pavlovian conditioning, who experience hallucinatory experiences of electric shock (Cole, 1939; Garvey, 1933), auditory tones (Ellson, 1941), and colors (Howells, 1944) in response to a CS.

Reality testing failure in rats and mice is typically produced using conditioned taste aversion. Conditioned taste aversion occurs across many species. When a taste is paired with illness, later exposure to that taste produces behavior that indicates the flavor has become unpleasant to the subject (Chambers, 2018). You might have experienced something similar if you have ever had food poisoning. You probably avoided the type of food altogether for some time, but if you had tasted it soon after the experience, it is likely that you would have experienced an unpleasant flavor. The effect has likely evolved to train us to avoid poisonous substances. Conditioned taste aversion is used to produce failures of reality testing in rats and mice in two distinct ways in the literature. Both involve first pairing the CS, typically either a tone or a scent, with a particular food, typically sucrose or flavored sucrose (the US). In the stimulus devaluation paradigm, the US is then paired with an injection that causes nausea. This causes the rodents to avoid the food for some time after. It is found that the CS alone will cause aversive behavior associated with conditioned taste aversion. Holland performs several variations on this study, aiming to demonstrate that the aversive behavior is caused by a sensory representation. Holland does not use the term reality testing to describe the effect, but he does endeavor to show that the rodents are mistaking a representation evoked by the CS for a

perception of the US, which is a failure of reality testing. Later researchers make the connection explicit.

In the representation-mediated taste aversion paradigm, the US is paired with a tone, but nausea is later paired with the CS, not the US. Nonetheless, rats and mice will later demonstrate aversion to the US. It seems that the representation evoked by the CS is vivid enough that it produces an association between the US and nausea. However, the effect only lasts briefly in normal rats and mice and after a handful of training sessions food consumption returns to normal. Manipulations that mimic schizophrenia, however, cause the effect to persist. Together these results suggest that representation-mediated taste aversion causes temporary reality testing failure in normal mice, but the ability eventually kicks in.

In what follows I will describe several of these studies in some detail, answering potential objections to the conclusions they draw.

2.1.1 Stimulus devaluation

Holland (1990) trains rats to form two distinct associations. First, two different audio tones are paired with two distinct flavors of sucrose solution (both delivered to the same clear cup). The CR is contact with the cup. Rats showed no preference for one flavor over the other. The rats were then given an injection that induces nausea in conjunction with solution 2. This is done in the absence of the associated tone. Then, in the absence of sucrose or toxin, the tones were presented. Production of the CR was greatly reduced for tone 2 but not for tone 1. Since the cup is the same in either case, the difference cannot be due solely to an association of the sucrose itself with nausea. Since the tone had never been paired with nausea, he concludes that the

behavioral effect of the tone is mediated by a representation it evokes. The representation evoked by the CS must be of the food (or the smell, taste, or some other property of the food). The value of the intentional object of the representation has been reduced by its association with nausea, and this is displayed in the rat's behavior.

However, this study does not show that the representation involved is sensory in nature. But there is reason to believe that it would be. The varieties of negative food associations and their corresponding behaviors are well-known. Rats react to bitter or otherwise unpleasant flavors with "aversive" behaviors – gaping, chin rubbing, head shaking and flailing of the forelimbs. Association of a food with a toxin or a bitter taste like quinine induces aversive behaviors, whereas association with electric shock simply reduces consumption. Aversive behavior, then, is indicative of a negative gustatory experience. When Holland's conditioned rats were later offered a non-flavored sucrose solution, tone 2 evoked aversive behaviors when the sucrose was ingested (this effect is also obtained by Delamater et al., 1986; Kerfoot et al., 2007). Holland concludes that the representation evoked by tone 2 was the taste of food two, to which the rats had developed an aversion by association with nausea. No such effect was produced by tone 1. Fry et al. (2020) show increased activity in gustatory areas of the brain in a similar experiment.

Further evidence for the sensory nature of the representation evoked by the CS comes from the sorts of combinatorial effects it produces. Holland (1990) trained rats to associate tone 1 with a sucrose solution and tone 2 with a saline solution, both of which produced ingestive behavior equally. He then paired nausea with a combination of the two solutions. It was found that a combination of the two tones produced aversive behavior where each tone separately did not. Holland found that the reverse worked as well. If nausea was induced with each flavor

separately but not with the combination, rats exhibited aversive behaviors in response to each tone separately but not to their combination. This suggests that the combinatorial nature of the representation is not that of classical logic. The conjunction of two bads ought not entail a good, unless the representations are combining in a way that creates something new. Sensory experience has this quality - red and blue make purple. Holland also claims that these results rule out a simple association of the tone with a motor response. The sensitivity of the behavior to combination suggests that the CS intervenes at the level of representation. Holland found that when rats were given electric shock instead of toxin-induced nausea the same effect was not observed. This is consistent with the hypothesis that the nausea changes the flavor of the food for the rat, whereas electric shock does not. This claim is supported by neural evidence. Later researchers (Kerfoot et al., 2007), show increased FOS expression (a protein associated with neural activity) in gustatory cortex in the devaluation condition.

Classical conditioning in humans suggests that these sorts of failures of reality testing are not limited to rats, supporting the claims about the sensory nature of the representations. Humans report sensory experiences of the unconditioned stimulus after receiving the conditioned stimulus alone. These include experiences of auditory tones (Ellson, 1941), electric shocks (Cole, 1939; Garvey, 1933), and shades of color (Howells, 1944). More recent studies replicate these findings (Powers et al., 2017), showing that patients with psychosis are more likely, after repeated pairings of a tone and a visual stimulus, to report hearing the tone when the visual stimulus is presented alone. Functional neuroimaging during these conditioned hallucinations confirms activity in sensory areas associated with hearing a tone.

Holland's studies are not immune to criticism. The different effects for different combinations of flavors could also suggest that the representation is not of flavor at all, but of

some other property of the stimulus. In the next section I will survey some more recent studies that use gene knockouts, neural evidence, and drug interventions to determine the nature of the representation involved in conditioned taste aversion.

2.1.2 Representation-mediated taste aversion

Representation-mediated taste aversion (RMTA) pairs a CS with sucrose (the US), as in the devaluation paradigm. But instead of devaluating the US, nausea is instead paired with the CS. Feeding behavior is subsequently reduced, even in the absence of the CS. This suggests that the representation evoked by the CS forms an association between nausea and the US. The effect extinguishes, however, after as few as 40 subsequent CS-food pairings, whereas the aversion behavior in devaluation paradigms does not extinguish after as many as 160 pairings (P. Holland, 1998). This suggests that in representation-mediated taste aversion, something in the nature of the representation, or in the subjects' attitude toward it, changes over time. The phenomenon is interpreted by researchers as a temporary failure of reality testing, which is subsequently regained. Early in the process the image of the US evoked by the CS is interpreted by the subject as perceptual. But after repeated pairings of CS and US, the subjects learn to distinguish the CS-evoked representation from actual sucrose. The effects produced by representation-mediated taste aversion are smaller than in the devaluation paradigm and can be difficult to produce at all without genetic or pharmacological manipulations. This is not surprising if the effect is indeed a failure and subsequent recovery of reality testing abilities. Everyday failures of reality testing are typically tentative and temporary. This is a necessary condition of any functioning cognitive

system, for if representations evoked by a CS were self-reinforcing, then extinction of the association would never occur, and chaos would reign.

Holland (2005) challenges a competing explanation of representation-mediated taste aversion by Pearce and Hall (1980), who suggest that a CS loses its "associability" or usefulness in new learning the more reliably it becomes associated with the US. This associability can typically be regained by interposing extinction trials (CS alone). Holland performs three experiments aimed at testing this explanation and finds that extinction trials caused no difference in learning between rats with extensive and minimal CS-US pairings. The minimally trained rats formed a stronger association between the representation evoked by the CS and the CS-paired illness even when the extensively trained rats received extinction trials aimed at increasing the associability of the CS by reducing its association with the US.

Although Holland's research forms the basis of the study of reality testing in rodents, he never uses the term. McDannald and Schoenbam (2009) are the first to explicitly interpret Holland's data as demonstrating a failure of reality testing. McDannald et al. (2011) study the neural basis of reality testing in the representation-mediated taste aversion paradigm. They compare normal rats with those who receive ventral hippocampal lesions at birth. This brain area is associated with imagery, and damage to it is associated with schizophrenic symptoms. Lesioned and normal rats performed identically when the US was devalued by nausea. But only lesioned rats reduced consumption when nausea was paired with the CS. Unlike Holland, who noted brief periods of RMTA in normal rats, McDannald et al. do not produce the effect in normal rats. McDannald et al. argue that the effect cannot be explained by facilitated learning in the lesioned rats, which is sometimes found after hippocampal damage (Bussey et al., 1998), since in the conditioning stages both the lesioned and normal rats learn at the same rate. This

study further strengthens the claim that the effect produced by RMTA is a failure of reality testing, as it is facilitated by factors associated with schizophrenia.

Kim and Koh (2016) provide evidence that the extinction of RMTA is a return of reality testing. They compare the performance of normal mice with that of genetically modified mice. Mice with a gene knockout (phospholipase C beta 1) are commonly used as a model of schizophrenia and exhibit behaviors associated with schizophrenia, including sensorimotor gating deficits, hyperactivity, social abnormalities, and working memory deficits. Despite these deficits, knockout mice form conditioned associations like normal mice. Kim and Koh were able to produce RMTA in both normal and knockout mice but found that knockout mice did not exhibit the extinction effect, whereas normal mice did with further training.

Koh et al. (2018) expose mice to ketamine during adolescence, a procedure that has been shown to produce symptoms of schizophrenia (Corlett et al., 2007; Krystal et al., 1994). Like McDannald et al. (2011), Koh et al. do not produce RMTA in normal mice but do produce the effect in the ketamine-treated mice. Treatment with an antipsychotic (dopamine antagonist risiperidone) reverses the effect. Ketamine treated mice also showed a greater response to a dopamine agonist (amphetamine), which indicates dysfunction of the dopaminergic system, also associated with schizophrenia.

2.1.3. Simple Conditioning

While later studies primarily use the representation-mediated taste aversion paradigm, a recent study provides strong evidence that Holland's earlier devaluation paradigm does produces failures of reality testing = but without using devaluation. Fry et al. (2020) use genetically

modified rats. The modification is associated with schizophrenia (disrupted-in-schizophrenia-1 or DISC1). After being trained to associate the presence of a sucrose solution with a tone, in the presence of the tone alone the modified mice will produce a licking behavior typically associated with sucrose with plain water alone. This suggests that the tone produces a gustatory representation that makes the plain water taste sugary. An extinction effect similar to Holland's RMTA conditions was found, where normal mice with minimal training showed a stronger hallucinatory effect than those trained more extensively. Modified mice did not show the extinction effect. The effect is eliminated by the injection of an antipsychotic (haloperidol) in both wild-type and modified mice and was augmented by social isolation in adolescence in the modified mice. The increased licking behavior was accompanied by increased activity in insular cortex, a brain area associated with gustatory sensation. Fry et al. do not pair nausea either with the US or the CS, instead directly testing the effect of the CS-US pairing on behavior. Increased licking with plain water suggests a hallucinatory taste, which is confirmed by the neural evidence and supported by the behavioral differences in the modified mice. The extinction under minimal training also suggests that reality testing does kick in, even without the relatively complicated representation-mediated taste aversion paradigm.

2.1.4 Analysis

These studies show that sensory representations of flavor can be induced that rats and mice take to be perceptual when in fact they are merely imagined. The rats experience a flavor in the absence of any stimulus that would normally produce that flavor, a conclusion which is supported not only by the behavioral evidence but by neural evidence. This is a failure of reality

testing. This failure can be produced in long-lasting and short-lived ways. The effect is typically long-lasting in stimulus devaluation conditions and short-lived in representation-mediated conditions, though the short-lived version can also be produced with straightforward CS-US pairings and no devaluation. The short-lived version demonstrates the resilience of the reality testing ability. Manipulations associated with schizophrenia increase the effect and subsequent manipulations that treat schizophrenia reduce it. As schizophrenia is associated with a failure of reality testing, this supports the conclusion that these studies induce failures and subsequent return of reality testing abilities. Thus, it does appear that rats and mice engage in reality testing.

Some philosophers deny that taste has an intentional object (Burge, 2010), which would imply that reality testing is an inappropriate way to describe the behaviors in these studies. I will say more about this in chapter 5, where I offer a sensorimotor account of reality testing in various modalities, including taste and smell. The point is the subject of some controversy. It seems natural to describe what is going on with the rodents in these studies as a failure of reality testing. If plain water starts tasting bad to you, something has gone wrong with your perceptual system. Tastes correlate with objects in the world in the predictable sorts of ways we associate with perception. Regardless, the rest of this chapter will offer arguments based in modalities other than taste.

2.2 Conceptual and Nonconceptual Content

I have argued that rats and mice possess reality testing abilities, which in chapter 1 I argued is a metacognitive ability. In sections 2.3 and 2.4 of this chapter I will argue that corvids, human infants, and possibly bees, also engage in reality testing. This argument will take a different

form, as the sort of explicit studies of reality testing explored in 2.1 only exist for rodents. Instead, I will argue that any creature that possesses faculties of perception and imagination or episodic memory must possess a reality testing ability. So the arguments in 2.3. and 2.4. will involve presenting evidence that corvids, human infants, and bees possess these faculties. In the course of that argument, it will be necessary to distinguish episodic memory from semantic memory, the latter of which involves information stored in a propositional format. I will argue the performance of certain memory tasks cannot be explained by semantic memory, because these creatures lack the concepts required to do so. So I need to say something about what I take concept possession to be and how concept possession can be determined in creatures who lack language. In section 2.5. I will argue that these same creatures lack mental state concepts, and thus that they employ nonconceptual reality testing.

What is nonconceptual content? Nonconceptual mental content is a form of mental representation. When a subject entertains a mental representation with a given content but lacks the concepts required to specify that content, that content is nonconceptual. The color content of perception is one common example. When choosing from among paint samples at the hardware store I can visually discriminate at least five shades of teal when they appear simultaneously in my visual field. But I, like most people, do not have five distinct teal concepts. Such concepts can be acquired. An employee at Dunn-Edwards Paint developed the shades and named them ‘Salamander’, ‘Pacific Foam’, and so on. Perhaps with practice my teal concepts could become that fine-grained. But I do not need those concepts to discriminate those shades in perception. Nor does this ability to discriminate shades presented simultaneously entail that I could identify Salamander later. I seem to have no stable representation of Salamander that outlasts the perceptual episode. Many take such stability to be a minimal condition on the possession of a

concept. If so, then there is nonconceptual content. Others take any categorization ability to be sufficient for concept possession. But if I do not possess the concept 'Salamander', what concept am I employing to categorize the shades as distinct from one another? It has been argued that the demonstrative concept 'that shade' would suffice (McDowell, 1994), though it is not clear how it would allow me to distinguish more than one of 'that shade'.

Whether an ability is nonconceptual, then, depends a great deal on what we take concepts to be. This essay is not an investigation of the nature of concepts, nor can it be a detailed defense of any one account of concepts. But there are reasons to favor one view in the context of reality testing – the account presented by Evans (1982), inspired by Strawson (1959), and elaborated by a group of philosophers through the 1990s and early 2000s, including John Campbell, Adrian Cussins, and J. L. Bermudez. Here I will explain that account of nonconceptual content and argue that it is the most appropriate one to employ in the current context.

Evans articulates the necessary conditions for the possession of conceptual content with his *generality constraint*.

> We cannot avoid thinking of a thought about an individual object *x*, to the effect that it is *F*, as the exercise of two separable capacities; one being the capacity to think of *x*, which could be equally exercised in thoughts about *x* to the effect that it is *G* or *H*; and the other being a conception of what it is to be *F*, which could be equally exercised in thoughts about other individuals, to the effect that they are *F*" (G. Evans, 1982, p. 75)

By 'thought' Evans means propositional mental content like belief. Beliefs are posited in philosophy and cognitive science to explain behavior. My walking to the fridge can be explained

by a desire to eat something delicious, a belief that there is pizza in the fridge, and a belief that pizza is delicious. This sort of psychological explanation requires that my beliefs and desires are composed in such a way as to allow for inference. That is, for my desire to eat something delicious, my belief that there is pizza in the fridge, and my belief that pizza is delicious to cause me to walk to the fridge, there must be a common conception of deliciousness that is operative in both the belief and the desire. It is the very sort of thing that I desire which I believe to be in the fridge. This requires that the same concept be employed in each case. Thus, concepts must be capable of being combined in at least as many ways as there are potential behaviors. For adult humans, this requires a great deal of generality. We can apply deliciousness to various objects and attribute other properties to the pizza – coldness, greasiness, and so on. In metacognition we can also conceive of others as desiring something delicious and calculate whether they will try to get our pizza.

For Evans, then, mental content qualifies as conceptual if it exhibits this extreme form of compositionality. If there are mental representations that do not satisfy the generality constraint, but nonetheless qualify as mental, these will be forms of nonconceptual content. As the generality constraint sets a rather high bar for conceptual content, this leaves a good deal of room for nonconceptual content to govern behavior in ways other than through belief-desire psychology. One could insist on a lower bar for conceptual content which entails that any creature whose behavior is susceptible to explanation in terms of mental representations possesses concepts. But as the paint-sample example shows, at least *some* of our behavior appears to be governed by nonconceptual content. Skills like walking and shooting baskets also seem less obviously susceptible to explanation in terms of propositional content. The tradition following Evans offers another sort of explanation.

Evans' own conception of nonconceptual content does not suffice as a way of explaining behavior in creatures who lack concepts. Following Strawson (1959), he conceives of nonconceptual content as the detection of properties in the environment without attributing them to any particular object. Strawson offers the example 'It is raining' as a sort of content that picks out an environmental feature without applying a predicate to a singular term in the standard propositional format. One might imagine a creature that can detect odors or other sensory information about the environment without conceiving them as properties of objects. Such "feature-placing" content, in Strawson's terms, does not satisfy the generality constraint. Since a feature is not attributed to any objects, there is no way it could be applied to *every* object, as the generality constraint requires. But it is not clear how Strawsonian/ Evansian feature-placing content could govern behavior in a way that approximates belief-desire psychology. Dretske (1981) characterizes the conceptual/nonconceptual distinction in terms of the amount of information carried. Concepts are "digital," in that they carry only information to the effect that a is F, whereas nonconceptual content is "analog," carrying far more information. This makes sense of our color example – being untutored in various shades of teal, my concepts in this realm are relatively coarse, whereas the content of my perception contains more distinctions than the concepts I possess. And perception carries even more information than that – light and shade, context effects, and so on. But from the standpoint of psychological explanation, this account of the conceptual-nonconceptual distinction also leaves nonconceptual content behaviorally inert. It does not appear to leave open the possibility that creatures could exist and survive whose behavior is facilitated purely by nonconceptual content.

Evans' account of conceptual content, as we will see, leaves open such a possibility, even though he does not cash it in. And importantly for our purposes, Evans' focus on inferential

abilities and the compositionality required helps to draw a clean line between different accounts of metacognition, and thus of reality testing. One class of accounts of metacognition requires various forms of inference, either in the form of deductions from a theory of mind or in "ascent routines" that transform first-order mental states into metacognitive states. If I am correct that reality testing is nonconceptual, then these accounts will be inadequate to account for it.

Later researchers have developed Evans' account of nonconceptual content. They also bear similarities to Gibsonian psychology and its notion of environmental affordances (Gibson, 1979), with the important difference that Gibson's theory is behavioristic and thus does not involve mental representation. Dummett (1993) develops an account of non-propositional cognition he calls "proto-thought." Proto-though is a form of perceptual content that presents possibilities for action. Dummett conceives of it as imaginary content overlaid on the perceptual content. Proto-thought goes beyond feature-placing content in that it sketches, albeit very minimally, a behaviorally explanatory form of nonconceptual content. On this account the content of the subject's perception relates directly to its opportunities for behavior. No inference is required from the perception of an opportunity for eating to the thought that one should eat it – the feature is presented as to be eaten. (An astute reader might notice that something is missing here. We must explain occasions when the creature does not eat in the presence of edibility. Bermudez fills in the gap shortly.) This form of content fails the generality constraint because the content does not even have the sort of predicate-object structure to allow for the generality constraint to be satisfied. The creature does not conceive of an object as being edible – it sees free-floating edibility.

Campbell (1995) describes a similar type of content he calls "causal-indexical" content, which represents properties of the environment only as they relate to the causal powers of the

subject. The subject-relative nature of these representations makes them nonconceptual according to the generality constraint. Not only does this content lack the sort of structure necessary for the attribution of such a property to any given object, since the creature does not conceive of objects, but it is also only applicable to a restricted set of circumstances. The content captures the fact that *this* action is afforded to *me* in *this* circumstance. It is not inconceivable that this sort of content could be applied to other objects (although again, they lack the requisite structure). However, the resulting content would rarely be true and peculiarly uninformative without the specific abilities that the creature itself brings to the table. Though it is not explicit in Evans' formulation, we will see that later theorists argue that a necessary condition for conceptual content is the ability to conceive of objects as existing independently of the self. This, they will argue, is required for the object-predicate structure of conceptual content. The indexical nature of Campbell's nonconceptual content brings out this failure.

Cussins and Bermudez fill in important gaps in the nonconceptual account of behavior. Cussins (1992) develops a normative standard for nonconceptual content. Rather than being assessable in terms of truth, nonconceptual content is assessable in terms of "fluency." If a certain action is afforded and the action progresses without mishap, we can say that the mental state is accurate. Bermudez (2007) points out the necessity of some form of desire. We cannot explain behavior purely in terms of perceived features or affordances, as creatures may choose whether or not to exploit these affordances depending on their desires. A perceived eating affordance alone will not prompt eating if the subject does not want to eat. Bermudez offers a distinction between nonconceptual "goal-desires" and conceptual "situation-desires." Situation desires are propositional, in that their content can be captured by that-clauses that specify a particular situation or state of affairs. A desire *that* I eat a pizza would be a situation-desire. But

we also speak of desires *for* certain things, such as a simple desire for food, or a desire to eat. These sorts of desires do not require that-clauses to specify their content. They are simple drives for survival. Some such drive must be posited in any creature whose behavior can be interpreted and predicted, but it is not necessary to posit propositional content. For the creature endowed only with nonconceptual content, the combination of a goal-desire and a perceived affordance that would satisfy that desire is sufficient to explain behavior. If disposed to eat thanks to a goal desire with that content, I will eat when such an affordance is presented, so long as no stronger desire/affordance pair exists. The presence of a predator may alter my behavior. Proto-desires fail to satisfy the generality constraint for the same reasons as other feature-placing content. They lack an object/predicate structure, and they are indexical in their content.

Bermudez's account is incomplete, however. He describes nonconceptual content as unstructured. One might think that this must follow from its failure to satisfy the generality constraint. But there is a middle ground between the extreme form of compositionality required by Evans and a more basic sort that even a creature who possesses only nonconceptual content will require. The creatures' psychology must match the affordance in the environment to the goal represented in the desire. A goal-desire representation of eating alone ought not engage the relevant motor responses or the creature will attempt to eat all the time. It must engage only when there is a perceptual representation whose content matches that of the goal desire. Even at the nonconceptual level there is a need for an ability to entertain the same content under different attitude types and to keep the two types distinct. The problem parallels the problem of reality testing, in which we can distinguish content entertained in imagination from the same type of content entertained in perception. It seems that even at the nonconceptual level some minimal amount of compositionality is required for psychological explanation, which in turn entails some

minimal amount of generality. The generality constraint is compatible with this, as Cussins (1992) points out. Cussins argues that the distinction between nonconceptual and conceptual content is a continuous one, without a bright line separating the two. He also argues for a kind of structure underlying nonconceptual content – a nested series of sensorimotor dispositions. Such an account might allow for the various levels of compositionality and generality it appears are required to explain behavior on the basis of nonconceptual content. Grush (2007) offers a detailed account of the compositional structure of spatial content that is compatible with that content's being nonconceptual. I will explore these accounts in detail in chapters 3 and 5, as I will make use of them in my account of reality testing.

I have offered one account of nonconceptual content that has its roots in Evans' generality constraint as the measure of conceptual content. This is not the only account of the conceptual-nonconceptual distinction, nor is it the only one that might be used to argue that reality testing is nonconceptual. Hume, for example, holds that mental images are distinguished from perceptions by being less vivid. If this were the basis of reality testing, and if the vividness of our representations were, like color properties, nonconceptual, then one might argue on this basis that reality testing is nonconceptual. This is not the view I favor, and I argue against such accounts in chapter 4. In chapter 5 I will argue that reality testing is accomplished on the basis of sensorimotor dispositions and so it naturally fits into the Evansian tradition, particularly as elaborated by Cussins and Grush. But there are other accounts of concepts. If to have a concept is simply to be able to distinguish Fs from non-Fs, then something like feature-placing content would suffice and Evans' generality constraint is false. As I have pointed out, setting such a low bar for concepts does little to help explain the role of concepts in behavior. It does nothing to explain why the creature eats, fights, or avoids Fs on different occasions. The accounts of the

conceptual/nonconceptual distinction which derive from Evans' prioritize psychological explanation. They allow for the possibility that the behavior of some creatures could be governed and explained purely terms of nonconceptual content. This is useful for at least two reasons. First, such creatures might be actual, and those creatures might include humans at certain stages of development or certain subsets of adult human behavior. It also provides a reductive basis for the explanation of the development of conceptual content in both ontogeny and phylogeny.

So while I think that the argument over what counts as conceptual content is more than a merely verbal dispute, those who are so disposed may simply think of it as a useful marker. If one objects that Evans draws the conceptuality line too high, then my thesis could always be rephrased as the claim that reality testing is a metacognitive, *noninferential* ability rather than a *nonconceptual* ability. I want to argue that metacognition, in the form of reality testing, is more phylogenetically and ontogenetically basic than has previously been posited. A helpful shorthand for this is to call it nonconceptual. But it would still be of interest even if one were to insist that all mental content were conceptual. Some minds are still more primitive than others, and my claim is that metacognition is quite primitive indeed.

2.2.1 General criteria for concepts in nonhuman animals

It is not obvious how Evans' generality constraint can be applied to determine whether an animal possesses a particular concept. In this section I will survey and evaluate various criteria researchers use to this purpose and argue for a set of criteria that is sufficient for our purposes.

Some researchers use the ability to categorize stimuli as evidence for concept possession. That is, if rats can be trained to distinguish triangles from squares in order to receive a reward,

some take this to be evidence that such rats have the concept TRIANGLE. Mere categorization abilities, however, are not sufficient to show that animals are using concepts rather than feature-placing content. The training may simply cause the rats to detect features in the environment that produce eating affordances. Other researchers set the bar too high, claiming that linguistic ability is the only evidence of conceptual content (Davidson, 1975). Some animals do arguably satisfy the linguistic criterion, for example apes who have been taught sign language and some parrots, but the criterion itself is suspect. Deaf children subject to language deprivation in the home typically exhibit slowed cognitive development (Hall, 2017), but there is no evidence that they lack concepts altogether. And anyhow it's entirely plausible that I possess color concepts that I can't name – I could probably make three or four relatively coarse discriminations of shades of teal from memory. If we concede, as I think we must, that there is some nonconceptual content that allows us to make discriminations and categorizations and that there is conceptual content that is not linguistic, we need a different criterion for pulling apart the conceptual and nonconceptual than the presence of language.

Allen (1999) has suggested that error detection provides evidence for the presence of concepts in animals. The ability to detect one's own errors, the reasoning goes, distinguishes mere lawlike stimulus-response patterns from the more flexible discrimination abilities afforded by concepts. Pigs display this ability, backing away from an initially mistaken choice (Keddy-Hector et al., 1999). More recent studies have shown impressive error-detection abilities in apes (Tomasello, 2023). Taking error detection as a starting point, Allen offers three criteria that he proposes are sufficient for attributing concepts to an animal O.

i. *O* systematically discriminates some *X*s from some non-*X*s; and

ii. *O* is capable of detecting some of its own discrimination errors between *X*s and non-*X*s; and iii. *O* is capable of learning to better discriminate *X*s from non-*X*s as a consequence of its capacity" (C. Allen, 1999, p. 37).

What these criteria amount to, however, is at most a way of detecting the presence of representations in animals. Error detection suggests that the animal has initially misrepresented the stimulus and is now representing it accurately. But it does nothing to suggest the presence of the sort of structured content that concepts facilitate. Error detection is commonly considered a form of metacognition and is often used as evidence for metacognitive abilities (see Yeung & Summerfield, 2012 for a review). In judging one's initial judgment false, one engages in metacognition. This can take propositional forms but can also occur in nonhuman animals, for whom the attribution of conceptual metacognition is controversial. It is certainly the case that metacognitive abilities entail the presence of first-order mental content. To claim that it implies *conceptual* content, however, is to beg the question against the possibility of nonconceptual metacognition – the point of this essay. And whether error detection truly counts as evidence of metacognition depends greatly on the details of the task, as entirely first-order interpretations are often available. Allen's criteria may be adequate for detecting the presence of mental states, but they fail to distinguish conceptual from nonconceptual content.

Some psychologists and ethologists set the bar higher than mere categorization, and instead take the ability to categorize *novel* stimuli as evidence for concept possession. That is, having been trained to distinguish equilateral triangles from rectangles on a set of stimuli, rats can continue to accurately categorize new examples of triangles even when the new examples are right rather than equilateral, are constructed of small circles, and so on (Fields, 1932). The ability

to generalize to new examples suggests that these rats have constructed an abstract triangle representation that is not tied to the specifics of the initial examples. Pigeons have demonstrated this ability for the category ‘tree’ as well (Herrnstein et al., 1976). However, some researchers account for this ability without concepts, labelling it ‘stimulus generalization’ (Chater & Heyes, 1994, p. 215). Stimulus generalization is the ability to categorize new stimuli based on their similarity to stored representations of earlier examples that received a reward. This ability does not require the animal to recognize anything further that unites the types of stimuli that received rewards. This representation may be nonconceptual - an image or feature-placing content. This view is supported by the failure of many animals to perform well on such tasks when the contingencies are reversed and the previously unrewarded items from which they were meant to discriminate the target items, say rectangles, are rewarded instead of the target items, say triangles. They also often fail to make the same discriminations when the required behavior changes, e.g., from pecking four different keys to pecking a single key at different rates. And even if animals could pass these tests, for our purposes it does not distinguish between nonconceptual forms of representation and conceptual forms. The ability to detect affordances coextensive with the human concept ‘tree’ or ‘triangle’ does not suffice for the sort of compositionality that distinguishes conceptual content from nonconceptual content and would allow for animals to make inferences using these representations (See Chater & Heyes, 1994, p. 216).

Of course, whether a stored image to which one can compare occurrent perceptions counts as a concept depends on one’s theory of concepts – it might qualify on an exemplar account (Medin & Schaffer, 1978). But on tests designed to detect more robust kinds of concepts the animals in question tend to fail. Pearce (1989) shows that pigeons do not possess prototype-

style concepts. Pearce trained pigeons to peck at histograms with a particular central tendency and predicted that if this ability were underwritten by a prototype-style concept that pigeons thus trained would subsequently peck more vigorously at histograms that were perfect examples of that tendency (3-3-3 or 5-5-5). The pigeons did not behave as predicted. There is also an inherent, and arguably insuperable, problem of attributing concepts to animals when our basis of comparison are human concepts. What is the likelihood that, if animals have concepts, that they look anything like the sorts of categories we assume in our experiments? Computational models of categorization suggest a wide variety of features that can be used for categorization that are highly abstract and correspond very poorly to human concepts.

There is one behavioral criterion for conceptual content that could be taken as evidence for satisfaction of the generality constraint. These are certain types of transfer tasks, and some animals do succeed at the sorts of transfer tasks pigeons fail. Transfer tasks take a learned ability and apply it to a novel context. For example, an animal might be trained to match red stimuli with other red stimuli for a reward. If the animal, without training, is then able to match squares with squares to obtain the reward, that is some indication that the animal understands the concept of matching. That is, it is an indication that the animal has satisfied the generality constraint, for it can apply the concept MATCHES to a variety of stimulus pairs. This method is often only useful for more abstract concepts, for with more basic concepts like TRIANGLE it could always be argued that the animals are using stimulus generalization. But the ability to transfer learned abilities will come in handy as a criterion for us, as mental state concepts are fairly abstract. Some of the concepts we will have to consider when determining whether animals and human infants employ episodic or semantic memory to accomplish certain tasks, like the concept of time, are similarly abstract.

In what follows I will argue that rats, corvids, and human infants possess faculties of perception, imagination, and episodic memory, and thus must also possess reality testing abilities. In some cases, it may be argued that a given memory task could be performed using procedural or semantic memory. Possession of these forms of memory does not necessarily entail reality testing abilities. I will use empirical research to argue that the form of memory involved is indeed episodic memory, and at times this will involve arguing that the creatures in question lack the concepts required to accomplish these tasks using semantic memory. I will then argue that these creatures lack the requisite mental state concepts to perform reality testing propositionally. I will take the primary behavioral criterion for the possession of a given concept to be successful performance on the sorts of transfer tasks described in the last paragraph.

2.3. Episodic Memory in Nonhuman Animals and Human Infants

I have offered evidence that reality testing exists in rats and mice. Later I will argue that rats and mice do not possess the concepts necessary to accomplish reality testing conceptually. Together these claims will entail that a nonconceptual form of reality testing exists. This, along with the premise defended in chapter 1 – that reality testing is a form of metacognition, entails that nonconceptual metacognition exists. But this bare existence claim would be less interesting if it were entirely limited to rats and mice. My stronger claim is that nonconceptual metacognition not only exists but is a necessary condition for higher forms of cognition. I will offer an a priori argument for this claim in chapter 3, but in the remaining pages of this chapter I will present empirical evidence that reality testing exists in a variety of nonhuman animals and in young infants. I will then argue that these creatures lack the necessary concepts to accomplish reality

testing conceptually. The explicit studies of reality testing in rats and mice have not been extended to other animals, and so this argument will take a different form. I have already argued that any creature able to employ distinct faculties of perception and imagination or episodic memory will require a reality testing ability. If this is the case, then I need only argue that a variety of creatures possess these faculties and the existence of reality testing in those creatures will follow.

2.3.1 Perception in infants and nonhuman animals

It is broadly accepted that all the creatures I will discuss possess a faculty of perception. In rats and mice, it is implied by the existence of the reality testing ability. In the studies I will describe in corvids and infants, much of what researchers try to demonstrate are behaviors that cannot be explained *merely* by perception and require the position of an additional faculty. In this way the perceptual ability of these creatures is assumed. There is really no single argument to be had that rodents, corvids, and infants perceive. This is partly because it is assumed in the literature, and partly because there is no uncontroversial definition of perception. Sometimes perception is defined as conscious sensation. Consciousness is notoriously difficult to assess behaviorally, though researchers do try. Sometimes perception is defined as sensation plus categorization. This will be adequately demonstrated in the studies described below. Some may take categorization as sufficient for the possession of concepts, which would in turn entail that perception requires concepts. This would not necessarily be a problem for my thesis, as I only claim that these creatures lack mental state concepts, but nonetheless I have more general reasons to want to leave open the possibility of broader nonconceptual skills. It would take me too far afield to

defend a particular account of the distinction between sensation and perception. In what follows, then, I will argue that apes, corvids, bees, and infants possess imagination and episodic memory. If we grant that they also perceive, on the basis of common sense and these cursory remarks, it should follow that they are capable of reality testing.

2.3.2. Episodic memory in nonhuman animals

Episodic memory is distinguished from procedural and semantic memory. In semantic memory we recall facts and in procedural memory we develop habits or skills. Episodic memory is memory of events. Episodic memory stores a subject's experiences for later recall (Tulving, 1972). This might seem to imply that episodic memory is necessarily conscious, but as consciousness is not directly measurable in nonhuman animals this issue is typically set aside in favor of other criteria. This leads some researchers to conservatively refer to the phenomena under study as "episodic-like" memory (Clayton & Dickinson, 1998). Memory of an event, it is reasoned, involves memory of "what, where, and when." Any one factor may not be necessary for a given episodic memory – I may remember seeing Nirvana live but not recall the venue. But the three together are often taken as sufficient for episodic memory. This is also questionable. I may recall that Napoleon lost the battle of Waterloo in June 1815 without having any episodic memory of it. Episodic memory is generally taken to be a first-person phenomenon, but even then, I might know facts that adequately locate an event, perhaps told by my mother of an event I no longer recall, without having episodic memory of the event.

Despite these caveats, the generally accepted behavioral criterion for episodic memory in nonhuman animals is an ability to discriminate unperceived stimuli along a temporal dimension

as well as location and physical type – "what, where, and when." Some researchers add other conditions, as we will see, but these too could in principle be accomplished by complex semantic memories. But researchers typically assume that the creatures in question lack the sort of conceptual repertoire that would be required to encode these various factors in semantic memory.

"What, when, where" memory is not the only way of distinguishing episodic memory from other forms of memory behaviorally. If we grant, and as I will later argue, that corvids, rats, and human infants lack the necessary concepts to perform certain tasks using semantic memory, then the most salient challenge is to distinguish episodic from procedural memory. So the question at issue is whether the behavior demonstrates memory of an event or a mere learned ability to respond to general kinds stimuli. If an animal is trained to press levers for food, then returning to press a lever after receiving food cannot be taken as evidence that the animal recalled a particular event. But if the apparatus is novel and there is no training involved, subsequent successful interactions with it suggest more strongly that the behavior is facilitated by memory of a single event. Evidence of this sort is common in studies of adult infants, but also exists in crows.

Finally, the behavioral evidence is boosted by neurological studies. Episodic memory is associated with the hippocampus, and animals whose hippocampal activity is suppressed show reduced performance on the behavioral tasks associated with episodic memory. In what follows I will survey studies that provide evidence for the existence of episodic memory in rats, corvids, and human infants.

Scrub jays search preferentially for cached grubs if they were hidden recently. But if enough time has passed that the grubs have probably decayed, the jays search for cached peanuts instead (Clayton & Dickinson, 1998). Researchers take this to indicate that the jays have a

memory of what is hidden in each location and when it was hidden – what, where, and when. Pigeons, in contrast, are able to make discriminations along the three dimensions separately, but training in one fails to improve performance on another. This is taken to indicate that pigeons fail to combine "what, where, and when" into a structured representation (Skov-Rackette et al., 2006; Watanabe, 2018). Clayton, Yu, and Dickinson (2003) show that the temporal properties of grubs are learned and not innate, since jays whose crickets are consistently and surreptitiously switched for fresh ones do not make the same discriminations.

Rats also show evidence of episodic memory. Babb and Crystal (2006) provided different flavors of food at different locations and time intervals, and rats proved able to remember when and where a type of food was available. When a particular flavor was paired with nausea, rats avoided that location at the appropriate time.

Panoz-Brown et al. (2018) test whether rats remember events in order. They trained rats to choose the penultimate or fourth-to-last odor presented in a random sequence. Performance was significantly above chance and remained so after a one-hour time delay and despite interference by performance of another task in the interim. Long-term memory is dependent on the hippocampus and Panoz-Brown et al. found that when the hippocampus was chemically suppressed performance decreased for the episodic memory task but not for the interference task.

Episodic memory requires not only recall of a particular fact, but the context provided by the original event. Panoz-Brown et al. (2016) trained rats to choose new smells over old, and rats were able to do so when a smell was new relative to the context (an arena with distinguishing markings), even though the smell had recently been presented in a different context. The researchers conclude that the rat was not simply encoding smell familiarity, but the details of the event in which the smell was encountered, which is indicative of episodic memory.

Episodic memory has also been attributed to bees and cuttlefish, but the differences from the research on rats, corvids, and infants is enlightening. Pahl et al. (2007) trained bees in two Y-mazes to choose the arm labelled by a colored grating (yellow and blue; horizontal and vertical) at different times of day (morning and afternoon). In a transfer task, bees had two maze options and the gratings were black and white instead of colored. Bees consistently chose the correct arm in the correct maze for the time of day. This indicates that bees can keep track of the contingencies (sugar water) associated with a particular property (grating direction) at a particular place (maze A or B) at a particular time (morning or night). This is taken to show that bees encode what, where, and when in memory. Cuttlefish learned to move toward visual cues only after 1 hour and 3 hours, respectively, to receive distinct food rewards. The researchers take this as evidence of episodic-like memory, as the cuttlefish retain information about food type, location, and time since last feeding (Jozet-Alves et al., 2013).

One might wonder whether behaviors like those exhibited by cuttlefish and bees are truly episodic memory. Is any sensitivity to environmental spatiotemporal contingencies episodic memory? Episodic memory is traditionally conceived as memory of single events, not necessarily learned patterns, which could be subsumed under procedural memory. I might expect the kitchen curtains to be closed each morning (but not in the evening) and automatically reach to open them. This is not quite the phenomenon we mean when we speak of episodic memory. Transfer tasks are designed to quiet this worry. In all but the cuttlefish experiment, the trained behaviors were then tested with new stimuli that reproduce certain properties of the training stimuli but only those of interest. Thus, when rats identify the second-to-last odor, that odor is different than those used in training. Only the property of being second-to-last is repeated. In this case the rat is clearly sensitive to the temporal order of events, which seems to require episodic

memory. The behavior in crows is convincing as well, as they frequently accomplish the tasks on the first try. In the bee case the transfer task uses a black-and-white pattern rather than a colored pattern. This design is less successful at isolating properties essential to episodic memory. It shows that bees are sensitive to the pattern as well as the color, but this might be accomplished by procedural memory.

Just as one might worry that in some cases procedural memory is sufficient to accomplish a task designed to demonstrate episodic memory, semantic memory might also be sufficient in some cases. Semantic memory, however, requires concepts. For rats to accomplish the smell recognition tasks using semantic memory would require concepts of the relevant odors, some number concepts, and a concept of time. Researchers are often eager to attribute different sorts of concepts to nonhuman animals, but their concept of 'concept' is often much more minimal than the one we are operating with. Mere categorization or discrimination will not be sufficient to demonstrate that the generality constraint has been satisfied.

Do rats have a concept of time? Rats can be trained to press a lever after three seconds for a reward, and even can track their own errors in timing, evidenced by their approaching the appropriate reward window at a rate greater than chance when performance is accurate (within 500ms) (Kononowicz et al., 2022). Such short durations can be represented as environmental affordances, however. Consider how you might approach the task. Your concept of the duration of a second - that it is a very short amount of time, one-sixtieth of a minute, and so on - is unlikely to help you discriminate the relevant half-second differences in the task. It might help if part of your concept of a second is that it is roughly the amount of time it takes you to say "one-and," but this is not useful without presupposing the correct rhythm - you could certainly say "one-and" in less than a second. What would probably happen is that you would develop a "feel"

for the correct length of time. This skill is just the sort of thing that is attributed to nonconceptual content. According to the generality constraint, the researchers would need to demonstrate that the skill developed transfers to other contexts – that the concept 'is three seconds in duration' can be applied to other events. If after training on the lever rats could accurately judge whether other events were three seconds in duration, this would provide more evidence that rats have the concept 'three seconds in duration'. For example, if the rats in Panoz-Brown et al.'s studies, after learning to pick out the fourth-to-last smell were then able to pick out the fourth-to-last tone without training, this would point toward an ability to apply 'fourth-to-last' to a variety of events or objects, and thus a partial satisfaction of the generality constraint. To my knowledge there is no successful application of such transfer tasks to studies of temporal concepts in rats or in corvids.

There is also a general implausibility to the idea that rats form propositional beliefs of the form 'the fourth-to-last smell was rewarded'. But implausibility is not evidence. There is additional evidence for my interpretation. The studies that show that the relevant memories are sensitive to context and that they correlate with activity in hippocampus also push us toward the interpretation that episodic memory is involved. The balance of evidence so far, then, is that rats and crows accomplish the sorts of tasks described in this section using episodic memory.

2.3.3 Episodic memory in human infants

Studies of infants that approximate the "what, when, where" paradigm sometimes take episodic memory to emerge around the same time as mental-state concepts. Hayne and Imuta (2011) asked 3- and 4-year-olds to perform a modified version of the scrub-jay caching task (Clayton &

Dickinson, 1998). After watching the researchers hide a different stuffed animal in three different rooms, the children were asked to recall what was hidden, where, and in what order. The task was both verbal and behavioral (finding the toys). 3-year-olds were less successful at the verbal task and at the "when" component of the behavioral task, suggesting that full "what, where, when" memory does not develop until age 4.

However, other studies push the emergence of episodic memory much earlier. Russell and Thompson (2003) demonstrated a similar ability in 21-month-olds. Infants were shown two toys being placed in two boxes, after which one was removed. The infants, after a 24-hour delay, were able to locate the box with the toy, suggesting that they not only recalled the "what" (the toy) and the "where" (the correct box), but also the "when" in the form of the sequence of placement and removal. Nakano and Kitazawa (2017) used eye-tracking to determine whether infants could anticipate salient events after having experienced them only once. 6-, 12-, 18-, and 24-month-old infants were shown a video of a person in an ape costume emerging from a door and attacking a man. One day later 18- and 24-month-old infants made anticipatory glances at the door from which the ape had emerged, whereas younger infants did not. Furthermore, the glances were more frequent in the scenes leading up to the attack, suggesting some awareness of a "when" component in the memory.

If we forego the "what, when, where" definition of episodic-like memory and adopt other criteria, however, episodic memory appears even earlier in infants. "What, where, and when" are not, after all, necessary for episodic memory, though they might be sufficient if we rule out semantic memory as the source of the ability. The task of creating behavioral tests of episodic memory in human infants differs from the case of adult animals. Infants' motor skills are less developed in some ways than, say, adult rats. But infants often require less training than animals,

and so evidence of memory of events can often be derived in different ways. Much can be inferred from where they look. Infants typically prefer to look at novel stimuli, and this forms the basis of the habituation/dishabituation paradigm from which much of our knowledge of cognitive development is drawn. Indeed, the entire paradigm relies on the fact that infants have some ability to recognize previously presented objects and scenarios. An infant can recognize its mother's voice shortly after birth (DeCasper & Fifer, 1980)[6], its mother's face at 3- to 4-days-old (Bushnell et al., 1989), and its mother's breast milk at 8- to 10-days-old (McFarlane, 1975). This alone is not clear evidence of episodic memory, as it could simply indicate a more general ability to detect environmental kinds. When evidence of recognition occurs after only brief exposure to a novel stimulus, however, this suggests that the memory is of a discrete event. Fagan (1970) demonstrated recognition of black-and-white patterns in five-month-olds after a five-hour delay and (1973) recognition of a face shown only for a couple minutes after two weeks (also in five-month-olds). Kagan and Hamburg (1981), using a similar paradigm with patterned cards, place the emergence of episodic memory ("recognition memory" in their parlance) at 8-9 months. 9-month-olds can imitate observed actions on novel objects even after a 24-hour delay (Meltzoff, 1988). Perris, Myers, and Clifton (1990) showed evidence that 6.5-month-olds can recall a single event (grasping a sounding object in the dark) up to 2 years later. It is not conceptually impossible that procedural memory could develop after a single exposure to an environmental kind, but to allow procedural memory to include such abilities is to blur the very distinction between procedural and episodic memory. If there is such a distinction to be made, then we must allow that recognition after a single event is recognition of that event, not merely of the kinds involved in the event.

[6] This study involved preferential sucking rather than gaze detection.

Schacter and Moscovitch (1984) argue that recognition after a single presentation of a stimulus could be interpreted as a form of priming, distinct from episodic memory. They propose a "parameter filter" method of distinguishing different types of memory in infants. Manipulation of variables that are known to disrupt procedural memory, but not declarative memory (which comprises semantic and episodic memory), could rule out the priming hypothesis. For example, amnesiacs who lack the ability to form episodic memories but are capable of procedural memory for tasks like the Tower of Hanoi puzzle can retain such skills after long periods of non-exposure. Traditionally declarative forms of memory like word recall fade over such periods. Manipulating the retention interval, then, is one "parameter filter" that can distinguish procedural from semantic and episodic memory. Surveying the literature with this method in mind, Schacter and Moscovitch conclude that there is evidence of episodic memory in 12-month-olds but not in infants younger than 7 months.

But later studies using this same principle have shown that manipulation of such parameter filters as retention interval, study time, and context do have effects on imitation and other abilities displayed by infants (e.g., learning to kick to move a mobile) as young as 2-7-months (Hayne, 2004; Rovee-Collier, 1997). In a review of the evidence to date, Mullally and Macguire (2014) conclude that declarative (i.e., non-procedural) memory appears in the middle of the first year of life, possibly as early as two months. Boller (1997) first trains 6-month-olds to kick in the presence of certain pairings of felt shapes by connecting their legs to a colorful mobile by a ribbon only for those pairs. After a three-week interval, infants forget the association and kicking behavior returns to baseline in the presence of the relevant pairs. But a single two-minute exposure to the moving mobile (not connected to the leg) will cause infants to remember the association and increase kicking. Increased kicking was also observed when infants were

exposed to a pair of shapes for only two minutes, later trained on one member of the pair, and then exposed to the other member. Barr et. al (2003) demonstrate a similar ability in 6-month-olds using puppets. The infants were shown two visually distinct puppets (a duck and a cow). Later a sequence of actions was performed on the first but not the second puppet. Infants reproduced the action on the second puppet 24-hours later, whereas control infants who had only been exposed to one puppet did not. The effect could also be obtained not with preexposure to two puppets, but to one puppet and a train merely in view. Experimental subjects later performed the actions on a novel puppet when merely exposed to the train again whereas controls did not. This suggests that the sorts of associations that bind the "what, where, and when" of episodic memory are present in early infancy as well. That these associations can be transferred to the second puppet suggests that they are more flexible than those that occur in procedural memory. Campanella and Rovee-Collier (2005) demonstrated the same ability at 6 months when the puppets had been introduced much earlier, at three months of age.

Adults, on average, cannot recall specific episodes in their lives from earlier than age three and one-half. But if allowed to freely recall events from their past (independently verified by the parents), younger children reveal memories of specific episodes from the first year of life, and as early as one month of age (Tustin & Hayne, 2010). Follow-up questions attempted to distinguish episodic memories from semantic memory ("Who else was there?", "How did you feel?").

The problem of distinguishing episodic from semantic memory also arises for studies in infants. Many of the studies I've reviewed avoid the issue by only distinguishing procedural from declarative memory. There seems now to be general agreement that declarative memory, as opposed to procedural memory, is in place in very young infants – perhaps as early as two

months and almost certainly by six months of age. Those studies that explicitly claim evidence of episodic memory vary, with claims ranging from as early as one month to as late as eighteen months of age. But it can always be argued that the tasks in question were accomplished using semantic memory. I should argue then, as I have done in the case of rodents, that the children in these studies do not possess the concepts necessary to accomplish the tasks using semantic memory. This is an arduous task, given the variety in age of the subjects and the variety of concepts involved in the various tasks. But I can make some general points that weigh in favor of interpreting performance of these tasks as evidence of episodic memory.

Piaget places children younger than two years of age in the "sensorimotor" stage of development, which is preconceptual. I will discuss Piaget's sensorimotor stage in chapter 5 when I argue for a sensorimotor account of reality testing. Studies that attribute concepts to younger children are typically not concerned with distinguishing conceptual content from nonconceptual, "feature-placing" content, and the results tend to be ambiguous between the two. An ability to categorize or to demonstrate surprise when stimuli behave in unexpected ways does not satisfy the generality constraint.

Beck (2012) argues that that ability to discriminate analogue magnitudes must be accomplished nonconceptually, for an interpretation of the ability as the application of number concepts would violate the generality constraint. That is, when a pigeon is able to tell that 38 pecks is fewer than 50, but not that 38 pecks is fewer than 40, it shows that number concepts are not in use. Similar limitations have been demonstrated in a variety of species, including human infants (and adult humans).

I mention Beck's work not because it demonstrates that human infants lack the concepts necessary to accomplish the memory tasks in question conceptually, but to show that the sort of

research necessary to demonstrate that is possible in principle, but difficult in application. It is certainly implausible that young infants possess the rich store of concepts required to form semantic memories like ‘the man squeezed the duck puppet but not the cow when the felt square and circle were presented together’. But to demonstrate that infants lack these concepts would require studies designed to show failure to satisfy the generality constraint for each of these concepts. Such studies do not currently exist. For now, we will have to make do with the general implausibility that a 6-month-old can possess such complex conceptual content.

2.4 Imagination

For some theorists (e.g., Michaelian, 2016), episodic memory just is a form of imagination, and to have one is to have the other. Memory, on this view, is not the retrieval of stored impressions but the imaginative reconstruction of past events. The converse claim, that imagination constitutively involves memory, is certainly true. Since the imagery entertained in imaginative episodes is, by definition, not perceptual it must be retrieved from memory. For my purposes, it would suffice to document either imagination or episodic memory in nonhuman animals and human infants as evidence of reality testing. There is more research on memory. Here I will briefly survey evidence that nonhuman animals and infants possess imagination.

2.4.1 Imagination in nonhuman animals

Mental rotation in pigeons is suggested by studies which show they can match rotated stimuli (Hamm et al., 1997) and predict the location of a rotating image (Neiworth & Rilling, 1987) (see

Hollard & Delius, 1982 for contrary evidence). Pigeons can also match abstract relational properties of a new stimulus (degree difference in orientation of a line) to a previous sample (different degrees of difference in hue) in the absence of the original sample, and this is presumably accomplished by forming a mental image of the original stimulus (Roitblat, 1980). Crows can choose an appropriate tool for a task consisting of a series of subgoals. They consistently choose the tool from among distractors (e.g., using a stick to retrieve a stone from one of several tubes containing different items) that will be appropriate to accomplish a further, unseen task (using the stone to release food from a lever apparatus) (Gruber et al., 2019). Evidence of future-planning has also been shown in jays. After only one trial they learn to cache food items in cages where food will not be provided the next morning over cages where it will (Raby et al., 2007). Like the tool-based planning in corvids, this suggests mental representation of events that are neither perceived nor remembered, since they are in the future, which suggests that they are represented in imagination.

Rats have been attributed spatial representations or "cognitive maps" that facilitate their impressive navigation abilities (Tolman, 1948). Placed in a circular water tank with a submerged platform and guided only by symbols placed on the walls, rats are reliably able to relocate the platform even when the tank is filled with an opaque liquid (Morris, 1981). The existence of cognitive maps has been challenged, and there is evidence that instead rats rely on viewpoint-specific visual representations for navigation (Whishaw, 1991). But in either case, the dominant theory is that rats do use representations for navigation rather than "dead reckoning." Blaisdell (2019) and Waldmann and colleagues (2012) show that rats can represent absent stimuli and rewards, which is arguably a form of imagination. And more importantly, the evidence surveyed

earlier in this section for reality testing in rats and mice shows that rats entertain imagined sensory representations.

2.4.2 Imagination in human infants

Some evidence of imagination in human infants does not speak to reality testing directly, as it involves adding information to what is present in perception rather than distinguishing the two. 8-month-olds show similar EEG activity to adults when viewing a Kanizsa square, and 6-month-olds to a lesser degree, suggesting that in this age range infants develop the ability to connect the gaps between "Pac Man" objects in imagination (Csibra, 2000). Four-month-olds expect, and thus presumably represent, complete objects even when they are partially occluded, as evidenced by longer looking at broken objects when the occluder is removed (Kellman & Spelke, 1983). 4-month-olds, after viewing a display of a ball passing behind an occluder, will look preferentially at displays with the occluder removed that show the ball disappearing and reappearing rather than a continuous trajectory. This suggests that in the occluded condition they represent the trajectory of the object as continuous (S. P. Johnson et al., 2003). Since in each of these cases the infants are representing things that are not there to be seen, it is reasonable to assume that they are represented in imagination. Another possibility is that infants are representing these conditions propositionally, but this would require relatively sophisticated concepts.

Boundary extension is the tendency to expand the boundaries of a perceived visual scene in memory. When asked to draw a previously viewed scene from memory, children as young as 6 years of age will draw a larger background (Quinn & Intraub, 2007). Studies using the habituation paradigm suggest that this phenomenon occurs as early as three months of age.

Infants looked preferentially at closeups of previously viewed scenes over wider shots, suggesting that the wide shot was familiar due to boundary extension in memory.

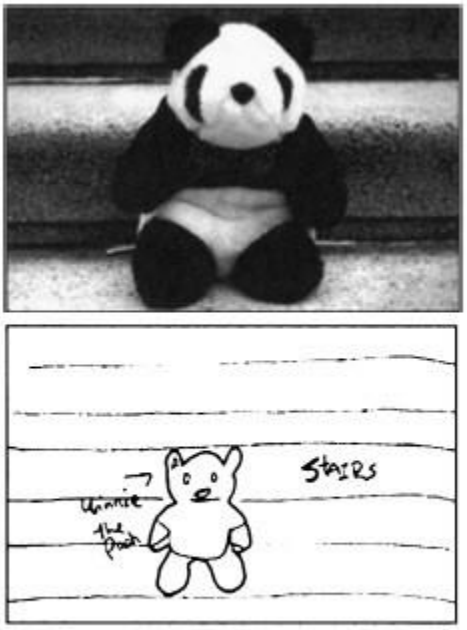

Figure 1. A single-object scene in which the object is not cropped by the edges of the picture (close-up view) and a representative drawing from memory (from Intraub et al., 1996, Figure 1, panels A and B).

Figure 2.1. Boundary extension (Quinn & Intraub, 2007, p. 325).

Much of the study of infant imagination is in the context of pretense. Children begin to engage in pretend play around 18-months (Friedman & Leslie, 2007; Lillard et al., 2010), much earlier than they are typically able to pass the false-belief task. It is probable, then, that the ability to distinguish between imagination and perception, as is required in pretending, is prior to theory of mind.

And there is evidence that the sort of imagination required for pretend play emerges earlier. Piaget (1962) reports the anecdotal observation of his 15-month-old daughter putting a blanket over her head and pretending to go to sleep. Onishi and colleagues (2007) produce

evidence that 14-month-olds understand the distinction between imagination and reality. When observing an adult pretending to pour water in one of two cups they look preferentially when the adult drinks from the cup that remains empty in the pretense. The effect extinguished when the cups were replaced with shoes but returned after a single trial. This suggests that the children could distinguish imagination from reality and the imaginative episode was disrupted by the novelty of the shoes. The children were not hallucinating water, they were imagining. Once they caught on to the shoe gag, they continued imagining.

2.5. Reality testing in nonhuman animals and infants is nonconceptual

I have argued that animals and human infants engage in reality testing. In the case of rats and mice, there is research that directly supports this conclusion. In other animals and human infants, the ability is implied by the ability to entertain episodic memory or imagery. If a creature can entertain these various types of experience without confusing memories or imagery for perception, then this implies a reality testing ability. In this section I will argue that this ability is nonconceptual. That is, that it is not accomplished using inferences from a theory of mind or ascent routines which involves mental state concepts. This will be established if I can argue that rats, corvids, and human infants lack mental state concepts.

2.5.1 Rats and corvids do not have a theory of mind

Some mental state concepts have been attributed to apes, corvids, and toddlers, but very little has been offered as evidence for such an ability in rats. Imitation is sometimes taken as evidence of

theory of mind, as it suggests the attribution of goals or intentions to the imitated. Imitation has been demonstrated in rats. After observing a conspecific push a joystick to obtain food, rats will do the same (C. M. Heyes et al., 1992). However, the inference from imitation ability to theory of mind is questionable. Observation of the relation between behavior and effect is likely adequate for imitation without attributing any mental states to the imitated (C. M. Heyes, 1998). Rats have also been shown to choose to free trapped conspecifics, even choosing to do so instead of obtaining food for themselves. This has been taken as a demonstration of empathy (Bartal et al., 2011). However, the choice needn't be based on the attribution of suffering or some other mental state to the other - physical confinement is a bad thing in itself. Self-directed metacognitive abilities have also been suggested in rats. This ability is tested by offering the option to decline a given trial in a memory task for a minimal reward when accurate performance would produce a large reward and inaccurate performance none. This is sometimes taken as of form of "knowing whether one knows" (Foote & Crystal, 2007). This point is the subject of much controversy, with scholars arguing that the ability is first-order (Carruthers, 2017) or nonconceptual (Proust, 2013). Rats also frequently fail at these sorts of metacognitive tasks (Carruthers, 2008; Smith, 2007). In any case, none of *this* potential evidence for the existence of mental state concepts in rats would explain their reality testing ability. Conceptual reality testing would require very specific concepts - namely concepts of perception and memory or imagination. There is no evidence to date that rats or mice possess these concepts.

A concept of perception has been attributed to corvids. Given the choice between a noisy caching location and a quiet one, jays will choose the quiet one when they are not observed by conspecifics, but do not bother when they are observed. When observing others caching, they vocalize less. This suggests that they attribute perceptual states to conspecifics (Shaw & Clayton,

2013). Small green bee-eaters also exhibit hesitance to enter their nest when a human observer watches, but not when the human is present but unable to see the nest (Smitha et al., 1999). Critics, however, have pointed out that the behavior could be explained by the birds' tracking correlations between head cues and consequent behavior (Bugnyar et al., 2016; C. Heyes, 2015).

The evidence that corvids possess the concept of perception is controversial. But even if we grant it, it does not suffice for reality testing. An ability to distinguish being seen or not does not suffice to distinguish various mental state types like perception, memory, and imagination in oneself. For the latter ability – reality testing – to be accomplished conceptually in the corvid would require a much more complete theory of mind than is evidenced in these studies.

2.5.2 Theory of mind in infants

Theory of mind is typically attributed to humans around age 4, when they pass the false belief task (Gopnik, 1993). I have surveyed evidence that shows capacities for imagination and episodic memory develop much earlier. A concept of perception, however, has been attributed to younger children. As in the case of corvids, however, we will see that this evidence is ambiguous and insufficient to attribute propositional reality testing abilities to infants.

In the first year of life infants learn to follow another's gaze, though the precise time is controversial (Butterworth & Cochcran, 1980; Hood et al., 1998; Morissette et al., 1995). Gaze following may only indicate a learned correlation between the direction of a gaze and objects of interest, or it may be an innate mechanism (Moore & Corkum, 1994). Butler, Caron, and Brooks (2000) intermittently placed opaque screens between the gazer and the object. They found that 18-month-olds looked preferentially when the gazer's view was not blocked, whereas 14-month-

olds showed no preference. This suggests that the older infants understand more than mere correlation between head position and object, they understand that visual access can be blocked. Franco and Butterworth (1996) show that when 12-16 month-olds point at an object they check the gaze of the person whose attention they are directing.

Flavell (2002) posits a difference between a 'level 1' understanding of perception and 'level 2'. Level 1 is attained by 2 ½ to 3 years of age and involves an understanding that others cannot always see what the child sees. Level 2 requires an ability to project the perspective of the other, for example understanding that a picture that appears right-side-up to the child will appear upside down to someone seated across a table. Full level 2 understanding does not develop until 4 ½ to 5 years of age.

Even if we accept the earliest age at which a concept of perception is claimed and place it around 12 months of age, this is much later than the development of episodic memory. There is less evidence for early development of imagination, so it is possible that the possession of a minimal concept of perception is coeval or even slightly earlier than the development of a faculty of imagination. Of course, if one understands episodic memory as a form of imagination, then we can push imagination much earlier. And my points regarding concepts of perception in corvids apply here as well. A concept of perception alone is insufficient for reality testing. The balance of evidence shows that infants require a reality testing ability earlier than they acquire the mental state concepts required to accomplish reality testing propositionally.

2.6 Conclusion

I have argued that reality testing is a nonconceptual ability. This is supported by empirical studies of reality testing in rats and mice and by the commonsense conclusion that any creature capable of both perception and imagination or episodic memory must have some means of telling them apart. I have argued that rats and mice, corvids, and human infants possess imagination or episodic memory. I have also argued that these same creatures lack the concepts required to accomplish reality testing conceptually. I conclude that reality testing is accomplished nonconceptually in these creatures. This chapter has focused on empirical evidence for the existence of nonconceptual reality testing. In the next chapter I consider some philosophical accounts of the nature of primitive mentality and argue that these presuppose a form of reality testing.

Chapter 3. Reality testing is necessary for conceptual content.

The generality constraint sets a relatively high bar for conceptual content, which in turn leaves a good deal of room for nonconceptual forms of mentality. The generality constraint defines a kind of structure for conceptual content which distinguishes between objects and their properties. This, many philosophers and psychologists have argued, requires *objectivity* – the ability to conceive of mind-independent objects. I will review several of the more prominent accounts and show that in each reality testing plays an implicit but essential role in the development of objectivity, and thus in the acquisition of conceptual content. This not only supports my claim that reality testing is nonconceptual, but also shows that reality testing is fundamental to forms of cognition like belief.

3.1. Objectivity is a necessary condition for conceptual content.

The generality constraint requires a form of compositionality that is not available to creatures capable only of nonconceptual content. This, in turn, allows for inferential abilities not available to simpler creatures. Feature-placing content allows for eating to be afforded on one occasion and avoidance afforded on another. Feature-placing content does not allow one to conceive of the same *object* as either edible or to be avoided. Call this ability to conceive of objects in a way that allows the generality constraint to be satisfied *objectivity*. If we accept the generality constraint, objectivity is a necessary condition for conceptual content.

Morgan's Canon (Morgan, 1894), a widely accepted methodological guideline in cognitive ethology, states that we should not attribute more complex mental abilities to a creature than are necessary to explain its behavior. Consider a creature endowed only with nonconceptual content. Call this creature 'Nan'. If we are entitled to attribute mental content to Nan at all, be it conceptual or nonconceptual, she must exhibit behavior that cannot be explained purely in terms of stimulus and response. If Nan eats every Snickers bar she encounters, then the behavior can be explained using only variables for stimulus (the Snickers) and response (eating). If Nan's behavior is more complex, however, such that her behavior cannot be predicted using only external variables, we may be driven to posit mental variables that affect the behavior. The most straightforward way of explaining such behavior is by attributing to Nan a kind of practical reasoning – Nan will not eat the Snickers if she does not desire to eat, believes the Snickers is poisoned, and so on. In its nonconceptual version, however, such reasoning must not employ the sorts of inferences that are contingent on satisfying the generality constraint. If Nan directly perceives things as 'to be eaten' or 'to be avoided' on different occasions, then she may eat or

avoid the same Snickers bar on different occasions, but she can't conceive of that *same* Snickers bar as having those different properties. Nan can't conceive of an *object* corresponding to the Snickers bar at all, if a conception of an object is a representation to which predicates can be applied in the way the generality constraint dictates.[7] As we have seen in chapter 2, feature-placing content can govern behavior without the inferential abilities afforded by conceptual content by virtue of its essential reference to opportunities for action. This has the benefit of obviating the need for instrumental beliefs such as those involved in the inference from 'I am hungry' to 'I should eat that' ('food satisfies hunger', 'that is food', etc.). An environmental feature is presented to the subject as something that invites action. But this entails that there is no place for a sharp distinction, from Nan's perspective, between her and her environment. Her perceptual world consists of invitations to exercise her own motor capacities (though this does not entail that she has a conception of herself in any robust sense). But because her mental content is indexical in this way, it cannot represent objects with properties independent of her and her motor capacities. So, Nan does not possess the kinds of standing, instrumental beliefs of the form 'eating satisfies hunger' or 'climbing reaches tall food' that are required for the types of inferences we attribute in propositional forms of psychological explanation.

For a representation to qualify as of an "object" in a way that can satisfy the generality constraint, it must, in a sense, fill a particular grammatical role. Even if certain predicates are never in fact applied to it, it must be conceived as the sort of thing to which any arbitrary property could be applied. None of this requires an explicit knowledge of grammar on Nan's

[7] Here, and at many other points, I must ignore many arguments against the generality constraint and its entailments. Couldn't Nan be able to conceive of an object that, as it turns out, she can only classify under one predicate? This seems unlikely, for anything that qualifies as an object in any robust sense will be conceived as having at least two spatial dimensions, and thus two properties. My discussion of Burge will show that, even if we reject the generality constraint, reality testing is required for more minimal accounts of objectivity.

part. Rather, it requires a particular conception of ontology on her part that corresponds to a certain grammar in her psychology. Thus, her mental representations might satisfy the generality constraint even if she *in fact* never applies non-indexical properties to perceived objects. So what is the connection between generality and objectivity, where the latter is conceived as self-independence? This link originates in Strawson (1959), who argues that a conception of perception-independence is the basis of objectivity. Even though a standing representation of an affordance is not in principle independent of the self, detaching the representation from perception introduces a kind of generality. To remember an eating feature and begin to navigate to it I must conceive of it as currently existing at another place than I am now, currently unperceived by me. This is not yet the full generality of the generality constraint, but it allows representations to be detached from a particular time and place, if not from a particular subject. In this way features begin to be conceived as something more metaphysically robust. This ability to conceive of mind-independent entities is called *objectivity*. There are other necessary conditions for mental content to satisfy the generality constraint, but objectivity will be my focus in this chapter. I will argue that conceiving of something as existing unperceived requires either memory or imagination, and thus a reality testing ability.

Objectivity thus conceived can occur at the level of nonconceptual content. That is, features can be conceived as existing unperceived just as well as full-fledged objects. But objective features are not yet objects, as they do not yet have the grammatical structure to allow inferences from ‘x is round’ ‘x is green’, ‘round, green things are edible’, ‘eating is desired’ to eating behavior. There may be further conditions on fully conceptual content besides objectivity, such as some conception of substance, or solidity, or of the causal relations that define substance and solidity. Even if it is not sufficient, however, objectivity is a necessary condition for

conceptual thought, for it is the beginning of the kind of structure that detaches a conception of a feature from the here and now.

There are other conceptions of objectivity, but these are not directly relevant to conceptual content. Objectivity is sometimes defined as being in-principle perceptible by other subjects. But this is irrelevant. I might be the only creature in the universe and still conceive of objects as independent of my perception. Strawson (1959) connects objectivity with the ability to reidentify a feature as having been previously perceived. But I may conceive of something as independent of my perception even if I never see it again, or never could see it again. The key is to conceive of it as a thing with a kind of independent existence.

Many authors have offered accounts of objectivity and I will consider several of them. I begin with Bermudez because he comes closest to recognizing the gap I want to address. I will argue, however, that he does not address it adequately. I will then survey a number of other accounts and point out where they run into the same problem. I will not endorse every aspect of each account, and in fact they contradict each other on some points. My strategy here is to survey the most prominent accounts of objectivity and point out that all of them imply that reality testing is nonconceptual. This suggests, even if it does not prove beyond doubt, that whatever the correct account of conceptual content turns out to be, reality testing will play a fundamental role.

3.2. Bermudez on objectivity

Bermudez's (1998) account of the development of objectivity from feature-placing content comes in the context of an account of self-consciousness. Objectivity, he argues, is a necessary condition for self-consciousness. Thinking of oneself requires, at minimum, a distinction

between the self and the non-self. A creature capable only of feature-placing content, which makes no distinction between self and world, would lack this ability. To make a distinction between self and world, Bermudez argues, one must be able to conceive of perception-independent things. These "things" may be conceived as perception-independent but not yet qualify as conceptions of physical objects, and so he refers to such self-independent entities as objects*.

To return to our creature endowed only with nonconceptual content, Nan, there are several ways that she might conceive of the distinction between self and world. In somatic proprioception her body responds to her will, whereas the rest of the world does not. Responsiveness to the will, then, can form a basis of the self-world distinction. There are also minimal sorts of self-specifying information available in Gibsonian "ecological perception" (Gibson, 1979). The invariant information I get from my nose in my field of vision contrasts with the varying information that is not coming from my body. But these ways of distinguishing the self, according to Bermudez, are relatively impoverished. The self, he claims, is fundamentally a contrastive notion - to have a rich conception of the self, one needs a rich conception of the non-self. This richness is not mere informational capacity. Dretske's account of nonconceptual content, for example, holds that conceptual content is informationally impoverished relative to nonconceptual content. Rather, we need to be able to conceive of things as susceptible to various types of predication – we need grammatical objects.

Following Strawson, Bermudez takes the ability to reidentify a feature as numerically identical to one previously perceived as a crucial step toward objectivity. Objectivity, and crucially for Bermudez the conception of things other than the self, requires the ability to conceive a distinction between one's experience and that of which one has experience. That is,

one must be able to view one's own experience as tracing a particular path through an objective world which contains many such possible paths. Nan begins to make this distinction between her own path and others when she can recognize a feature as the same as one previously encountered—a place where the paths intersect. In order to be able to reidentify a feature not as a new, qualitatively identical one, but as a numerically identical one previously perceived entails conceiving of that feature as having existed in the interim between perceptions of it.

Bermudez and Strawson sometimes speak as if reidentification is logically necessary for objectivity, but as I have described it, it is merely sufficient. I think this is the better interpretation. Reidentification of features entails a conception of an objective world independent of Nan, but it is not logically required. Mandik (1998) points out that another conception is also possible. Nan could exist in an environment in which nothing is ever encountered more than once, and yet each thing that is encountered is conceived of as the sort of thing that *could* be reencountered. In this case reidentification would not be a necessary condition for objectivity. The sufficiency condition might be questioned as well. If Nan conceived of mind-independent entities that appear and disappear from existence willy-nilly, reidentification wouldn't entail continued existence unperceived. Strawson is of course interested in *our* sort of objectivity, and perhaps reidentification is both necessary and sufficient for that, but Bermudez and Strawson do not make it clear why it is necessary at *this* stage. Later I will consider Evans' critique of Strawson and will explore what I think is a better account of the necessary conditions for objectivity. This account would not require reidentification of particulars but does point toward the necessity of metacognition. Let us accept Bermudez's account for now, perhaps as a plausible account of how objectivity develops in this world. How can Nan accomplish this act of recognition of identity across time? At a minimum this requires a capacity of memory.

Bermudez claims that "unconscious" memory will not suffice for reidentification of features. His distinction between conscious and unconscious memory maps closely onto the distinction between procedural and episodic memory. I prefer the latter, as consciousness tends to mystify things. To recognize a feature as identical to one previously perceived, it is not sufficient that one's later behavior merely be modified by the previous perception, as would be the case in procedural memory. Recognition, he claims, is something more. Reidentification of a feature requires two representations of the feature, one in episodic memory and one in perception.

Bermudez does not recognize it explicitly, but reidentification is a problem that only metacognition can solve. For identification to take place between the object of a memory and the object of a perception, it is not enough to note that $f = f$, where f is some feature. This trivial sort of identification of an object with itself does not to suffice to identify the object as having been *previously* perceived. This is the job of memory. Reidentification is to conceive that Perceived-f = Remembered-f. This requires some conception of perception and memory, the ability to entertain the two contents simultaneously, and the ability to at the same time keep them distinct. This is reality testing.

Bermudez also implicitly seems to recognize this. He suggests that conscious memory is distinguished from other representations by the presence of a "feeling of familiarity." This could be interpreted as a sort of account of reality testing. But it must be developed. In chapter 4 I will consider various ways Bermudez's brief remarks might be developed into an account of reality testing and will conclude that in each case the account fails. For now, we can conclude that Bermudez's account of objectivity as a necessary condition for conceptual content implies that reality testing is a necessary condition for objectivity.

3.3 Strawson on objectivity

Although Strawson is in many ways the father of the modern debate on objectivity, his (1959) treatment of reidentification of particulars, which he takes to be necessary for objectivity, does not consider the role of memory in any detail. Nor is his argument framed in terms of conceptual and nonconceptual content. Though he coins the term 'feature-placing content', none of his discussion of reidentification as a condition for objectivity is couched in terms of it. Bermudez's account is in many ways the natural extension of Strawson's account into these areas. There is, however, other work by Strawson that is explicit on the issue.

Strawson (1982) considers the concept of imagination in Hume, Kant, and Wittgenstein, and he describes it as playing a role much like memory does for Bermudez. When dealing with these historical figures issues of interpretation are difficult, so I won't attempt to argue that Strawson's reading is correct, only that it reveals important aspects of Strawson's conception of objectivity. I will focus on his treatment of Kant. He reads Kant as claiming that imagination is necessary for "seeing-as." He describes seeing-as in two ways — (1) seeing an object as a member of a kind and (2) seeing something as a physical object at all. My discussion so far has been focused on (2), but (1) is also a necessary condition for conceptual content as defined by the generality constraint. That is, to see an object as a member of a kind is to attribute a property to it in the sense implied by the generality constraint.

To see something as a dog, says Strawson's Kant, is to see it as potentially barking and moving. This potential barking and moving, he claims, is represented in imagination. But, he points out, we don't typically see the dog as *actually* barking and moving, though we might

make this mistake if we are "particularly timid" (Strawson, 1982, p. 89). In seeing something as a dog we entertain an amalgam of perceptual content and imaginative content. Typically, we are able to keep these distinct, but this ability is not infallible. It should be apparent that what is needed here is a form of reality testing. Strawson's account tends toward the metaphorical, describing the imaginary aspects as "saturating" or "alive in" the perception (p. 89). When speaking of reidentification of an object as one previously perceived, he similarly describes past perceptions as "alive" in the current perception of the recognized thing, facilitating the reidentification (p. 89). This presumably is something like perceiving a newly clean-shaven friend and identifying him with the familiar, bearded person. Images or memories of his past bearded version are present in or alongside the perception. The metaphors are just that, but we can assume that the images or memories oughtn't be too tightly woven into the perception if one is to avoid mistaking the man's actual, current state as a bearded one. It must be the case that we do keep the two types of representation distinct, superimposing the imagined or remembered on the perceived in the act of recognition or categorization, but not confusing the imaginative and the perceived elements in normal cases.

If this sort of seeing-as is necessary for conceptual content, Strawson's account implies that reality testing is necessary for seeing-as, and thus that reality testing is nonconceptual.

3.4 Dummett on proto-thought.

In chapter 2 I introduced Dummett's (1993) account of nonconceptual content, which he calls "proto-thought." His conception is importantly different than that of the other authors I discuss here, as it does *not* generate the need for reality testing. Proto-thought does not make a

distinction between perception and imagination or memory, though it does invoke separate modalities of content. I mention him briefly to illustrate the boundaries of the problem – at what point does reality testing become necessary? Dummett describes proto-thought in affordance-style terms as perception overlaid with imagination of the behavioral possibilities afforded by the thing perceived. But this description alone, which invokes both perception and imagination, does not yet generate the problem for there is no need at this stage to distinguish the two forms of representation. That is, even if the perceived thing's edibility or to-be-avoidedness is added by the imagination, from the creature's point of view it is simply perceived content. If the creature acts on the imagined edibility as if it were perceived, that is perfectly fine, since the imagined content is only presented when the feature is present. One may wonder why such a creature needs a distinct faculty of imagination, if it is not used as such. But this does set the boundaries of the problem. It is not the existence of distinct modalities of content per se that requires metacognition, but rather the ability to treat them *as* distinct.

3.5. Campbell on objectivity

John Campbell (1995) holds that to conceive of something as an object is to conceive of it as having certain intrinsic causal powers. These powers can, for our purposes, be roughly identified with Lockean primary properties. Like Strawson, and unlike Bermudez, his account of reidentification is of objects, not of features. As such it does not make reidentification a necessary condition on objectivity. And so even though his account of reidentification requires both memory and imagination - one imaginatively reconstructs the causal process by which a

remembered object might reach its current form (p. 6) – that does not entail that reality testing is a necessary condition for objectivity.

Campbell (1994) does, however, address the necessary conditions for objectivity when he considers the nonconceptual abilities of rats. He describes studies in which rats reliably locate a visually imperceptible platform submerged in a pool of water. The arena in which the pool is located has symbols located at various points which rats use as landmarks to locate the unseen platform. Campbell proposes that the rats are capable only of feature-placing content, conceiving of the landmarks not as objects but as features. The platform must have some minimal causal structure for the rat, since the point is to causally interact with it, but Campbell claims it is not yet a physical object for the rat. The platform itself is represented non-objectively, and thus nonconceptually, and crucially its location is not represented in perception, since it cannot be seen. Campbell does not spell this out, but the rats must be representing the platform in some non-perceptual modality such as memory or imagination. So here again we appear to have, at the nonconceptual level, an ability to distinguish two types of representations. The rat does not dive immediately upon activation of the memory of the platform – it knows that it is a different sort of representation than perception. The memory guides the rat, and thus is present, but does not activate the same sorts of behaviors that a perception of the platform would (diving). It is crucial that the rat not confuse perception with imagination or it will never reach the actual platform. The ability to find the platform implies a reality testing ability at the nonconceptual level.

3.6. Grush, Cussins, and O'Keefe on objectivity

Grush (2000) claims that objectivity is achieved when a creature is able to identify a feature represented as existing in an egocentric bodily space with a feature represented as existing in an allocentric space. Bodily space is represented in perception while allocentric space is represented in the imagination. I will offer a more detailed exposition of Grush's account in chapter 5. For the moment it suffices to point out that the identification of the same feature in perception and in imagination requires reality testing. Thus, Grush's account of objectivity also implies that reality testing is prior to objectivity. Grush is ultimately noncommittal on the conceptual-nonconceptual distinction but concedes that his notion of content in this context maps onto what is typically called nonconceptual content (2000, p. 60). So, on Grush's account, reality testing is nonconceptual.

Cussins' (1992) account of objectivity and nonconceptual content makes use of the notion of a "cognitive trail" which corresponds roughly to a feature or affordance (again, more on Cussins in Chapter 5). His account makes less explicit use of imagination or memory, but imagination is mentioned as a faculty that also tracks "cognitive trails." In context, this does not appear to be a Dummett-style use of the imagination in which imagination is inextricably fused into the feature-placing content. Rather, its use in the context of navigating trails tracks more closely with Campell's or Grush's use of imagination, in which the locations of unperceived features are represented in imagination. Again, reality testing is implicitly assumed at the nonconceptual level.

Even the biologist, O'Keefe, when he attempts to account for the possibility of objectivity in rats, concludes that an objective, non-egocentric conception of things in the environment would require an allocentric map housed in imagination (1994, pp. 39–40). His comments on the subject are brief and appear to presage some aspects of Grush's account, and as

such produce the same problem of distinguishing imagined and perceived content and the same need for reality testing prior to objectivity. Again, reality testing is required for objectivity and is thus nonconceptual.

3.7 Evans on Objectivity

Two more authors require slightly more sustained treatment because they offer a counterpoint to the accounts I have described thus far. Evans (1985) argues against much of what Strawson, and consequently Bermudez, claims about the necessary conditions on objectivity. Burge (2010) takes a radically different approach. I will argue that reality testing is a necessary condition for objectivity on these accounts as well.

Evans takes issue with several important points assumed by Strawson, and consequently by Bermudez, in their accounts of objectivity. Bermudez invokes memory as necessary for objectivity because his account of the ability to conceive of self-independent objects assumes, with Strawson, that objectivity requires the ability to conceive of one's experience as tracing one of many possible spatiotemporal paths through a self-independent world. Memory preserves the moment when one's own path intersects with that of another object. But we have also noted that there is more than one way to conceive of objects as existing independently. Evans points out that distinct spatiotemporal paths are not necessary to conceive of objects as existing independently of our perceiving them. An object might occur at the same time and place as my current perceptual episode, but I might be perceptually unreceptive to it (for example if my eyes are closed or I am inattentive). The ability to reidentify things across time or to identify them as existing in both an egocentric and allocentric space might be *sufficient* for conceiving them as

existing unperceived, and it may even be necessary on any psychologically or biologically plausible account of how humans typically do it. But, as Mandik also points out, it's not necessary. It is logically possible to conceive of an object as independent of my perception even if I have never seen it before and may never see it again.

The reidentification requirement, for Strawson, was intended as a vindication of the Kantian thesis that a conception of space is a necessary condition for objectivity. Evans (1985) questions this Kantian thesis and describes a possible account of objectivity that does not rely on reidentification. He describes objectivity as "… be[ing] able to understand the hypothesis, even if, in fact, he never believes it to be the case, that the phenomena of which he has experience should occur unperceived (p. 261)." One way to understand this is to conceive of some relation such that *if* one were that relation to the thing it *would* be perceived. "In the formulation of the condition there lies a theory, or the form of a theory, of perception (p. 262)." Evans is explicitly talking about feature-placing content here, so this theory cannot be very sophisticated, but it must be of the form "If it is true that it is now phi-ing, then it must be the case that if the condition is satisfied, he will perceive it to be phi (p. 263)." If the condition turns out to be necessarily spatial, then Strawson's Kantian approach is vindicated.

One possible condition that has already been suggested is that of receptivity: If it is true that it is now phi-ing, then it must be the case that if I am receptive, I will perceive it to be phi. In addition to being possibly false in the case that one is receptive but not near the phi-ing, this condition is potentially circular. To be receptive is, arguably, simply to be ready to perceive. If it turns out that any non-spatial account is similarly circular, then Strawson is vindicated. Or the circularity might be that to be receptive is to assume an objective world one is receptive to. Evans argues that in this form the condition is not circular but holistic, and that in any case any

spatial account of objectivity will be similarly holistic. To what are we receptive? Evans argues that this sort of theory of perception will require a distinction between primary and secondary properties. Our experience can then be distinguished from the cause of that experience. But Evans carefully argues that our theory of primary qualities cannot be constructed from experience of secondary qualities. This would not provide us with true objectivity, but simply a list of counterfactuals involving experience. Evans leaves us here with little to say about how objectivity could be possible. It is unclear whether his points are meant as a refutation of the Kantian approach to objectivity or as offering another possible route. And surely we are creatures who gain objectivity without direct access to the primary qualities of objects. Evans' project could be completed if he could explain how our conception of primary qualities comes about without reference to experience. This is something that Burge attempts, and it is to him that I now turn.

3.8. Burge on objectivity

Burge (2010) argues that most accounts of objectivity require abilities that are far too sophisticated. In part this is because he only considers accounts that assume, or that he interprets as assuming, the possession of propositional content. Our discussion has shown that there are many accounts of objectivity that begin with nonconceptual, feature-placing content. Indeed, if objectivity is a necessary condition for conceptual content, then this must be the case. Rather than critique Burge's analysis of the literature on objectivity, however, I want to consider his own proposal, which might be interpreted as lying in the direction that Evans sketches out. This is ironic since Evans is one of Burge's primary targets of critique. Evans argues against an

empirical route to objectivity, in which extra-experiential objects are constructed out of experience. Evans is not explicit about what follows from this, but one option is an innate mechanism that guarantees objectivity.

Burge argues that the subpersonal and, he claims, non-representational "formation laws" involved in psychological phenomena like size constancy are sufficient to guarantee objectivity. When a percept is underdetermined by sensory stimulation, say when retinal stimulation is consistent with both a small object in the foreground or a large object in the background, psychologists have posited and experimentally supported the existence of heuristics that determine how the sensation will be interpreted and thus the nature of the perceptual experience. Thus, he claims, science has determined that objectivity occurs in early visual processing before any form of representation occurs. Of course, there are scientists who would refer to operations in early visual processing as involving representations of, say, edges, but Burge disputes those scientists (while nonetheless holding that the verdict of science on matters like size constancy and its implications for objectivity are beyond dispute by philosophers).

Burge claims that objective representation in perception is the most basic form of representation (though he allows that perceptual memory and perceptual imagination are not objective, and implies, as we shall see, that they are nearly as basic). As such he has no use for feature-placing representation. But this also means that he must offer a different account of primitive cognition. Recall that feature-placing content allows for action without instrumental belief by placing affordances for action in the perceptual state. Burge then seems to be committed to the claim that all creatures capable of perception are also capable of propositional inference.

But I needn't critique Burge's account to make my point. In fact, he strengthens it by demonstrating the same need for metacognition within a theory of objectivity that is vastly different than the authors I have discussed so far. The need for reality testing arises when the same content is represented in distinct modalities which must be coordinated yet kept distinct. There are various points in Burge's account where this is required. Burge holds that perception of three-dimensional bodies, as opposed to perception of mere shapes, requires a coordination of perception and perceptual memory. Likewise any perceptual representation that involves a temporal element, such as perception of motion, involves memory. He also holds that memory is involved in the transition from perception to perceptual belief. There are even passages that suggest memory is necessary for perception simpliciter (p. 527). Given his examples, it also seems clear that the kind of memory involved is conscious, episodic memory. To take one example, the task to be performed in perception of motion is the connection of different temporal stages of a consciously perceived event. This is not procedural memory, as it is memory of a particular event, and not semantic memory because it is imagistic. I have already argued that any creature capable of both perception and episodic memory requires a reality testing ability, and thus a capacity for metacognition. So even if Burge is right about how basic objectivity is, the need to entertain the same content in distinct modalities, and thus the need for metacognition, arises at that stage. Indeed, it seems likely that any account of primitive mentality will involve a need to entertain the same content in memory and perception, and thus involve a need for metacognition.

I have argued that a wide range of authors concerned with setting out the necessary conditions for objectivity, i.e., the ability to conceive of mind-independent things, leave a gap in their accounts. These authors, either explicitly or implicitly, all require more than one mode of

entertaining a given content, be it perception, memory, or imagination. But they typically offer no account of how a creature endowed with only nonconceptual content could manage to distinguish these various mental modalities. In chapter 5 I will offer such an account, but first I will discuss extant accounts of reality testing and argue that they are insufficient to do the job.

Chapter 4. What reality testing is not.

While there is little philosophical research devoted exclusively to reality testing, there are accounts of the ability to distinguish among the propositional attitudes. This is a more general phenomenon than reality testing, as it includes attitudes like belief in addition to perception, memory, and imagination. Many such accounts assume the possession of conceptual content, so they won't do for the forms of reality testing I am concerned with. But there are other accounts that do not assume conceptual abilities and might be extended to explain reality testing. Here I will argue that the extant accounts fail.

4.1. Conceptual reality testing

There are, of course, conceptual forms of reality testing. I might wonder to myself whether the sound I am experiencing is perceived or imagined. I might reason that I am alone in the house, have not left the stereo on, and conclude that the sound is imagined. When we engage in this form of reality testing, we make inferences from propositional states. I will briefly survey the most prominent propositional accounts of reality testing and show that they can't be used to

explain nonconceptual reality testing. I will then move on to the accounts that might appear more promising.

4.1.1 Theory-theory

Theory-theorists (Gopnik, 1993) claim that self-directed metacognition, or introspection, is no different in principle than other-directed metacognition, or mindreading. We all possess a theory of mind we use to make inferences from observed behavior to draw conclusions about the sorts of mental states people, including ourselves, are in. The point is best illustrated, and is largely supported, by evidence from research on child development. This sort of study presents children with a scenario in which a character, call her Lisa, sees an object, say a cookie, being placed in a container. While Lisa is not looking, but in view of the subject, the cookie is moved to a new container. When asked where Lisa will look for the cookie, 5-year-olds will say the original, now empty, container. Younger children will say the new one. This is taken to show that younger children lack the concept of false belief – which is to say that they lack the concept of belief altogether. To predict and explain the behavior of others, children must conceive of mental states that can fail to track the state of the world. That is, they must develop a theory of mind. Once this theory is in place, children can apply it to their own mental lives. Theory-theory, then, holds that a prerequisite for metacognition is the acquisition of mental concepts and their integration into a theory of mind that can be used to make inferences about one's own mental states and the mental states of others.

When I determine whether a sound I experienced was perceived or imagined, I might employ a theory of mind. Since perceptions have external sources, I perform a sort of disjunctive

syllogism, eliminating the possible sources of the sound. If I fail to locate one, I conclude that the experience does not have an external source. This means my sound experience has an internal source, which my theory tells me is imagination. Thus, the sound is imagined. Adult humans, and potentially some apes, may perform some reality testing in this manner. But the creatures surveyed in chapter 2 do not, including young humans. It is also implausible that adult humans perform *all* reality testing in this way. I am certainly not conscious of making such inferences very frequently. If instead the inferences are made unconsciously, this would be computationally laborious, given the sheer amount of reality testing that I must perform from moment to moment. So theory-theory is not sufficient to explain nonconceptual reality testing.

4.1.2. Transparency theory

Transparency theory (Byrne, 2018) holds that I come to know my own mental states by coming to know first-order propositions about the world. By performing an "ascent routine" that applies epistemic operators to those first-order propositions, I produce metacognitive states. Do I believe that the cat is on the mat? First, I ask whether the cat is on the mat. If it is, then I believe that it is. Ascent routines operate on propositional mental states, applying mental-state operators to those propositions. As such, they clearly do not offer a nonconceptual account of reality testing. Byrne posits other ascent routines that allow me to determine other sorts of attitudes besides belief – if x is desirable, then I desire x. The ascent routine for visual perception relies on "v-facts" – facts that can be known only visually, like an object's particular shade of color. The existence of such v-facts can be used to infer through an ascent routine that one is seeing an object with the properties specified by the v-facts. Byrne also offers ascent routines for memory

and imagination. Like Hume (1739), he appeals to the idea that these modalities are "degraded" relative to perception. These ascent routines appeal to the existence of v-facts along with their degradation. These ascent routines might, then, be used as a form of reality testing, but of a clearly conceptual sort. Transparency theory is thus inadequate to explain nonconceptual reality testing. The Humean approach to distinguishing memory and imagination from perception is arguably nonconceptual, and later in this chapter I will argue that it is also insufficient for reality testing.

4.1.3 Modularity theory

Modular accounts of metacognition offer functional "boxologies" composed of mental modules that explain our metacognitive abilities, the most prominent being proposed by Nichols and Stich (2003). The boxes in a boxology represent causal role, so a proposition is in the belief box when it is poised to play the sort of causal role beliefs play in our behavior.

Some modules are involved in third-person metacognition, like those that detect where conspecifics are looking or detect the desires of others. Other modules apply to first-person metacognition. There are modules responsible for first-order states such as perception, which in turn feed into modules for belief. My believing that the cat is on the mat consists in the proposition 'the cat is on the mat' being in the belief box. That same proposition in the desire box would constitute a desire that the cat is on the mat. Metacognitive modules take first-order mental content as an input and apply epistemic operators. Metacognitive content is produced when a proposition from, say, the belief box, feeds into a metacognitive module that affixes an "I believe" operator to that proposition.

Nichols and Stich do not offer a detailed account of reality testing, but there are indications of what they might suggest. They concede that further modules are necessary in order to apply 'I perceive' operators to visual content. This entails quite a proliferation of metacognitive modules if the account is to capture every sort of metacognitive operator (I want, I hope, I suppose, etc.). There are questions about whether this is proliferation is plausible, and even more fundamental questions about whether the boxological account is coherent when we lose the box metaphor and attempt to understand how such a system might be implemented in a brain. I will consider some of these questions later in this chapter.

But regardless, it is clear this form of reality testing requires propositional thought. The application of an epistemic operator implies a propositional format. In general, however, since the boxology represents causal role, one might posit a nonconceptual mechanism for implementing a similar account. One might, for example, incorporate Goldman's neural account of reality testing, discussed below, with a boxological account of reality testing. Or one might interpret it as I do in chapter 5, in terms of sensorimotor skills. These might be interpreted as nonconceptual implementations of the general functional picture painted by Nichols and Stich. This fact, however, reinforces a general problem with the abstract boxological approach. It is consistent with such diverse types of realizations, "realizations" that are themselves functional accounts, that it is arguably uninformative.

4.2. Nonconceptual reality testing

In this section I will survey accounts of reality testing that do not require the possession of conceptual content on the part of the subject. I will conclude, however, that these accounts have serious problems.

4.2.1 Phenomenal qualities

Phenomenal qualities like the color content of perception are also taken by many to be nonconceptual content. Therefore any account of reality testing that made use of phenomenal qualities might be used to explain nonconceptual reality testing. I will examine accounts of reality testing that rely on vividness, cognitive phenomenology, and feelings of reality and familiarity and conclude that they fail.

4.2.1.1 Humean vividness

Imagination and episodic memory are similar to perception in that they involve sensory qualities like shape, color, pitch, and so on. This similarity is the very reason for the existence of reality testing – the perceptible properties of perception and imagination or episodic memory can be difficult to distinguish. Some of these properties, we have seen, are taken to be nonconceptual – determinate shades of color, for example. Hume claims that the phenomenology associated with mental imagery can be distinguished from that of perception because the former is less vivid. Byrne (2018) offers essentially the same claim, but substitutes ‘degraded’ and ‘transformed’ for ‘less vivid’. If this proposal could be made to work, it would explain how creatures who lack

mental state concepts could perform reality testing – they become sensitive to the sensory or "phenomenal" properties that distinguish imagination and episodic memory from perception.

There are many problems with Hume's proposal (See McGinn, 2006 for a sustained critique). First, it is unclear what vividness means – it cannot be brightness or degree of attention, as images can be brighter or better attended than a similar percept without a failure of reality testing. And even if we accept an intuitive notion of 'vividness' as basic and without need of analysis, there are obvious counterexamples both from everyday life and empirical research. Much of our perceptual experience is degraded, for example when we hear a poor-quality audio recording, when lighting is dim, or when we forget our glasses. Hume's account predicts that we should always mistake these degraded perceptions for mental imagery or memories. More often we take them to be dim, fuzzy, or otherwise degraded percepts.

Despite these commonsense objections, the Humean account of imagery has drawn empirical support from a particular interpretation of the Perky effect (Perky, 1910). Perky placed subjects in front of a blank screen and asked them to imagine an object projected onto it. When an image of that same object type was surreptitiously and gradually projected on the screen, subjects mistook the perceived object for an imagined object, offering descriptions of their imagined objects that matched the projected image. The more general phenomenon of perceptual abilities being inhibited by simultaneous imagination has become known as the Perky effect. A Humean explanation of the effect is that a sufficiently degraded perceptual stimulus, the projected image at a stage of partial brightness, is indistinguishable from imagination, since the two types of experience differ only on the dimension of vividness. Thus, subjects will mistake faint projections for imagined objects. As for the more general effect, the Humean might claim

that confusion between similar imagined and perceived objects inhibits accurate detection of perceptual stimuli.

Segal and Fusella (1970) offer an explanation consistent with the Humean approach, though framed in terms of signal detection theory (SDT). If an object of perception, an external "signal," generates an internal signal (or representation) in the visual system which must then be distinguished from noise in the visual system, this signal might be confused with the internal signal generated by imagining a similar object. "This model assumes that both a perception and an image have an internal representation and S makes his sensory decision on the basis of these internal representations; when they are very similar or when the imaged signal is very strong, discrimination of the physical signal becomes more difficult" (p. 464). The way in which the representations are "similar" is not explicitly described as Humean vividness but is compatible with it. Indeed, the SDT framework offers a more precise account of what vividness could be – signal-to-noise ratio. Contextual effects could be recruited to explain the rest. Context might bias our reality decisions in favor of imagination or perception. In Perky's studies we are biased toward imagination, and thus when a perception of comparable signal-to-noise ratio is presented we mistake it for imagination. The converse would be predicted as well, though it would be more difficult to surreptitiously introduce mental images while subjects are concentrated on a perceptual task. These contextual effects might explain why we do not typically make reality testing errors when perception is degraded – my turning on the radio leads me to expect a perceived song and not an imagined one, even if the resulting sound is fuzzy or otherwise degraded. The Humean account, then, is not so easily dismissed. The question of its truth is an empirical one, and it has been subjected to empirical scrutiny.

Reeves (1980) finds that the Perky effect does not occur in an early, parallel stage of visual processing but does occur in a later, serial stage. The two stages are distinguished by the effect of backward masking. Subjects viewed a pattern of dots, followed by another pattern that "masks" the previous one by overlapping it. Masking disrupts serial processing, capping the number of objects that can be processed. In imagery trials subjects are asked to maintain a visual image while performing the task of reporting numbers of dots displayed. Subjects in masking conditions showed reduced accuracy (d') relative to non-imagery trials. There was no difference in accuracy between imagery and non-imagery conditions in unmasked trials. Reeves concludes that it is not similarity of the image to the stimulus which produces the Perky effect but a processing bottleneck. Producing imagery reduces the overall processing capacity available, which reduces accuracy at these later stages of perceptual processing.

This hypothesis predicts that imagery can improve perceptual performance if imagery is used to suppress stimuli that would otherwise interfere with processing. Reeves found just this effect when the task was to find a dot out of place. Imagery that matched the normal pattern of dots weakened processing of irrelevant dots and made the signal for the odd dot stronger, increasing accuracy or d'. This evidence undermines the Humean claim that imagery reduces perceptual accuracy due to similarity of the representation. Instead, imagery reduces perceptual accuracy in some conditions by creating a processing bottleneck, but this same bottleneck can be exploited to increase perceptual accuracy under the right conditions. Controversy remains about many aspects of the Perky effect, but a Humean explanation is no longer a contender (See Waller et al., 2012 for a review). Thus, the Perky effect does not support the Humean account of reality testing.

4.2.1.2 Cognitive phenomenology

There are many conceivable ways that reality testing could be performed on the basis of sensory or phenomenal properties. David Pitt (2004) holds that we are aware of what mental content we entertain at any given moment in virtue of that thought's distinctive, proprietary phenomenology. His account focuses on content, not on the attitude taken toward a given content. But one might also claim (indeed, one probably must in order to make sense of Pitt's account) that there is also a distinct, proprietary phenomenology of taking a particular attitude toward some content. This would account for our ability to know that we believe that p rather than hope that p or fear that p. Feelings of presence, feelings of reality, and feelings of familiarity have been posited to account for our ability to know that we are in the presence of an object, that we are experiencing an actual world, and that a given mental state is a memory, respectively. If such feelings reliably accompany memory, imagination, or perception, this might be the basis of reality testing, and it would be nonconceptual.

Pitt's view runs into the problem that many of our mental states are unconscious. Knowing what I believe in some cases requires accessing a standing, unconscious belief and making it occurrent. But if I identify that content on the basis of its phenomenology, how can I identify the unconscious belief to be made conscious, if phenomenology is necessarily conscious and the belief is by hypothesis unconscious? Proponents of cognitive phenomenology tend to either deny the existence of unconscious mental content or claim, seemingly paradoxically, that there is unconscious phenomenology. But this problem need not arise for a similar account of reality testing. I have made no claim that reality testing always involves distinguishing among

conscious states, but a proponent of a phenomenal account of reality testing might plausibly insist on it. If so, then these particular issues do not arise.

But there are other problems for such an account. This more general problem can be effectively illustrated in relation to Bermudez's account, with which the reader should now be familiar. I will do so in the next section. In general, my point is this. If a feeling reliably co-occurs with some mental state attitude, the cognitive system must still be able to identify the attitude in order to produce the correct feeling. This requires the very ability that the feeling was meant to explain. The feeling simply reports the results of unconscious cognitive operations to the conscious subject. Pitt might go further, however, and identify the mental attitude with the feeling (if he can make sense of unconscious feelings). But then the problem is explaining how reality testing could ever fail, and why it seems to fail in systematic ways. The fact that reality testing fails shows that our imaginary states are not indelibly tagged as imaginary. There must be cues we follow that may sometimes lead us astray.

4.2.1.3. Feelings of reality/familiarity

Bermudez (1998) claims that conscious memory involves a feeling of familiarity. This feeling is meant in some way to accompany conscious memory and to distinguish its content from other types of mental content, like perception. He does not flesh out the details, but I will consider the different ways a feeling of familiarity might function and argue that none are sufficient to account for nonconceptual reality testing.

The interpretation of the feeling of familiarity most directly suggested by Bermudez's account is as a feeling of recognition. That is, he invokes the feeling of familiarity not as a

general means of reality testing per se, but as a means of recognizing features as having been previously perceived. This is a first step toward reidentification of particulars, which Bermudez takes to be a necessary condition for objectivity. On this interpretation feelings of familiarity serve as a mark of recognition, not as a general mark of mnemicity. That is, it might be the case that memories do not typically occasion feelings of familiarity, rather such feelings only occur when there is a match between a memory and a perception, facilitating recognition. This would allow for reidentification of features, but the problem is that it does so only by assuming it. For a feeling of familiarity to occur only when a feature is recognized presupposes that the cognitive system can do just what we were puzzling over how to do—determine that two nonconceptual representations with the same object were the same yet somehow different, and on that basis produce a feeling of familiarity. But this is just what the feeling of familiarity was meant to solve. And it is also insufficient as a means of reality testing, for which we need more than recognition. Memories occur in the absence of their objects as well, and we must not mistake them for perceptions.

Perhaps, then, it would be better if we take feelings of familiarity to be a mark of mnemicity – the feeling accompanies every memory. First, this seems inadequate for recognition on Bermudez's account. Bermudez holds that nonconceptual representations are unstructured. This entails that the feeling is not attached to the memory itself but merely coincides with the representation (or at any rate the representation's becoming conscious). But mere coincidence will not be sufficient. Recognition of a perceived feature would involve a conjunction of three mental states—the perception of a feature, the memory of the feature now perceived, and the feeling of familiarity that accompanies the memory. This would entail that any time a perception and a memory coincide, regardless of their content, a feeling of familiarity would also coincide.

A memory of an eating affordance that happens to coincide with a completely different eating affordance (or a fleeing affordance for that matter) will be indistinguishable from real recognition. The problem also arises for reality testing if we think of the feeling of familiarity as merely coincident with the memory. The feeling does not allow the creature to determine which coincident representation is the memory – the feeling is equally coincident with any perception that happens to occur at the same time. Some experimentation might solve the problem in highly constrained cases. Hold representation A constant while changing representation B. If the feeling persists, representation A is the memory. This method would prove impractical in ecologically valid circumstances.

Anyhow, unlike Bermudez, we are not committed to the claim that nonconceptual representation is unstructured. It must simply not be structured like a conceptual representation. If the feeling of familiarity were attached to the representation in such a way that it indicated to the creature entertaining it which representation was familiar, it might be a better candidate for the basis of reality testing. And phenomenal states can be composed of parts. The taste of wine may be described as comprising nuttiness, sweetness, and the like. The case of feelings of familiarity is complicated by the fact that it would necessarily involve multimodal phenomenology – whatever modality feelings of familiarity belong to (perhaps cognitive phenomenology), along with the various elements of an episodic memory. But it does not seem to be impossible in principle.

The question is how the feeling of familiarity comes to be part of the memory. There seem to be two possibilities, and neither relies on the phenomenal quality. Rather, the phenomenal quality serves as a signal to the subject, but the cognitive system must do a bit of work to ensure that the signal is generally accurate. The feeling of familiarity may be attached to

the memory in virtue of some feature of its content that distinguishes it as a memory. This, of course, is the whole problem reality testing is meant to solve and thus invoking a feeling of familiarity simply pushes the problem back. The other possibility is that the feeling of familiarity becomes part of the memory as a function of the architecture of the cognitive system. Representations that originate in memory will, barring some malfunction, always have feelings of familiarity as part of their content. This sort of approach, in fact, might be invoked to solve the problem of reality testing regardless of whether there is a phenomenal component involved. The functional architecture of the mind might be such that representations originating in memory or imagination are either marked from their inception as being memories or images, or the functional architecture of the mind might be such that no marking is needed – the functional pathways keep such representations isolated, or only allow interactions that would not allow confusion. I will consider this approach in section 4.2.3. Anyhow, it would make the phenomenal component superfluous except as a signal to the conscious subject of reality testing that has already been accomplished at a lower level.

4.2.2 Neural properties

Goldman (2006) is a simulation theorist of third-person metacognition or "mindreading." Simulation theory holds that we draw conclusions about the mental states of others by simulating their mental states in our own minds. We do not, however, simulate our own mental states. Goldman proposes a kind of inner-sense account of self-directed metacognition, but only as regards the ability to distinguish mental attitudes. We determine the content of those states, he claims, in another way. His account of our ability to determine what attitude we are taking

toward a given content could serve as an account of reality testing. Since the mechanism he proposes does not require concepts, it is worth considering here.

Goldman claims that we directly introspect neural properties, and that these are the basis of mental attitude individuation. His remarks are brief, so a bit of interpretation is needed. He claims that introspection is a quasi-perceptual mechanism and that we determine our own mental attitudes by (quasi) perceiving the neural types involved. He offers an analogy with pain perception. Particular types of pain sensations, he claims, implicate particular types of nociceptors, and on this basis we can individuate pain types. He suggests that we distinguish attitude types in the same way. He does not identify the introspective content involved in determining the pain type with the pain *sensations*, but with perception of the nociceptor types themselves. The sensation provides the content and the introspection of the neuron itself provides the pain type. This perception of the nociceptor is not conscious - clearly this is not how we experience the individuation of our mental state types. But there is, for Goldman, a subconscious, quasi-perceptual modality that perceives the neuron or group of neurons involved in my headache and identifies its type. This, he claims, is analogous to how we distinguish beliefs from desires and other mental states. Perhaps, though Goldman does not go this far, this is also the way we distinguish perception from imagination and memory. If these neural types can be represented as features, then the process could be nonconceptual.

Goldman's is a strange picture. It appears to make levels of explanation intersect that are typically kept distinct. A traditional picture of the relation between psychological states and their neural realizers would posit interaction at the neural level that manifests as interaction of psychological states at a higher level. The latter can be reduced to or explained in terms of the former. It is odd to posit that a psychological event, like introspection, be directed at a neural

event. The picture is made starker by his account of how neurons are typed. In other sections he argues that functional properties cannot be the properties perceived by inner sense, since these are largely non-occurrent, dispositional properties. But if this is to be applied to neural properties, then it seems the inner sense must respond to intrinsic properties of the neuron, such as axon diameter or myelination. Firing rate and transmission speed, for example, are functional properties and must be ruled out.

This commits Goldman to a brand of type physicalism that few currently hold. Even in the case of pain, which is meant to be the concrete basis of his metaphorical gesture toward the nature of propositional attitude introspection, his account conflicts with the evidence. The dominant account determines pain types not by nociceptor types (which are themselves determined in part by their functional properties) but by patterns of activity, as evidenced by phenomena like phantom limb pain, where pains exist in the absence of the relevant nociceptor (Melzack & Wall, 1988). Neurons are notoriously plastic, and many researchers doubt that a mental state type will have a reliable correlation with any given neural type (Aizawa & Gillett, 2009), though there are still proponents of identity theory (Polger & Shapiro, 2016). But the case is even worse for Goldman, since his is not merely a metaphysical claim, but a claim about how humans identify their own mental state types. An identity theorist only concerned with metaphysics could make use of dispositional properties when individuating neural types. But Goldman, in his critique of functionalist accounts of mental state type individuation, denies that dispositional properties are introspectable. If so, the subject in first-person metacognition has a much more difficult task even than the beleaguered identity theorist, in that she must distinguish her own mental state types on the basis of their occurrent properties alone. I am not aware of any identity theorists who hold that this is possible.

4.2.3 Architectural properties

In section 4.1 I considered Nichols and Stich's (2003) modularity theory as a functionalist account of reality testing that operates by appending mental state operators to propositions. As this approach employs mental state concepts, it cannot explain nonconceptual reality testing. But one could imagine similar systems that operate on nonconceptual representations. The functional "boxes" that house each type of mental state keep representational types separated. So long as the imagination box and its functional inputs and outputs preclude, say, eating behavior when an image of food is entertained, the problem of reality testing is dissolved. This sort of picture complicates Nichols and Stich's account, as it requires distinct modules for each sort of propositional attitude and distinct functional pathways for each. If a central reasoning module cannot distinguish beliefs, desires, hopes, and memories on the basis of their propositional operators, this extra work must be accomplished by the architectural properties of the cognitive system. This picture might be criticized for its complexity, but the brain is complex.

There is a sense in which something like this architectural solution to the problem of reality testing is obviously the case. For any physicalist, reality testing is accomplished through the causal properties of a series of neural events. At this level of description an explanation of reality testing needn't invoke mental content at all. But we must also inquire whether reality testing demands explanation at the psychological level (pace Goldman and his unorthodox blending of levels of explanation). There may be eliminativists who insist that no neural processes demand explanation at the psychological level (Churchland, 1981). Most of us, however, hold that some (though probably not all) neural processes can be usefully characterized

at the psychological level as well. Is reality testing one of these? The studies described in section 2 shed light on this. Recall that failures of reality testing in rats and mice were theorized in terms of conditioned stimuli evoking certain types of representations and those representations evoking certain types of behavior. Where genetic, surgical, and neurotropic interventions were used, these were characterized fairly crudely as interventions correlated with symptoms of schizophrenia – as levers known to turn a certain psychological state on and off. Eliminativists tend to hold that empirical research ought to be our guide to whether psychological explanation is appropriate. Current research into reality testing appears to rely heavily on psychological explanation and the invocation of representational content. It is therefore reasonable to insist that we seek an explanation of reality testing at the level of representational content while conceding that there is another level of explanation as well.

Is it plausible, then, to posit an architectural solution to reality testing at the psychological level? The various "boxes" that contain mental states can be defined functionally in terms of their inputs and downstream effects. But there is a fundamental problem of integrating nonconceptual content into a boxological picture of cognition. Nichols and Stich's boxology, for example, has beliefs and desires flowing from their respective boxes into a single practical reasoning module. The practical reasoning module must distinguish beliefs from desires in order to perform its operations adequately, and for Nichols and Stich this is done on the basis of propositional operators. For an entirely nonconceptual psychology to be modeled in this way we must assume some other type of compositionality that would enable goal desires to combine with perceptions of affordances. Bermudez's view of nonconceptual content as unstructured seems incompatible with a modular psychology of this sort. One might invent a new type of boxology for nonconceptual content based on a new type of structure. This would require another book.

But for our purposes we can simplify the task. I have not assumed that rats and the other creatures I considered in Ch. 2 lack propositional content altogether. Rather, I have only claimed that they lack mental state concepts. If so, the task for the proponent of the architectural dissolution of the problem of reality testing is somewhat more tractable. We might consider whether a creature with propositional beliefs and desires might accomplish reality testing by segregating their memory and imagination functionally. If representations in the memory or imagination box do not lead to actions, whereas representations in the perception box do, there might be no need for a form of reality testing sensitive to properties of the content of the representations.

It is in this sense that a functional/architectural account could be said to "dissolve" the problem of reality testing. It is not that we don't eat imagined pizza because we have solved the reality testing problem, rather the reality testing problem is dissolved because we don't eat imagined pizza. There are two ways to interpret this approach. On a strong dispositionalist interpretation there is nothing more to these mental states than their tendency to produce certain downstream effects given certain inputs. These sorts of accounts are typically associated with behaviorism (though see Schwitzgebel, 2002) and tend to end in circularity (Chomsky, 1959).

The more plausible interpretation is that the differences in downstream effects between perception and imagination or episodic memory are due to a segregation of the boxes and their outputs. If the imagination box does not connect to motor outputs, then reality testing may not be necessary. But of course, this picture is too simplistic. It is not true that imagination and memory never produce action. A memory can make us wince or head off in the direction of a lost wallet. Imagining a loved one in danger might induce panic and movement to ensure that loved one is okay. Of course, these actions are quite different than those would be occasioned by, say,

perceiving my wallet. A cognitive system could be conceived with two distinct and largely identical motor systems – one reserved for responding to perceptual states and the other for memory or imagination (perhaps there are three). There is no neural evidence to suggest this. And anyhow it wouldn't be sufficient. If we suppose, as we have been, that the creature in question has some propositional mental states then, as in humans, information originating in imagination and memory may be used in *inferences* that influence action. This would require not only distinct motor systems but distinct systems of central processing. But such inferences, if they are to produce appropriate action, must often involve information derived from perception, and the central processor must know which is which. Reality testing seems to be required.

4.2.4 Responsiveness to the will

McGinn (2006), while primarily concerned with offering a metaphysical account of imagination, makes some points that could constitute an account of reality testing. He claims that a subject can distinguish her own perceptual representations from imagination because the latter is responsive to the will in a way that perception is not. He concedes that some imagery is not willed, as when a song pops into your head unbidden, but claims that all imagery is "in principle" subject to the will. By this he means that it "makes sense" to will an image to start or cease, whereas it does not make sense to will the same for a perception. Since, he claims, we are aware of when we are willing or not, this suffices to distinguish imagination from perception.

The "in principle" bit may be relevant to the metaphysical distinction he attempts to make, but it is of little use in reality testing. Unwilled imagery is frequent enough that without some method of distinguishing it from perception the subject will be in trouble. And even when

one is aware of willing their imagery, problems arise that parallel those for Bermudez's feelings of familiarity. Does the awareness of willing merely co-occur with the imagining, or is it more intimately related? The former cannot be the case, for then willfully walking while perceiving should cause a failure of reality testing. If instead the willing attaches to the imagining like an operator, it is difficult to see how this could apply to nonconceptual forms of reality testing. McGinn says that being willed is "somehow" imprinted on the phenomenology of imagining. This seems of a kind with Pitt's cognitive phenomenological approach and the same objections apply. There is a difference, in that McGinn does not hold that being willed is part of the intentional content, whereas Pitt holds that the phenomenology is constitutive of the intentional content. But the claim is underspecified as it stands. If willing is imprinted only on imagery that is willed, this still does not solve the problem of reality testing for unwilled imagery. If he is claiming that all imagery is imprinted as being willed in principle, this seems to amount to a corollary of the feeling of mnemicity we discussed earlier. And the same problems apply. The phenomenology may report to the subject the result of reality testing, but if it is consistently accurate it must either be the result of a subpersonal reality testing process based on content, which pushes the problem back, or a result of a purely architectural process, which has been critiqued in section 4.2.3.

There is, however, something right about McGinn's view. In chapter 5 I will argue that there are differences in the way imagery changes relative to motor commands as opposed to perception. I will describe this in terms of distinct sets of sensorimotor contingencies associated with imagery and perception. McGinn makes other interesting points that gesture toward such an account but consistently couches them at the level of phenomenology. He states that imagined objects are not "felt" as being in the same spatial relation to the physical eyes as perceived

objects. Again, this is on to something, but to focus on the phenomenology is to miss the explanatory dimension. If there is a distinction in the spatial phenomenology of imagination and perception, this is a result of the difference in spatial relations as represented.

4.2.5 Miscellaneous points by McGinn

McGinn points out a collection of other ways that a subject might distinguish her perceptions from her mental imagery that are, in one way or another, connected to his general point about the will. If true, each might provide a basis for reality testing.

McGinn claims that perception always occurs in a visual field, whereas imagination needn't. We can imagine the Eiffel tower without imagining its surroundings, whereas perception always comes in a context. Empirical research pushes against this conception of perception. We veridically perceive only small portions of the visual field, typically areas that are the focus of attention and are thus foveated (Mack & Rock, 1998). We miss many details at the periphery. Nor is it entirely clear whether, when I imagine the Eiffel tower, it has no background or whether it simply has a black background.

A related point is his claim that perception is "saturated," in that at every point in the visual field there is a phenomenal quality. The same empirical evidence pushes against this claim. Perception appears to involve sampling the environment for relevant information at a given time, leaving many gaps in our perceptual field (Simons & Levin, 1997). He also claims that imagination, but not perception, is attention dependent. Again, perception is in fact attention dependent as well, and unattended portions of the visual field are prone to detection errors (Mack & Rock, 1998).

Finally, and more interestingly for my account, McGinn borrows from Sartre (1966) the claim that images are not related spatially to one's body in the same way as percepts. He seems to claim that there is no relation at all, but this is not the case. In chapter 5 I will present evidence that shows a relation between saccades and the character of visual imagery. It is not the absence of a spatial relation but, as I will argue in chapter 5, a difference in the sensorimotor contingencies of that relation by which we distinguish imagination from perception.

4.3 Summing up

In this section I have argued against extant accounts of reality testing. I have dismissed, for our purposes, accounts of reality testing that require the subject to employ mental state concepts. While such forms of reality testing certainly exist, they do not occur in rats, infants, or birds, but these creatures are capable of reality testing. I then argued against accounts of reality testing that posit phenomenal differences between perception and imagination or episodic memory. These include Hume's vividness, Pitt's cognitive phenomenology, Bermudez's feelings of familiarity, and feelings of reality. I then argued that Goldman's account, on which reality testing is determined by introspection of neural properties, fails. I also argued against purely functional/architectural accounts which would make reality testing unnecessary. Finally, I argued against McGinn's "responsiveness to the will" account along with other suggestions by McGinn.

Chapter 5. Reality testing is a sensorimotor skill.

In this chapter I will develop an original account of reality testing. The account must explain reality testing abilities in creatures that lack mental state concepts. It must also avoid the problems I have pointed out for other accounts in the previous chapter. This is accomplished in part by keeping the account content-based – reality testing is not described purely at the neural or architectural level. Instead, differences in content between perceived and imagined or remembered states guide our ability to discriminate them. The content involved is not primarily phenomenal content, however, as reality testing is often accomplished unconsciously. Rather, I will argue, it is the changes in sensory content contingent on particular motor commands (and sometimes cognitive "commands") which differ between perception and imagination or episodic memory. Such "sensorimotor contingencies" have been employed in theories of perception, but their utility for reality testing has not been exploited. Imagined and remembered content does not change in the same way as perceived content when we move. In what follows I will use the terms 'sensory input', 'sensory content', and 'sensation' to include the content involved in imagination and memory. Sensation in this sense need not involve stimulation of peripheral transducers like the retina or eardrum.

In sections 5.1 and 5.2 I will explain the notion of a 'sensorimotor contingency' as it is used in extant sensorimotor accounts of perception and of imagination, respectively. No such account has been proposed for reality testing, and I will do so in section 5.3. My account will require abandoning some theoretical commitments of previous sensorimotor theorists, and I will argue that this is appropriate. I will develop my account of reality testing by describing some of the fundamental differences between the sensorimotor contingencies associated with perception and those associated with imagination or memory in each sensory modality. The description will not be exhaustive, but it will be sufficient to provide the basis of a novel account of reality

testing. I will argue that the differences in sensorimotor contingencies between perception and imagination/memory are used to construct distinct "spaces" into which we sort our sensations. My account makes predictions that, I will argue, are supported by the extant empirical evidence and that suggest further research. I will offer accounts of reality testing in different sensory modalities separately, starting first with audition, then vision, and finishing with a brief discussion of other modalities. In section 5.4 I will consider contexts in which reality testing failure is the norm. Dreaming is one such context. There are also accounts of perception on which much of what we consider perception includes imagined content. I will argue that my account of reality testing is consistent with the empirical evidence in each case.

5.1. Sensorimotor accounts of perception

In this section I will present a roughly chronological survey of some important sensorimotor accounts of perception. The survey is not exhaustive, but the accounts are chosen to illustrate the nature of sensorimotor contingencies and their usefulness in explaining certain aspects of perception and imagination. I have also chosen accounts that stress the nonconceptual nature of sensorimotor expertise and thus its availability to creatures who lack mental state concepts.

Helmholtz (1924) invokes sensorimotor contingencies to explain visual stability. Our eyes perform small movements, called 'saccades', almost constantly, yet our visual experience is of a relatively stable scene. Helmholtz proposed that the perceptual system takes into account not only information from sensory transducers, in this case the retina, but also motor information, like the movement of the eye. Movements of an image on the retina that are consistent with a given motor command for the eye and are cancelled out at the level of perception, while those

that cannot be accounted for by the movement of the eye are perceived as movement of the perceptual object. Thus, the content that we perceive is not a simple, bottom-up process of receiving information from the world. Rather, our perceptual experience is the result of our encoding of certain relationships between motor commands and the changes in sensory information that typically results. The process is typically unconscious, and the format in which the contingencies are encoded is likely nonconceptual, as the problem must be solved by any creature with a capacity to move its eyes.

Piaget (1952) proposed a sensorimotor stage of development in infants that precedes the acquisition of concepts. While not yet capable of what he terms "symbolic" thought, the infant is capable of intelligent interaction with its environment, implicitly learning the contingencies between certain actions and corresponding sensory inputs. The ability begins with reflexes, like sucking, but develops into more complex activities like pushing aside an obstructing object to grasp a desired one. This ability to move an "object" to access another needn't involve objectivity in the sense discussed in chapter 3. Instead, for Piaget, this form of intelligence represents a learned association between motor commands and sensory input. These ways of categorizing the world in action-guiding terms rather than conceptually Piaget terms 'sensorimotor schema'. These schemes, once learned, can be combined into more complex schema. Thus, innate responses like sucking and eye movement along with innate motivations that include a desire for cognitive stimulation will cause the cognitive system to gradually learn patterns of activity that produce interesting or useful sensory inputs.[8] These patterns are the infant's first sensorimotor schema and Piaget attributes the intelligent behavior observed in the first 18 months of life to this form of mental representation. Mental representation of this type

[8] The "desire" invoked here is presumably a goal desire of the sort described by Bermudez.

can be complex but is essentially action-involving and therefore subject-relative, and so does not satisfy the generality constraint. Examples of such sensorimotor abilities in infants include shaking a toy to produce a rattling sound, pushing aside an object to access a toy, pushing on an adult to produce some reaction, turning the head or moving the eyes to encounter a visual stimulus, grasping a toy under visual guidance, and turning the head in the direction of a sound. By varying and combining these and other basic schemata, infants master a wide variety of sensorimotor contingencies that embody a significant amount of "know-how" and enable relatively complex, intelligent behavior without the need for conceptual content.

Cussins' (1992) account of nonconceptual content uses sensorimotor contingencies to develop the concept of a "feature space," which I will use in my account of reality testing. Cussins defines nonconceptual content as representational content that is not truth-evaluable. The existence of such content, he claims, is evidenced by certain abilities. In the case of a coffee mug, nonconceptual content is exploited in our ability to "grasp the mug or otherwise to locate it, to track the mug through space and time, and to be selectively sensitive (in judgment and action and memory) to changes in the mug's features" (p. 655). Adequate characterizations of the content involved in this "experience-based knowing-how" (p. 656) will not contain propositional 'that' clauses, but instead will include descriptions of the mug as "as graspable, as locatable, as being such as to resist manual-pressure, as being drinkable-from, as being push-and-then-fallable" (p. 658). The mastery of such skills is a relation not to objects, but to Strawsonian features - a "necessarily local (and context-dependent) object, and hence no object at all" (p. 659). While one *could* attempt to describe these skills using propositions with truth conditions, Cussins argues that it is more apt to invoke "fluency" as the relevant normative standard. An activity is fluent if the specific, subject-relative activity invited by the perceived feature is in fact

accomplished without mishap. The skill is not one of handling mugs, but of this subject handling this thing for this particular purpose on this occasion. Conditions of fluency are not truth conditions, but rather "experiential activity threshold conditions" (p. 666). Fluency is an evaluative standard for representations of opportunities for skilled engagement with some feature.

Cussins introduces the notion of a "cognitive trail," which is a development of skill in exploiting some feature. Cognitive trails become more elaborate as the creature's experience with the feature expands. This notion bears some similarity to the concatenation of sensorimotor routines posited by Piaget. I may master grasping the coffee mug, which in turn leads to a mastery of bringing the mug toward my face, tipping it at the appropriate angle, and so forth. Nonconceptual content, then, just is "(the experiential presentation of) cognitive trails" (p. 673). As the cognitive trails become more elaborate and exhaustive of the possibilities of exploiting the feature (i.e., "intersecting"), they come to define a "feature space." In grasping the mug, I find myself at an intersection of cognitive trails, one that affords drinking and another that affords throwing.

Although he largely sees himself as elaborating Evans' account of conceptual content, Cussins does not posit a single, binary criterion for the distinction between conceptual and nonconceptual content. Instead, he holds that nonconceptual and conceptual content are continuous with each other. He describes two axes along which they differ. One axis is the density of the cognitive trails which constitute a feature space. As I come to conceive a feature as affording a wide range of possible actions, this cluster of affordances in the environment begins to acquire a kind of generality insofar as a large number of proto-properties seem to inhere in a single location, resembling an object in space. This dense cluster of sensorimotor contingencies

constitutes the feature space of the mug, which in turn may be nested in a higher-order feature space of the room, and so on. When a creature can reliably move from one feature to the next, be this from an eating affordance to a drinking affordance in space, or from a grasping affordance to a throwing affordance in one's own bodily space, that creature develops a skill for navigating the relevant feature-space. A sufficiently elaborate feature space is one necessary condition for conceptual content.[9]

An elaborate feature space is not yet a concept until it is also "chunked" – this is the second axis of the conceptual-nonconceptual continuum. The phenomenon of chunking was initially posited in memory research. A phone number, for example, is easier to remember if the ten digits are grouped into chunks. The area code can be broken off from the other digits and because of its repeated instantiation in various phone numbers becomes a representation all its own of a geographical area. Features of edibility, graspability-upon-reaching, and so on might ultimately become chunked into a fruit concept. Concepts provide the advantage of reducing cognitive effort. Each mug no longer needs to be encountered as foreign territory to be explored. Instead, it can be treated as containing a similar set of behavioral affordances as any other mug. But should the mug violate our expectations, perhaps by being made of Styrofoam and thus being too light to afford proper throwing, the faux mug can be unchunked and treated as a territory to be explored through cognitive trails. In this way concepts can also be revised.

We can visualize the continuum from nonconceptual to conceptual content then as a two-dimensional Cartesian plane, with exhaustive skill in navigating a feature space on one axis and the stability or "chunking" of the proto-concepts thus defined on the other. Fully conceptual

[9] Adult humans can, of course, entertain very sparse concepts like 'colored' or 'has a property'. Cussins is explaining the grounds for having concepts at all. Sparse concepts arguably develop later in phylogeny and ontogeny.

thought will occupy one corner, but there is much space for thought that is less than fully conceptual, but not completely unstructured. Recall that one problem for Bermudez's account of reidentification of particulars, a necessary condition for objectivity, was that he conceived of nonconceptual content as unstructured, thus not allowing for content-based reality testing. Reidentification of particulars required an ability to distinguish one's memories from one's current perceptions in order to identify the perceived feature as identical to the remembered feature. But without some form of structure, it was unclear how the representations could be of the same feature yet recognizable to the subject as being entertained in different modalities.

While Cussins' account suggests room for structured nonconceptual content, it does not offer an obvious solution to the problem of reality testing. Grush offers an account of the genesis of spatial content that, when applied to Cussins' notion of a feature space, suggests a solution to the problem. Grush's (2000) skill-based account of spatial content is illustrated by two examples. First is a sensory substitution device, the "sonic guide," used to provide a quasi-visual experience to the blind. The device emits a supersonic frequency that is reflected from nearby objects and converted into an auditory tone for the subject. The tone will vary according to the properties of the objects at which the device is directed. Distant objects will produce a lower tone, solid objects a smooth tone, and so on. It takes some practice, but a subject can become so adept at using the device that she stops consciously inferring the properties of objects in her environment from properties of the tone. Once the subject has "learned" the device, the tone becomes transparent, and the objects represented are "directly" perceived. That is, from the subject's point of view she is immediately presented with, say, the presence of an object in a particular position relative to her body, rather than a c-sharp at a particular volume. The sensory substitution device provides, arguably, similar spatial content to that which a sighted person acquires visually. Such

a thing is possible if we conceive of perceptual content not as simple sensation, but as emerging from a complex interaction between information provided to the senses and information provided from the motor system. Learning to use the sonic guide consists in learning these sensorimotor contingencies, coordinating one's actions with the information provided in sonic form.

Sensory and motor information vary along dimensions that Grush calls "manifolds." Auditory pitch represents one manifold, and we possess an ability to make discriminations regarding information in that manifold. The motor commands to the muscles controlling the elbow are another manifold, opening or closing the arm to various positions in a single dimension. Spatial perceptual content, for Grush, arises when two or more manifolds are *coordinated* in a way such that a higher-order manifold is constituted. Grush distinguishes two types of coordination, c-coordination and r-coordination. C-coordination is simply connecting two manifolds by identifying a common feature, as when we can connect a map of California to a map of Nevada by identifying Barstow on both. Using c-coordination we can link the elbow manifold to that of the wrist, and so on for the entire body. More interesting for our purposes, however, is r-coordination, which takes two manifolds and produces a higher-order manifold.

To illustrate the concept of a higher-order manifold, Grush uses the example of a single eye able to move in two dimensions and a point of light directed at it. Because the eye can move, any pattern of retinal stimulation is compatible with a variety of positions for the light. The light might be directly in front of the eye, but if the eye is pointed up, the stimulation will occur in one part of the retina and if the eye is pointed down, the stimulation will occur in a different part. But by plotting eye position against location of retinal stimulation we can derive a higher-order curve that represents all such combinations compatible with a single position for the light. When two or more manifolds are coordinated in this way, such that they produce a higher-order manifold, they

are r-coordinated. R-coordination, for Grush, produces spatial "content," whereas a mere manifold only contains "information."

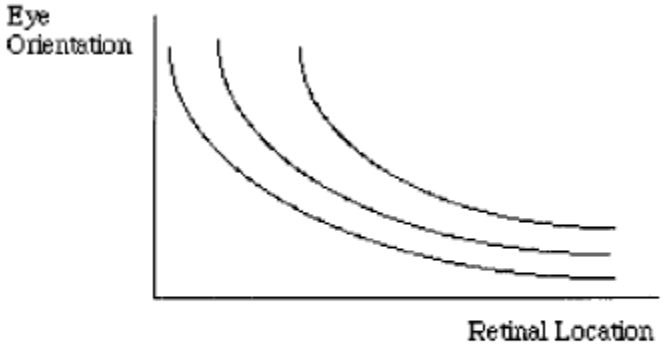

Figure 1. Schematic plot of combinations of eye orientation and location of stimulation on retina (for the simplified two-dimensional eye described in the text). All points on a given contour correspond to combinations of orientation and location that indicate the same direction in head-centered space.

Figure 5.1 Higher-order manifolds

Note that the same two manifolds can constitute different higher-order manifolds if the contingencies between them vary. This is not explicit in Grush, but it follows from his account. If the retinal stimulation does not change when the eye moves, or changes in different ways, we will have a different curve and thus a different sort of higher-order manifold. Imagination and memory, I will argue in section 5.3, are higher-order manifolds distinct from the higher-order manifold that is perception. This is due to the differences in the sensorimotor contingencies for each. An afterimage, for example, appears to move as the eye moves. A perceptual object does not. When I speak of perception and imagination/memory as distinct feature spaces, I take a feature space to be a type of higher-order manifold – higher by several orders of magnitude.

Grush goes on to posit a nested series of r-coordinated, higher-order manifolds, all c-coordinated with each other cumulatively constituting an egocentric space. We receive sensory information along various kinds of manifolds—visual, auditory, tactile, etc. There are also a great number of manifolds implicated in action. The elbow, for example, can be moved along

one dimension and the motor commands implicated in this movement constitute a manifold. Other joints with more degrees of freedom will create more complex manifolds. All of these must be c-coordinated relative to some relatively fixed point, and this is the torso. When these various manifolds are coordinated, I will be able to respond to a fly on my knee by swatting it with my left hand by coordinating my visual and tactile sensory manifolds with the various motor manifolds that connect my hand to my arm to my torso, and on to my leg and my knee. Content in this space will be of the feature-placing variety, in that the fly is not conceived as an insect or even an object per se. Rather it is a "to be swatted thusly," or something to this effect (keeping in mind that no concept of hands as objects is being invoked either).

This egocentric space and the complex coordination of manifolds involved is not sufficient for objectivity. And here Grush has a different story to tell than Bermudez. For Grush what gets us from egocentric space to objective space is the c-coordination of an allocentric map in imagination with the egocentric map in perception. This allows the creature to identify the same feature as existing in perception, but also existing unperceived as an item on the allocentric map. This accomplishes for Grush what recognition using conscious memory accomplished for Bermudez. I have argued in chapter 3 that this implicitly assumes a reality testing ability. And, as I have begun to suggest, Grush's account of spatial content also contains a level of structure that allows for an explanation of this reality testing ability – imagination and perception are distinct higher-order manifolds, constituted by distinct sets of sensorimotor contingencies. These higher-order manifolds are compositional, and thus may contain constituent parts that are type-identical to those contained in other higher-order manifolds. The manifold for eye position in imagination is the same manifold for eye position in perception. What differs is the behavior of sensory inputs relative to change of position.

5.2. Sensorimotor accounts of imagination

Neisser (1976, 1978) and Thomas (1999) offer sensorimotor accounts of imagination. These are not accounts of reality testing per se, but both authors have something to say on the subject. A typical sensorimotor account of perception holds that perception consists of sensory inputs, motor commands, and most importantly knowledge of the contingencies between the two. Neisser and Thomas claim that imagination consists of the same motor commands and sensorimotor knowledge involved in perception without the corresponding sensory input.

> Imagery is experienced when a schema that is not directly relevant to the exploration of the current environment is allowed at least partial control of the exploratory apparatus… we imagine, say, a cat by going through (some of) the motions of examining something and finding that it is a cat, even though there is no cat (and perhaps nothing relevant at all) to be examined (N. J. T. Thomas, 1999, p. 218).

By "schema" Thomas means the sensorimotor contingencies for a given type of object, say a cat. The term is borrowed from Neisser (1978), who also claims that imagination is an "unfulfilled" schema, in which a motor command is performed and a particular sensory input is expected, but does not arrive.[10]

[10] Neisser (1978, p. 170) applies this account only to deliberate imagination and not to hallucination. He does not appear to have an account of hallucination, and the restriction seems to brush away the problem of reality testing by stipulation.

Thomas and Neisser's accounts of imagery, while innovative, offer little more than broad outlines. As such, it is difficult to make sense of certain aspects of the account. Consider Thomas's sensorimotor account of perception:

> During normal perception the current schema activates certain instruments which then proceed to make their tests. The results of the tests are reported back to the schema and contribute to determining which instruments will next be activated, and so on in a continuous cyclical process. (1999, pp. 222–223)

Imagination is meant to consist in the same process without the sensory input. But this would seem to eliminate the "results" of the test, which leaves us unable to determine which "instrument" will next be activated, thus eliminating cyclical process. Elsewhere he states that:

> …there can be a continuum of cases between the extremes of veridical seeing and "pure" imagery (where the imaginal experience incorporates no aspect of what is before us). In the former, all the perceptual tests ordered by the schema are actually carried out, and all results returned by the perceptual instruments are given their full weight in determining the course of subsequent testing. In "pure" imagery, either the ordered tests are not carried out at all, or else any results that are returned by the instruments are completely ignored. (N. J. T. Thomas, 1999, p. 233)

This passage is even more puzzling. Not only is no sensation received to trigger the next step in the cyclical process, but the perceptual tests are not carried out at all. It is not altogether clear

what is meant here by “perceptual test.” If the test is the execution of a motor command, this would contradict empirical evidence he frequently cites of the similarity in eye movements between perception and imagination. If the execution of the test is a process of matching the sensory information to a prediction, then we still have the original problem that the similarity to sensorimotor perception fails – there is no cyclical process at all.

The previous quote is taken from a passage in which Thomas uses the phenomenon of pretense in children as support for his theory. A child acts as if a doll is smiling, and even though the doll never smiles, the child proceeds as if it were. This type of activity might be termed ‘imagination’, but in a far different sense than that involved in reality testing. There is no visual experience of a smiling doll, the child simply acts as though the doll is smiling. Thomas’ account is intended to capture a wide variety of phenomena commonly termed ‘imagination’. But this comes at the price of plausibility when it comes to visual imagery. If the same motor routines that are claimed to be sufficient for imagery are also employed in episodes that lack imagery, this is a problem for the account without some further explanation.

This discussion reveals a sense in which my account of reality testing is at odds not only with Neisser and Thomas’s sensorimotor accounts of imagination, but with sensorimotor accounts of perception as well. Thomas and Neisser claim that imagination does not involve sensorimotor interaction with a mental image, but rather posit imagination as the product of motor activities and sensorimotor knowledge in the absence of sensation. This is perhaps motivated by a general tendency of sensorimotor theorists to reduce the number of representations involved in perception. Noe and O’Regan (2001), for example, argue at length that perception does not involve a detailed representation of the visual scene, as the world itself is available to be sampled in vision as needed. Phenomena like change blindness, our tendency to

miss subtle changes in the visual scene, are meant to support this claim – a detailed visual representation of the visual scene would help us catch such changes. Something like this may well be the case in perception, but no sensorimotor theorist manages to reduce the number of visual representations involved in perception to zero. Neisser and Thomas may be trying to do just that in the case of imagination. If, for Noe and O'Regan, the world fills in where the representation normally would, Thomas and Neisser take away both the representation and the world, leaving nothing. In my account of the sensorimotor contingencies involved in imagination, there is sensory experience of a mental image. Mastery of the sensorimotor contingencies that govern our interaction with those images does not constitute the image, but it does enable reality testing.

Thomas calls accounts like mine, on which imagination involves interaction with a mental image, "pictorialism." He associates the view primarily with Kosslyn (1980). Thomas cites as evidence against pictorialism studies that show subjects are unable to perform the Gestalt switch involved in viewing a Necker cube or a duck/rabbit when those ambiguous figures are imagined. This, he claims, shows that there is no image being interacted with in imagination. He does acknowledge, however, studies that provide evidence to the contrary (p. 226). I seem to be able to make the switch myself, perhaps revealing that familiarity with the figures is an important variable. Thomas's other empirical argument against pictorialism is that many of the effects taken to support the pictorial theory of imagery (time taken in mental rotation experiments, size/inspection time effects, etc.) exist in congenitally blind people. But this is not so much a refutation of pictorial theory as a potential undermining of certain forms of empirical support for the theory. If Thomas is correct that congenitally blind subjects could have no pictorial imagery, which seems to be the assumption, then it does imply that similar behavior in sighted subjects

doesn't necessarily suggest the existence of imagery. But it is not evidence against the existence of mental pictures. In any case, picture theory is arguably still the default view in cognitive science.

Despite my criticism, Neisser and Thomas have done a great service in developing this account. If we abandon the anti-representationalist commitments of the sensorimotor theorist – claims that are implausible anyhow, particularly when it comes to phenomenal content - there is much to be gained from sensorimotor accounts of perception and imagination in explaining reality testing.

5.3. Sensorimotor reality testing

In this section I will present a sensorimotor account of reality testing. The sensorimotor contingencies associated with perception differ from those associated with imagination and episodic memory. It is our ability to detect those differences, I contend, that explains our reality testing ability. I will explain how these contingencies differ for perception and imagination/ episodic memory, constituting distinct feature spaces. I will not take a stand on whether imagination constitutes a distinct feature space from that of episodic memory. Each sensory modality involves fundamentally different sensorimotor contingencies, and so each merits its own discussion, followed by some discussion of how these modalities can interact in reality testing. I will focus primarily on audition and vision, as there is far more research on these modalities, but I will offer some comments on smell and taste. The sensorimotor contingencies described will in some cases be derived from commonsense reflection and in other cases from empirical research. I do not claim that my descriptions of the sensorimotor contingencies

involved in reality testing are exhaustive, but they are sufficient to provide a framework for a novel account of reality testing. Finally, I will consider cases where reality testing fails regularly - in dreams and when there is imagined content in perception – and show that these cases are consistent with my account.

5.3.1 Auditory reality testing

Consider the scenario from chapter 1. You are lying in bed, unsure whether you are hearing your roommate's television or imagining it. The example is deliberately auditory, as we shall see that this kind of failure of reality testing is more common than visual hallucination. This is because the human auditory system is not as effective at spatial localization as the visual system (Lotto & Holt, 2011). Unlike the retina, spatial information is not directly presented on the eardrum. Instead, the brain uses a variety of cues to extract the position of a sound source in space (see Grothe et al., 2010 for a review). The properties of a given auditory sensation will change as the head moves, and it is ambiguity in these sensorimotor contingencies that produces the failure of reality testing. It is, in part, your relative immobility in bed that makes the scenario plausible.

The basic auditory cues for spatial location are few. The outer ear attenuates certain frequencies of the incoming sound wave depending on the angle of its approach. This monaural cue indicates whether a sound source is parallel to the ear, above, or below it. Other cues for location are binaural. Interaural time difference is used to detect location in the horizontal plane. A typical sound wave will reach each ear at a slightly different time, and submillisecond differences can be detected by the brain. And finally, when one ear lies in the auditory shadow of the head there will be interaural differences in sound intensity, which the brain can also use to

locate a sound source. The latter strategy tends to dominate in higher frequency ranges (2,000 Hz and up). These binaural strategies can detect differences in location as small as a few degrees.

Imagined sounds do not enter the ears, and thus there is no change in frequency attenuation, interaural time difference, or interaural sound level when the head is turned. In many circumstances, a very slight turn of the head and the concomitant change in sensory information is sufficient to indicate that the sound is perceived. But when the sound is faint, distant, and reverberates off walls and other surfaces, the differences in the information entering each ear can be negligible or ambiguous. Differences in wave phase, the basis of interaural time difference detection, are differences in the amplitude of the wave at a given point in the cycle. These differences are negligible when the sound is faint. Faint sounds also make interaural intensity difference difficult to distinguish. The interaural time difference is also confounded when the sound source is distant, as there is a minimum threshold of one degree of spatial separation for detection of spatial difference.

When a sound wave bounces off the walls of a room, changes in head position may not change the phase information received by each ear in ways that are simple enough to be detected by the auditory system. In simple situations the intensity ratio between the direct and reflected sound can indicate the location of the source, but a faint sound reflected through halls, walls, and the like will not allow the brain to distinguish between the sound source and its reflection, as all the waves will be reflections. These reverberations can also confound the monaural location system based on frequency attenuation, as the sound will enter the ear from multiple directions.

In this way audition is fundamentally different than vision. In vision nearly all the information we receive is from reflected light, but the speed of light guarantees that these reflections will reach the eye at essentially the same time, and the lens assists by organizing this

light into a coherent spatial layout on the retina. Audition is much more informationally dependent on the original source and is much more easily confused by reverberation. If one were to get out of bed and move closer to the possible source of the sound, reality testing would likely kick in. If perceived, the sound would get louder and if imagined, it would not. But when lying in bed these factors conspire to leave the situation ambiguous between an imagined and a perceived sound, despite the subtle head movements that are still available.

The cues available to the auditory system leave so much room for ambiguity that there are more ambiguous auditory events than there are failures of reality testing. I do not regularly mistake an owl's faint hoot for an imagined hoot when lying in bed at night. This is not just a problem for reality testing. Ambiguity permeates auditory perception and auditory "scene analysis," the everyday parsing of auditory information into a coherent picture of the world, solves such problems regularly. Timbre is one example. The character, or timbre, of a sound (say a musical instrument) is not determined by the note being played, but by the many overtones that accompany the note. This distinguishes middle C played on a trumpet from middle C played on a clarinet. But the basic monaural and binaural cues are demonstrably insufficient to account for our ability to discriminate overtones (Nordmark, 1970). Similar puzzles abound in auditory research. Speech provides a wealth of such examples, including our ability to fill in missing syllables in speech and to disambiguate other syllables deliberately designed to be ambiguous at the level of known auditory cues (Liberman & Mattingly, 1985; Mann, 1980). Bregman (1994) distinguishes "primitive" auditory heuristics from "schema-based" auditory scene analysis. While the former are likely innate, schema-based analysis is learned and allows for auditory scene analysis when the primitive cues are inadequate.

Primitive auditory heuristics for grouping sounds include frequency proximity, spectral similarity, and correlations of changes in acoustic properties. That is, single sound source will tend to produce sounds within a particular frequency band, produce similar overtones from moment to moment, and the various properties of a single sound will change together, as when the source moves from one room to another. Primitive heuristics can accomplish quite sophisticated discrimination tasks, and they frequently parallel Gestalt principles of visual scene analysis (Bregman, 1994). Such Gestalt principles include the assumption that sounds similar in quality derive from the same object even when they are interrupted by other sounds, much like occluded visual objects (Fig. 2). Likewise, three pure tones may be assigned to two sources in different combinations depending on context clues, which include closeness of frequency of the tones and simultaneity of onset (Fig. 3) (Bregman, 1994, p. 30).

Figure 1.15
Tonal glides of the type used by Dannenbring (1976). Left: the stimulus with gaps. Right: the stimulus when the gaps are filled with noise.

Figure 5.2 (from Bregman, 1994, p. 28)

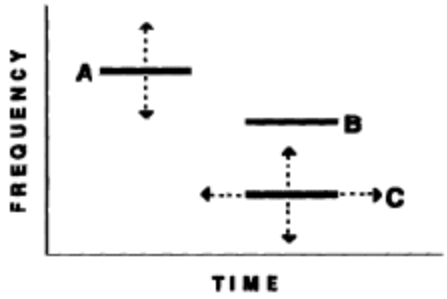

Figure 1.16
Stimulus used by Bregman and Pinker (1978).
A, B, and C are pure tone components.

Figure 5.3 (from Bregman, 1994, p. 29)

But primitive heuristics are not sufficient to explain more sophisticated problems of scene analysis. Schema-based organization makes use of prior learning when the primitive cues are insufficient to resolve ambiguity (Bregman, 1994; see Lotto & Holt, 2011 for a review). Dowling shows that a familiar tune can be more easily picked out of distracting tones of similar volume than a novel one (Bey & McAdams, 2002; Dowling, 1973). Familiarity also facilitates the illusory continuity of a tune through interrupting noise (DeWitt & Samuel, 1986). Context effects have been demonstrated for "streaming," which is the tendency for an alternating tone, when sped up, to be heard as two simultaneous tones. The necessary frequency difference for streaming to occur can be altered by presenting different patterns of tones before the trial (Snyder et al., 2008). Ihlefeld and Shinn-Cunningham (2008) show that primitive cues for object location provide different results when attention is focused on different properties of the sound. Schema-based analysis will sometimes make use of the same cues as primitive analysis, but in different ways. Van Noorden (1975) shows that intention can alter the frequency separation required for integration or separation of a series of tones. When we are trying to hear two tones

as coming from the same source, we allow for a wider range of frequencies to count as coming from the same source than the primitive system would otherwise. And the converse is true when we are trying to hear two tones as coming from different source. Likewise, when we have the schema of a particular melody in mind, we will ignore frequency separations that would normally indicate different sound sources. Hofman, Van Riswick, and Van Opstal (1998) demonstrate that by modifying human subjects' outer ears with molds they can temporarily disrupt their sound localization abilities, but once the new sensorimotor contingencies are learned, the ability is regained. Holt and Lotto (2006) created an artificial auditory "feature space" (my words, not theirs) by creating synthetic auditory objects composed of sets of modulating tones and training subjects to distinguish one set from another. They then determined the primary cue subjects were using to track the objects (perceived "central frequency" of each set) and were able to manipulate the predictive utility of that cue and induce subjects to shift their weightings in favor of another cue ("modulation frequency" – the frequency with which the tones of the entire set modulated up and down in frequency).

Something like schema-based organization must play an important role in reality testing, and it is one reason I employ the feature space construct. Not every ambiguous sound experience will result in a failure of reality testing, because context will lead us to place the experience into the proper feature space. The characteristic low-battery beep of a smoke detector will not necessarily be mistaken for imagination, even if similar conditions obtain to the example of the roommate's television. Once the question arises, however, as to whether the beep is imagined, and if the volume and other relevant properties of the sound are in a specific range, confusion may ensue. Likewise, the experience of a tune stuck in one's head under conditions of restricted movement may be at certain points qualitatively indistinguishable from that tune being played

faintly on a stereo. Various contextual clues will tend to disambiguate the nature of the experience, however. If the tune is familiar, frequently played in imagination, or the ambiguous experience occurs in the context of a series of unambiguous experience, failure of reality testing is less likely, and it is more likely that the experience will be placed in the appropriate feature space. Bregman (Bregman, 1994, p. 402) notes that schemata can be nested, where lower-order schema become part of higher-order schema, echoing Grush and Cussins' accounts of sensorimotor content. The imagination and perception feature spaces, on my view, are essentially higher-order schemata constructed on the basis of sensorimotor contingencies, but once constructed they have an attractive power of their own that may resolve ambiguity in sensorimotor contingencies. Importantly, a schema needn't consist of a series of beliefs. It is simply a higher-order sensorimotor contingency.

The sensorimotor approach to reality testing resolves some of the trickier problems in the literature on reality testing and makes some predictions. To take one example, a song playing in one's head differs in several ways from songs played on the radio. Contra McGinn and Hume, the imagined song cannot be distinguished by its being willed - often the song reappears, even when we strongly desire that it stop. Nor is it distinguished by its lack of vividness – songs overheard in the environment are frequently faint and noisy. But a song in imagination does differ in significant ways from the same song played in the environment. In the age of recorded media, a song played in the environment will typically have the same key and tempo, whereas in imagination the key and tempo may shift significantly on different occasions. This is an artifact of a contemporary age and predicts that in environments without tuners, sheet music, or recordings it would be more difficult to distinguish imagined from perceived songs. Given the

evidence that subjects' sensorimotor contingencies can be altered (Hofman et al., 1998; Holt & Lotto, 2006) such failures of reality testing might be produced in the lab as well.

Of course, McGinn is correct that it is possible to will an imagined song to stop in a different way than a perceived song, which offers additional information that the song belongs to the imaginary feature-space, even if responsiveness to the will alone is not sufficient. The effects of the will can be viewed as a kind of "sensoricognitive" contingency. There is no reason in principle why reality testing ought not be responsive to sensoricognitive contingencies as well as sensorimotor contingencies. It might be the case that willing is accompanied by some motor activity as well, in which case sensoricognitive contingencies might be a type of sensorimotor contingencies. But I needn't try to prove that cognitive "actions" like willing involve motor commands. If there is a nonconceptual analogue of willing, then a creature without mental state concepts might still learn the sensory contingencies related to it. To take Cussins' framework, mastering the sensory contingencies associated with willing would create cognitive trails that define a feature. Where I go beyond Cussins is that in this case the feature thus defined is mental rather than environmental. Sensoricognitive contingenices do not presuppose metacognition in the same way that reality testing does. The subject needn't understand willing as a distinct sort of phenomenon from other motor commands. What it is for a creature possessed only of nonconceptual content to engage in the kind of metacognition involved in reality testing is for that creature to construct distinct feature spaces for perception and imagination/ episodic memory. In this way they understand and behave toward perceived food as perceived and imagined food as imagined, though that understanding does not involve concepts of perception or imagination.

While the owl outside my window is unlikely to be sorted into an imagination schema, as it is rare that I imagine owls, voices are more likely to result in failures of reality testing. This is because one's own inner voice is frequently entertained in imagination. My account of reality testing predicts, then, that when reality testing fails it will frequently manifest as confusion about the source of voices, as the inner voice is one of the most common items in auditory imagination. Among patients with a schizophrenia diagnosis, studies have shown between 25% and 94% report vocal auditory hallucinations, and the prevalence is also relatively high (10-39%) in the general population (Shergill et al., 1998). Auditory hallucination is also more common than nonauditory hallucination, and the latter rarely occurs without the former (Bentall, 1990).

In the early 20th century, auditory hallucinations were theorized as thoughts "becoming sensory" (Mayer-Gross & Stein, 1928), that is, as a form reality testing failure. More recent theorists have revived this approach (Sterzer et al., 2016). Auditory hallucination takes many forms, not all of which are failures of reality testing, and some of which are very different kinds of reality testing failures. In schizophrenic hallucination there is often an auditory experience of a voice located in the mind but attributed by the subject to another agent. In some cases, the patient does not report *hearing* the voice, instead reporting it as a purely cognitive experience. In other cases, patients experience their own thoughts as perceptual objects "projected" into physical space ("thought echo") (Pienkos et al., 2019).

There are many competing theories of schizophrenic hallucination. A common theme, however, is that the source of these abnormal experiences is a higher-level abnormality in applying proper context to an experience. Often schizophrenia manifests not in hallucination but in a heightened importance or salience attributed to everyday experiences, which has been explained as a disruption of context (Matussek, 1987). Matussek explains the phenomenon as a

kind of Gestalt shift, in which certain sensory qualities dominate the overall perceptual experience and result in an attribution of increased significance. Visual "perceptual organization" tests on schizophrenics show disruptions in the ability to make figure-ground discriminations in visual stimuli (Panton et al., 2016) and inability to see stimuli as Gestalts (Pienkos et al., 2019). Failure to attribute the proper context to an experience, on my account, could result in a failure to place an experience in the proper feature space. An experience that is ambiguous with respect to its lower-level cues – say an inner monologue when one is alone and relatively still – would normally be automatically placed in the imagination feature space since no one else is present. But a schizophrenic's shift of focus to the details of the auditory qualities of the inner voice rather than the overall context might cause the sensation to be placed in the perceptual feature space.

Auditory hallucination experienced as another agent's inner speech inserted into the patient's mind can also be interpreted as a failure at the level of schemata. The schizophrenic knows that the voice is in her head and not in physical space but claims that the source of the voice is not her but some other agent. This is frequently taken as evidence for a McGinn-style view which places the heuristic for reality testing as responsiveness to the will. However, another explanation is possible. Voices are likely to engender failures of reality testing at the level of schema because we frequently experience both inner and outer voices in a range of situations. But primitive auditory scene analysis will succeed at reality testing imagined voices in situations where the subject is able to move. An impairment of schema-based reality testing, then, might result in borderline cases where the voice is accurately determined to be inner by primitive heuristics, but disruption of schema-based reality testing still places it in the perceptual feature space. The conflict is resolved either by categorizing the experience as an external agent inside

the head or as “thought echo,” where the inner voice escapes into the external world. If the context-failure account of schizophrenia is correct, then my account of reality testing predicts that the inner-voice insertion form of auditory hallucination should be more frequent when the subject is in an environment that reduces ambiguity at the level of primitive heuristics – fewer walls and less restricted movement.

Other accounts of schizophrenia identify the relevant contextual failure as a distortion of the sense of self, in which the subject, and thus her experiences, feel unreal or, conversely, to be the only reality (Pienkos et al., 2019). In either case, we can predict an impairment of one’s ability to sort sensory experience into the appropriate feature space. ‘Self’ is a broad and ambiguous term, but one approach would be to identify the self in part with the sensorimotor bodily space defined by Grush. Distortion of the boundaries of that space, which is defined by a set of sensorimotor contingencies, would affect the subject’s ability to properly sort sensations into “inner” and “outer” feature spaces.

Bentall (1990) explicitly identifies the reality testing failures in schizophrenic hallucination as a metacognitive impairment. He argues that distortions at the level of high-level schemata may be paired with a “criterion shift,” which causes schizophrenic subjects to make hasty categorizations of experience on the basis of the distorted Gestalt, which could result in failures of reality testing even when lower-level cues are not ambiguous. Techniques for reducing auditory hallucinations appear to tackle the problem at both levels. “Time-out” strategies attempt to shift the gestalt by placing the patient in a different context, interestingly, sometimes by placing the subject alone in a room or lying down (Shergill et al., 1998). Such situations might increase the ambiguity of low-level cues, but the shift in context can be sufficient to change the overall auditory gestalt. Other techniques tackle the misinterpreted

subvocalization itself by humming or reading aloud (Green & Kinsbourne, 1989), or simply opening the mouth (Bick & Kinsbourne, 1987), imposing different motor cues than those associated with inner speech.

Lower-level sensorimotor cues may also be distorted in schizophrenia. The failure of the sense of self is sometimes connected with anomalies in the sense of time (Pienkos et al., 2019), which cause predictive failures (more on this is section 5.3.5). This is essentially a direct distortion of the subject's understanding of the sensorimotor contingencies themselves, which could create failures of reality testing at all levels – both primitive and schematic.

In this section I have introduced my account of reality testing by introducing the sensorimotor contingencies associated with spatial localization in audition. These include the three basic strategies of frequency attenuation by the outer ear, interaural time difference, and interaural intensity difference. There are other "primitive" heuristics for assigning auditory qualities to spatial objects as well, including grouping by similarity in quality, tone, etc. There are also learned schema by which we categorize auditory objects. I identified these schemata, both lower- and higher-level as feature spaces. Among the auditory feature spaces are spaces for perceived sounds and imagined sounds. I then surveyed a number of ways that this sorting of auditory experiences into the perception and imagination feature spaces can fail and how these failures are failures of reality testing.

5.3.2 Visual reality testing

Debate over the representational format of visual imagery has spurred a great deal of research demonstrating similarities between visual perception and visual imagery. Sensorimotor reality

testing, however, requires that there be sensorimotor differences between perception and imagery. Neisser and Thomas's sensorimotor accounts of imagery hold that there is no difference in motor activity between perception and imagery, only the presence or absence of the object of perception. But there are differences, some obvious and some more subtle. In my discussion of audition, I began with the most subtle, primitive sensorimotor cues that differentiate between perception and imagination and worked up to larger schemata. In this section it is more illustrative to work backward. The most obvious differences between visual perception and imagination can be seen in the context of gross locomotion and high-level schemata. Once the broad distinctions between the feature-spaces of visual perception and imagination are sketched out, we can return to the issue of finer motor behavior like saccades and survey some recent research that reveals more subtle differences between perception and imagination.

The most obvious distinction between visual perception and imagination has no parallel in audition – we can close our eyes but not our ears.[11] Perceived objects disappear when we close our eyes, whereas visual imagery can persist, and is often enhanced, when the eyes are closed. Having one's eyes closed is an obvious cue that the images one entertains are not perceived (except when dreaming, see section 5.4.1). Mental images can also be made to disappear, but by different means – opening one's eyes or willing the image to cease. The appearance of perceptual objects is also subject to the will, but in different ways. I can induce a perceptual object to disappear, but it will require closing my eyes, moving the object myself, calling someone else to move it, or moving my body so that the object is not in my perceptual field. No single contingency alone can be picked out as the distinguishing factor between visual perception and imagination (as McGinn attempts to do with the will, see Ch. 4). But the set of relevant

[11] We can put our fingers in our ears, of course, but we do this infrequently.

contingencies differs between the two. Of course, in practice we rarely if ever need to move an object to determine whether it is perceived. A subconscious detection of the disappearance of the object when we blink, or as we shall see later a saccade or subtle move of the head, would be sufficient in most cases to sort the experience into the perceptual feature space.

We do experience visual imagery with our eyes open, and even then there are sensorimotor differences from perception. Afterimages offer a clear example. If I stare at a pink sheet of paper for 20 seconds and then move my gaze toward a white wall, I will see an image of a green square. As I move my eyes or my head, the afterimage moves. I may even walk around the room, but the image does not change any of its visual properties, apart from a slow fading from existence. The same is true of floaters and phosphenes.[12] Compare this to a perception of a green sheet of paper. A sufficient turn of the head will bring it out of view entirely. Walking around the paper will cause its visual properties to change from rectangular when viewed dead-on to a rhomboid shape when viewed form an angle. Any failure of reality testing occasioned by an afterimage will be brief, generally lasting only until the next saccade of significant size or turn of the head, at which point the differences in sensorimotor contingencies from typical perceptual objects will be revealed.

Other large-scale differences between perception and imagination include the flexibility of space in imagination. I may imagine the Empire State Building and then scan across the Atlantic to the Eiffel Tower in a split second in imagination. I can imagine the visual experience of traversing the route from my bedroom to the front door much more quickly than I could perceive it, as the latter would require actually walking the route.

[12] My argument does not depend on a metaphysical claim that afterimages, floaters, and phosphenes are products of the imagination and not of perception. I only claim that human reality testing classes them as such.

Differences in sensorimotor contingencies at this larger scale are obvious and suffice to accomplish a great deal of reality testing. It is unsurprising, then, that failures of visual reality testing are relatively rare. Failures of auditory reality testing are more frequent for a few obvious reasons. There is no auditory equivalent of blinking, and our visual systems are inherently spatial whereas our auditory systems must reconstruct spatial information on the basis of temporal information.

And unlike the ears, the eyes constantly perform tiny saccadic motions that help disambiguate information and prevent failures of reality testing. Sensorimotor accounts of perception add the advantages of a temporal element to vision by positing that our visual experience is built up by a series of foveations – fixations of the fovea on some aspect of the visual scene. Even on such accounts, each foveation does generate information that is essentially spatial. This means less inference is required to generate spatial information and less inference means less opportunity for ambiguity. While the case of the noisy roommate might be sufficient to trick the auditory system, the visual system has much more information to go on, even when lying relatively motionless in bed. But there is enough ambiguity in the visual information received that failures of reality testing do occur and can be reliably induced in a laboratory setting, the most famous example being the Perky effect.

In chapter 4 I discussed the The Perky effect. The Perky effect induces failures of reality testing by asking subjects to imagine an object on a blank screen and surreptitiously projecting a similar image on the screen. Rather than mistaking imagination for reality, subjects mistake reality for imagination. This is a sort of inverse of visual illusion, where the visual experience of an object is distorted or masked by the imagination. Here the projected perceptual object alters or masks the mental images - the subjects report their images to have properties matching those of

the projected image, even if their mental images initially lacked those properties (e.g., a vertical orientation reported even if the object was originally imagined as horizontal). The illusion is induced by requiring subjects to sit relatively still and concentrate on a blank screen. This eliminates some of the sensorimotor contingencies used in reality testing. A turn of the head might separate the projected banana from the imagined one (or eliminate the mental image altogether, as mental imagery is notoriously fleeting). The contingencies that remain are saccades over a constrained space. More importantly, the subject is primed to expect that any banana-like sensations will be imagined. The induced ambiguity of the lower-level cues along with the priming of the higher-level schema produce the effect.

One might wonder, however, how even under these constrained circumstances subjects might confuse a horizontal projection of a banana with an imagined vertical banana. Surely the relevant sensorimotor contingencies would differ in either case. But this would depend on how finely distinguished the sensorimotor contingencies are. That is, the relevant sensorimotor contingency might be that, when imagining a banana, a certain percentage of our fixations ought to produce yellow imagery and that they should do so in an oblong spatial area, with no rigid expectations about just where that yellowness might be and how it might be configured. More specific instructions, or individual differences in imagery ability, might make the contingencies more specific.

More ambiguity in sensorimotor contingencies involved in the Perky Effect is produced by the lack of depth cues. The image was projected on a flat screen, and the subjects were directed to imagine an object on the flat screen as well. We rarely endeavor to entertain a mental image with eyes open while moving, but a useful proxy is, again, the phenomenon of the afterimage. If the production of an afterimage were not deliberate – say I had been studying a

square, mostly green, abstract painting intently before moving my gaze to a wall in the gallery – I might temporarily be confused as to whether the square, pink image was another piece of art, particularly if I were not a philosopher very accustomed to inducing and pondering afterimages. But the experience would be disambiguated quite easily by moving toward the image. There would be no "looming up" of the pink square (Lee & Kalmus, 1980). Rather the size would remain constant, whereas perceived objects will project a progressively larger image on the retina as I come closer. This difference in sensorimotor contingencies allows me to sort the experience into the imagination feature space. If subjects were allowed to move forward or backward, the Perky Effect would likely not occur.

The debates between Kosslyn and Pylyshyn regarding the representational format of mental imagery contain a wealth of data regarding the similarities between mental imagery and perception (Kosslyn, 1980). Many studies show that reaction times in imagination tasks resemble those in perceptual tasks. Traversing locations on an imagined map takes longer when those locations are farther apart (Kosslyn et al., 1978) and judgments of whether a perceived arrow points toward a remembered dot take longer the further the arrow is from the original location of the dot (Borst & Kosslyn, 2008; Finke & Pinker, 1983). Subjects report similarities in eye-movement between perception and imagination in such tasks, and eye-tracking experiments confirm the similarities (Brandt & Stark, 1997; Holsanova et al., 1999). Such similarities should induce failures of visual reality testing if other cues are not available, and this is precisely what occurs in the Perky effect.

There are, however, differences in saccade patterns between perception and imagination that could be exploited for reality testing. Borst, Kosslyn, and Denis (2006), for example, give subjects the same task of locating a remembered dot on a square screen in both an "eyes-closed"

and “eyes-open” conditions. They found no correlation in the scanning speed between the tasks, suggesting a different pattern of saccades. Allen (M. Allen, 2019) finds that subjects scan significantly faster when asked to trace the trajectory of an imagined dot against a real screen with eyes open than when they trace the same trajectory with eyes closed. Allen does not offer a suggestion as to why the eyes-closed case is slower, but there is an explanation that accords with a sensorimotor account. In the eyes open case there is less work for the imagination to do – one need only imagine the dot and trace it across the perceived field. In the eyes-closed condition, there is more processing involved as the field itself must be imagined as well, slowing the entire process. In the eyes-open case the world does much of the work. While it is sometimes the case that longer distances can be traversed more quickly in imagination - I can imagine going down my stairs and into the kitchen much faster than I can perceive it. But when more detail and precision is required in what is imagined, it appears that imagination can be slower than perception in traversing distances.

Gurtner et al. (2019, 2021) find similar results to Allen’s unpublished study, using eye-tracking to show that subjects perform more repeated fixations on the same areas in visual imagery than in perception. Repeated fixation is required to maintain the image, as mental images fade more quickly than perceptual images (Gurtner et al., 2019; Hassabis & Maguire, 2009; Kosslyn, 1991; Theeuwes et al., 2009). This slows imagery task response times in eyes-closed conditions compared to eyes-open conditions. Here the differences in behavior reflect an underlying difference in sensorimotor contingencies. A saccade and re-fixation in perception will typically produce sensory input in the form of a clear image, or at any rate an image that was as clear as it was during the last saccade. If mental images fade over time, then a similar saccade in imagination conditions will produce sensory input that is degraded relative to the last fixation

and must be refreshed by memory. This difference allows us to distinguish mental imagery from perception when immobility produces ambiguity in other cues.

Gurtner et al. (2021) find another difference between visual perception and visual imagery – mental images are more "fragile" than perceptions, meaning that they are more easily disrupted by concurrent perceptual inputs like flashes of light or other surprising visual inputs. This leads to a general strategy of fewer and less spread-out eye fixations in imagery, along with the more frequent recurrence to the same locations required for maintenance.

Contra Neisser and Thomas, then, the motor routines associated with visual imagery differ in important ways from those associated with perception. Imagining a cat is not simply a reenactment of seeing a cat without the cat. Imagining a cat requires a different temporal pattern of saccades, involving more frequent refixations because the sensory inputs differ systematically from those in perception, namely fading quickly without constant maintenance by working memory. This evidence supports my claim that Neisser and Thomas are wrong to attempt to do away with representations altogether – imagination provides sensory content and successful reality testing involves the recognition of how that content differs vis-à-vis our motor commands from perceptual sensory content.

But if imagination can be distinguished from perception in virtue of these saccadic sensorimotor contingencies, then why should the Perky Effect obtain at all? When the subject is before the screen, imagining a banana, there are only a few forms of sensory feedback a given saccade might produce. A saccade might sometimes encounter blank screen. If the imagery at that location has faded, and it happens to be an area at which one was imagining yellow, then a command to refresh the imaginary sensation at that location will ensue, but no expectation will be violated. At other times a saccade might encounter a faintly projected image – a yellow

sensation. If the expectation is that the image should have faded by now in that area, this will be inconsistent with the sensorimotor contingencies. But whether this is registered by the sensory system will depend on whether and how precisely the timing of the fading image is monitored by the brain. It is entirely consistent with the data that this information is monitored somewhat imprecisely, in which case no expectation is violated by a yellow sensation in an area of the visual field where yellow sensations had recently been generated by imagination, even if the sensation would normally have faded by now. If the images are only yellow on white, there is no strong violation of expectations, even if the projected image is configured differently than the mental image. If there is a relatively thin outline of black, there may be some violations of sensorimotor expectations – encountering the black tip of the banana where one expects yellow or a blank screen – but occasional violations of expectations will be discounted by any real-world, noisy system. And anyhow, any subtle violations of sensorimotor expectations would likely to be insufficient to trump the contextual effect of being told that one's task is to imagine a banana and, prior to the gradual introduction of the image, the majority of the saccades being consistent with imagination.

I should say more about context effects or "schemata." One way that the term has been employed is by Neisser, who takes the word to be more or less synonymous with a pattern of motor routines associated with some perceptual experience. To imagine a cat, he claims, is to employ the cat schema in the absence of a cat. Neisser's account predicts far more failure of reality testing than we experience, since at best our perceptual schemata can only be accurate in very familiar environments. The world is often unpredictable, and so there will be frequent mismatch between the perceptual schema employed at any given time and the sensory input. If failure of match were interpreted as imagination, then the results could be disastrous. The

sudden, unexpected appearance of a child or a car where empty street is expected would presumably induce one to interpret the street, or at any rate that portion of it, as imagined. The problem is the assumption that the same schemata are in play in both perception and imagination. What is necessary are two distinct feature spaces, each with their own sets of sensorimotor contingencies. Schemata, as I use the term, are priors regarding which feature space we ought to sort an ambiguous stimulus into. Thus, we do not take the auditory experience of the owl's hoot to be imagined even when we are lying still in bed and unable to perform the motor routines that would determine the source of the experience. When driving, a context of safety-consciousness biases us to assume that all stimuli are perceptual, thus we are likely to swerve to avoid an imagined coyote but unlikely to interpret an unexpectedly occupied stretch of road as imagined. Context effects have been demonstrated experimentally for propositional reality testing in children (see Woolley & Van Reet, 2006 for a review), but there has been little research on reality testing in its nonconceptual form. There are also numerous context effects in visual perception (see Todorović, 2010 for a review). It should be expected, and everyday examples confirm, that visual reality testing also exhibits such effects.

5.3.3 Reality testing in other modalities

Taste and olfaction are often considered to be quite different from senses like vision and audition. Some claim they lack distal objects and that they consist only in an immediate sensation. If this account is correct, then one could argue that there is no need for reality testing in these modalities, as the internal/external distinction just does not exist for smells and taste. That would not be a problem for my view, so long as reality testing exists in some modality. But

if that account is wrong, then at least a preliminary account of reality testing for taste and smell is required. Millar (2021) has argued for a sensorimotor account of taste and smell that includes a notion of perceptual objects in these modalities. The view is at least as plausible as the opposing one – the perceptual object of a rose smell is the rose itself. My discussion of representation-mediated taste aversion in chapter 2 also provides evidence that reality testing does exist for taste. No research exists on reality testing for olfaction that I am aware of, so my comments will be somewhat anecdotal and speculative. But I hope to sketch out a plausible framework for extending my account of reality testing to olfaction.

The cues available to locate the source of a smell are less subtle than those available for vision and audition. The mind does not appear to calculate differences in input between the nostrils, but simple variations in intensity are available. A perceived smell will intensify as you approach its object, while an imagined smell will not. Differences in sensory character other than intensity are available as well. The sensory character of a smell or taste is largely a result of the chemical composition of the object. In taste the chemical composition can be explored more thoroughly by motions of the tongue and the mouth. More complex, learned contextual cues are available as well. For example, priming effects have been demonstrated for flavor. Subjects will claim an identical flavor for different drinks poured from the same jug (Woods et al., 2010).

Failures of reality testing for smell are rarely noticed in daily life. It can be impossible, or at least impractical, to verify whether there is an actual source of a strange smell experience. Perhaps someone is cooking roast beef next door, but we will never search thoroughly enough to confirm it. Vision provides a specific location to search, whereas locating the source of a smell is more labor intensive. It is also more difficult, if not impossible, to will an imagined smell or taste experience to occur. But I can relate a personal anecdote of olfactory reality testing failure that

illustrates how my account can be extended to smell. A sinus infection left my sense of smell impaired for some weeks. During this time, the impairment did not eliminate any smell experience. Rather, I only experienced one smell, a yeasty sort of scent that never went away, changing only in intensity. My initial reaction was to interpret the smell as of something in the world, a puzzling experience as no such object could be located. This is what contextual cues demanded – smell experience is rarely imagined, or rarely determined to be so. But it was if I were smelling the odor of the entire world, since every smell was the same. Eventually, the lack of variation in quality or correlation with any known smells from my past did cause the odor to present itself as imagined. The experience was so unlike other smells that it was sorted into a distinct feature space. Gradually, the variations in intensity and their general correlation with my nearness to objects in the world snapped the experience back into the perceptual feature space and the smell was presented as the smell *of* a given object – it was just that all objects had a similar smell. The experience is analogous to the experience Grush describes of learning the sonic guide, albeit an impoverished version involving only one manifold along which information can vary. It is unsurprising, then, that Millar (2021) cites intensity as the primary cue that the olfactory and gustatory systems use as the basis of the perceptual constancy that undergirds the objectivity of those modalities.

In Chapter 2 I argued that studies of representation-mediated taste aversion in rodents show that temporary gustatory reality testing failures can be induced by classical conditioning. After pairing a flavor with a tone and then with nausea, the rats hallucinate a flavor in plain sucrose solution when exposed to the tone. The effect would dissipate eventually. Here the sensory contingencies involved are multimodal – an auditory cue is associated with a flavor. Our understanding of the sensorimotor cues involved in flavor is currently limited. Intensity is one

cue, but the sensorimotor contingencies related to the character of the flavor are not well understood. It is known that flavor is experienced as weaker when the mouth does not move (Burdach & Doty, 1987), and that the more complicated sorts of mouth movements involved in wine tasting can clarify the perception of flavor (which also involves olfaction) (Buettner, 2001). Swallowing also enhances flavor (Burdach & Doty, 1987). If it is correct that these motions reveal aspects of the flavor that are not apparent without such motion, it would follow that distinct flavors will initially seem more similar before such motions are accomplished. It seems, then, that in the case of RMTA in rodents we have a case of a strong priming for a certain flavor experience along with an inherent ambiguity in sensorimotor contingencies involved in taste (as opposed to, say, vision). Likely the dominant flavor in each case is of sugar, increasing the ambiguity. The priming from the tone, along with limited exploration through mouth movements because of the priming for nausea (we would rather not savor a taste we expect to find nauseating) trumps the somewhat more subtle difference in flavor, causing the failure of reality testing. There is also likely overlap in the chemical constituents of the two flavors, creating more ambiguity until extended tasting can sample a broader array of those chemicals. The more experience the rats have with moving the flavored sucrose in the mouth, swallowing it, smelling it, and so forth, more aspects of the flavor are revealed, the ambiguity is eventually overcome, and the hallucinatory effect diminishes – thus the extinction effect observed in studies of RMTA. Similar effects have been demonstrated in humans, where drinks were judged to be more similar by manipulating expectations (e.g., pouring from the same jug). There is a limit, however, where differences were too pronounced to produce the effect (Woods et al., 2010).

5.4. When Failure of Reality Testing is Normal

5.4.1 Dreams

My account of reality testing so far has sought to explain how it is that humans are so overwhelmingly accurate at distinguishing between perception and imagination in everyday life. We seem to exist in two distinct worlds, constantly bombarded by experience both perceptual and imaginary, yet we rarely find ourselves confused on such matters. Indeed, the two worlds often complement each other, as when performance on an "eyes-open" visual imagination task is enhanced by perceptual props. Dreaming seems the complete reverse. Our dream experiences are bizarre and sometimes nonsensical, but we seem not to doubt their reality while they are experienced. Dreams, therefore, would seem to be a dramatic failure of reality testing. None of what we experience in dreams is perceived – our eyes are closed and the objects we experience are not there. And yet most of the time we take our dream experiences to be perceptual, at least until we awake. There are also those who experience "lucid" dreams, in which they understand that they are dreaming while dreaming. In this section I will argue that the phenomenon of dreaming is consistent with my account of reality testing.

If what happens in dreams is that we mistake imagination for perception, then at first blush it would seem my account should predict that the sensorimotor contingencies involved in dreaming resemble those of perception more than imagination. There is evidence that something like this is the case. LaBerge, Baird, and Zimbardo (2018) find that eye movements in dreaming and perception exhibit smooth tracking when subjects track the movement of their (dream) thumb in a straight line or circle, whereas in imagination there are repeated saccades. However, this study by necessity involved subjects engaged in "lucid" dreaming. The subjects signal to the

researchers that they are lucid dreaming with a pattern of eye movements before carrying out instructions. They then performed several eye-tracking exercises in their dreams, following a circular or motion of their dream-thumbs. It was found that the eye movements were smooth, like those in normal perception, as opposed to broken and saccadic, like in imagination. The researchers speculate that the reason for the difference in saccade patterns is that dream imagery is more vivid since it lacks competing inputs from the outside world. But the evidence is also consistent with the claim that the eye movements are in part constitutive of the content. The smoother the movement, the more perception-like the imagery. If smooth eye movements are partially constitutive of the perceptual feature space, then it is unsurprising that we mistake dreams for perception. This tendency might in fact be adaptive if dreams serve some adaptive function. It has been suggested that dreams allow us to regulate our emotions in novel situations, by pre-experiencing them and thus making them less novel (Scarpelli et al., 2019). This function would be better served if we believe that we are experiencing what happens in dreams. But the Laberge et al. study does not provide clear evidence in any case, since subjects are aware that they are dreaming, and so in these cases the similarity in eye movements does not induce reality testing failure.

Dreaming, however, is a nonstandard case of reality testing failure. Many of our cognitive and motor abilities are diminished when asleep – this is rather the point, to allow us some rest. But this also means that our reality testing abilities may be diminished. It is well-known that the motor neurons are inhibited during REM sleep, a phenomenon known as "motor atonia." In my own experience, fine motor actions in dreams are difficult or impossible – I have never successfully put on my socks in a dream. It is quite likely that atonia alters the sensorimotor contingencies in dreams and disrupts our reality testing abilities. And while the

brain is active during REM sleep, our normal cognitive abilities are not in peak form. Wild aberrations of the laws of physics and logic abound without surprise in dreams. It is likely that our understanding of sensorimotor contingencies involved in reality testing are likewise diminished.

But the phenomenon of lucid dreaming does suggest that the altered sensorimotor contingencies can be learned and consistently exploited to perform accurate reality testing. Dreams might be more accurately described as a distinct feature space, distinct from perception and imagination, with its own set of sensorimotor contingencies.

5.4.2. Imagination in perception

It may be the case that many of the experiences we "accurately" sort into the perception feature space also involve imagination. In Strawson's account of objectivity, recognition required a mix of imagination and perception. In recognizing the bearded friend, we imagined the clean-shaven version overlaid on to him. Dummett's proto-thought involved a similar process. We might mistakenly perceive a person as aggressive by infusing our perception of her with some imagined facial expression. Dummett's version might be considered a failure of reality testing, and if such projection is common in everyday perception, then failures of reality testing might be more common than I have presumed. Indeed, there are accounts of perception that involve systematic failures of reality testing. Predictive processing accounts of perception posit that the majority of what we call perception is generated by the imagination in a kind of "controlled hallucination" (Clark, 2015). If such accounts are correct, then failure of reality testing is the norm. This might seem at odds with my account. But whether reality testing failure is rare or common is not at the

core of my account. Rather, it is the claim that reality testing is accomplished by an understanding of sensorimotor contingencies. In this section I will argue that my account of reality testing is compatible with predictive processing accounts of perception, in which case it is also compatible with a wide variety of accounts of imagination in perception.

Predictive processing accounts of perception hold that typical perception does not involve the bottom-up processing of every bit of the visual scene. Instead, the visual system scans for unexpected inputs from the world, but the vast majority of what we experience is predicted top-down. It is only in cases of mismatch that we replace the prediction with new information from the world. Imagination and memory are taken to be the source of the imagery produced by the top-down predictions. Thus, predictive processing accounts of perception entail that typical perception is mostly imagination, and thus that reality testing failure is the norm. This in fact helps to clarify what failure of reality testing is. It is not merely a non-veridical experience – the predictions are typically accurate. Rather, reality testing is a failure to identify which cognitive module – perception or imagination – is the source of the experience. Predictive processing accounts do not do away with this distinction between modules, they simply claim that their outputs are often mingled.

The process of prediction involved is often described in propositional terms. The predictions are modeled as Bayesian priors, which are in turn framed as beliefs about the probabilities of a given stimulus obtaining (Williams, 2020). The question then arises as to whether the "imaginary" content added to perception by these predictions is conceptual or nonconceptual. If it is conceptual, then it is not our concern. There is a vast literature on conceptual content in perception, and it may represent a failure of conceptual reality testing in some cases. The relevant question for my account is, if there is nonconceptual imagined content

in everyday perception, can my account explain the systematic failure of reality testing implied by this phenomenon? Let us suppose, for the moment, that there is nonconceptual content in the predictions as well.

As noted, the basic idea behind predictive processing accounts of perception is that sensory input is primarily for error detection. The majority of our experience, it is claimed, is predicted top-down, and bottom-up sensory input is experienced only when top-down predictions turn out false. As such, it makes sense that reality testing would fail in these cases. On my account, imagination is mistaken for perception when there is ambiguity between the sensorimotor contingencies associated with each. Predictive processing accounts hypothesize that in typical perception these match – the predicted experiences are indistinguishable from perception at the level of sensorimotor contingencies. So, it is not a problem for my account that reality testing would routinely fail in these circumstances – such failure would be predicted.

A trickier question is how, when the predictions do not match, the surprising sensory input is seamlessly integrated into perception without a conscious experience of reality testing failure. This is because the predictions of predictive processing accounts are more fine-grained than the sorts of sensorimotor contingencies involved in reality testing. Reality testing predicts how a perceived image will move when I turn my head as opposed to an imagined image. It will not necessarily predict the color information that should hit my retina. This means that a failure in prediction for predictive processing will not necessarily be a failure in prediction for reality testing. The unpredicted data will often still be consistent with the broader sensorimotor predictions of the reality testing faculty. Context, or higher-order schemata, will factor as well, outweighing more severe inconsistencies and resolving ambiguity. An unexpected object on my

desk will not immediately result in my sorting it as imagined, so long as it behaves like a perceptual object.

Predictive processing accounts of the failures of reality testing in psychotic hallucination have proliferated in recent years. Psychosis is sometimes associated in these theories with a greater weight being placed on sensory information and less on the predictive model (Sterzer et al., 2018; Vinckier et al., 2016). For example, the sensory qualities of the inner voice are assigned a greater weight and the experience is taken as a spoken voice even though the subject's predictive model would place the experience in its proper feature space if the relevant contextual factors were properly weighted. Other accounts locate the problem in the predictive model, where false predictions along with a discounting of sensory error signals lead to non-veridical experiences (Elliott et al., 1995; Schultz et al., 1997). Griffin and Fletcher (2017) speculate that both dysfunctions occur, with overweighting of sensory information occurring early in the process, eventually producing a defective predictive model, which in later stages becomes inflexible. In any case, such accounts are consistent with my account of the reality testing failures in schizophrenia and with the dominant theories of schizophrenic hallucination outlined in section 5.3.1. There we noted that some theories of schizophrenia posit an undue focus on aberrant sensory information at the expense of the overall gestalt. This is consistent with the version of the predictive processing account on which too much weight is placed on sensory information at the expense of the predictive model. Underweighting of sensory information can result in reality testing failures as well. The schemata of my account are not the "controlled hallucinations" produced by the predictive model in predictive processing accounts, but schemata do allow that sensory information can be mistakenly taken as imagined on the basis of context.

5.5. Conclusion

In this chapter I have proposed an account of reality testing that does not require mental state concepts. Instead, learned associations among motor movements and resultant sensory inputs aggregate into a nested series of higher-level schemata. Different sets of such sensorimotor contingencies and schemata exist for each sensory modality, and I have described some of the basic contingencies for audition, vision, smell, and taste. Two higher-level, multimodal schemata are the "features spaces" of perception and imagination. Sensorimotor pairings that are compatible with both feature spaces will sometimes induce failures of reality testing. Typically, however, such ambiguities are resolved by contextual features. In schizophrenia the ability to employ context is impaired, and there is a concomitant increase in reality testing failures. Other cases of systematic reality testing failure include dreams and when (or if) perceptual content is partially constituted by imagined content. I have argued that my account of reality testing is consistent with each.

Chapter 6. Epistemic feelings, metacognition, and the Lima problem

Proust (2013) is one of two theorists, to my knowledge, who posit the existence of nonconceptual metacognition, the other being Gallagher (2004), who holds that mindreading is nonconceptual. Proust is the only theorist who offers a fully-developed account of nonconceptual metacognition, and many of her claims are shared by other researchers, including Dokic (2012), Arango-Munoz (2011, 2013, 2014a, 2014b; 2019), and Koriat (1993, 1997, 2000, 2007). This general view of the functional role and representational format of certain metacognitive states is

called ‘evaluativism’. Evaluativism holds that states like feelings of knowing and tip-of-the-tongue experiences are metacognitive and, in Proust’s version, nonconceptual. However, it also claims that these forms of metacognition represent mental states only indirectly. Nonconceptual metacognitive states, on this view, are directly responsive to heuristics like reaction time. In this chapter I will critique evaluativism, arguing that it is unmotivated, unsupported, and ill-conceived. I will then offer a phenomenological analysis of a tip of the tongue experience, arguing that such experiences are better explained in terms of feature spaces.

6.1. The Lima Problem

Do you know the capital of Peru? You may have come up with the answer, “Lima,” and felt confident that this was correct. You may have failed to come up with the answer immediately but felt that you *did* know it. You may have even felt that the answer was on the tip of your tongue.[13] Each of these is an experience that researchers have termed an ‘epistemic feeling’. *Feelings of confidence* are retrospective, indicating that a previous response was correct. *Feelings of knowing* are prospective, indicating that a correct response could be given. *Tip-of-the-tongue experiences* (TOTs) indicate that the correct response is tantalizingly close. These feelings have been dubbed ‘epistemic’ because they assist us in the retrieval and management of mental states that admit of accuracy. They tell us when our judgments were correct, when our memories or other cognitive capacities have the potential for successful performance on some task, and how likely that success may be. If epistemic feelings are about our epistemic states in the intentional sense of ‘about’, then they are also metacognitive states - they take other mental states as their

[13] This example is taken from Dokic (2012)

intentional objects. This has some phenomenological support. A TOT, for example, seems to be telling us that a memory is there to be retrieved. And empirical studies show that epistemic feelings are typically correct (for TOTs, see A. S. Brown, 1991; for feelings of knowing, J. Hart, 1965; Koriat, 1993, pp. 609–610; for confidence, Siedlecka et al., 2016), which suggests that epistemic feelings do function to relay information about the content of our own minds.

This poses a problem for some prominent accounts of metacognition. Epistemic feelings inform us about mental content that we cannot access, a puzzle for theorists who explain metacognition in terms of direct access to our minds (e.g., Locke, 1689; Lycan, 1996) or redeployment of first-order content (e.g., Byrne, 2005). And the work is done by feelings, which are not traditional elements of a theory of mind (e.g., Carruthers, 2011; Gopnik, 1993). Call this set of challenges the Lima problem.

The Lima problem motivates an account of epistemic feelings known as 'evaluativism'. Evaluativists claim that epistemic feelings reveal a distinct metacognitive mechanism, which they associate with "System 1" of dual-process theory (Arango-Muñoz, 2011; Koriat et al., 2008; Proust, 2013).[14] Among the proponents of evaluativism are Proust (2013), Dokic (2012), Arango-Munoz (2011, 2013, 2014a, 2014b; 2019), and Koriat (1993, 1997, 2000, 2007).[15] While these researchers' accounts of epistemic feelings differ in some respects, they all commit to a functional claim, which in turn supports their claims of a distinct mechanism. They claim that the information epistemic feelings provide about our mental states is not derived directly from those states but from heuristics. Call this claim *core evaluativism*. Core evaluativism is supported by a

[14] Some evaluativists (e.g., Proust, 2013) reserve the term 'metacognition' for the System 1 version. This form of metacognition is described in detail in section 2.3. I will use 'metacognition' to denote any intentional relation between mental states or processes. Thus, I will refer to theory-theoretic processes (section 2.2) as 'metacognition', whereas evaluativists would use 'metarepresentation'.

[15] Dokic (2012, p. 312, note 16) stops short of committing to the existence of a distinct mechanism, but does commit to what I will call 'core evaluativism'.

body of empirical research that purports to show that these heuristics are the causal factors to which epistemic feelings are sensitive (e.g., Koriat, 1993; Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992). The most commonly cited heuristics include cue familiarity, recall of related information, and processing fluency.

I will argue that core evaluativism is unmotivated, unsupported, and ill-conceived. The structure of the article is as follows. In section 2 I will introduce the phenomenon of epistemic feelings and discuss a range of views about their nature. I will then consider the Lima problem and argue that it is no problem at all, eliminating one source of motivation for evaluativism. But evaluativists also offer empirical arguments in support of core evaluativism. In section 3 I will describe the three primary heuristics posited as the causal inputs to epistemic feelings– cue familiarity, related information, and fluency. In section 4 I will evaluate the relevant empirical evidence and conclude that it is does not support core evaluativism. I will then re-examine the proposed heuristics and argue that, as conceived, they are not adequately distinguished from the content they are claimed to replace. One exception is fluency conceived as reaction time, but I will argue that this heuristic is flawed in a different way. In section 5 I will conclude by arguing that it is not necessary to posit a distinct mechanism for epistemic feelings, and that instead emphasis should be placed on the possibilities for nonconceptual metacognition suggested by epistemic feelings. I will then offer a phenomenological analysis of a tip-of-the-tongue experience which suggests that the kind of nonconceptual metacognition involved is best understood by the "feature space" account developed in Ch. 5.

6.2. Epistemic feelings, the Lima problem, and evaluativism

6.2.1. Epistemic feelings

My argument will not require a strict definition of epistemic feelings. Nor have I encountered an adequate one in the literature.[16] I have described a few epistemic feelings – feelings of confidence, feelings of knowing, and tip-of-the-tongue experiences. Other examples include feelings of familiarity, feelings of ease of learning, perceptual confidence, and many more (see de Sousa, 2009; Dellantonio & Pastore, 2019; Dokic, 2012 for longer lists).

Beyond a degree of overlap in the extension of the term, there is little agreement on the characteristic properties of epistemic feelings – even their epistemic nature and their status as feelings. Epistemic feelings are "epistemic" because they are involved in our cognitive processes in a way that admits of accuracy, but the manner of their involvement is a matter of dispute. Some theorists, including evaluativists, characterize many epistemic feelings as metacognitive, in that they take epistemic states as their intentional objects. A TOT or a feeling of knowing (FOK), for example, can tell us that we *know* the capital of Peru. Others take their epistemic role to be primarily first-order, aiding us in decision-making when purely rational deliberation is too time-consuming or costly (Carruthers, 2017; de Sousa, 2009). DeSousa (2009), for example, describes fear as an epistemic feeling that estimates objective degrees of risk in the environment. Carruthers (2017) denies even that FOKs are metacognitive.

Some epistemic feelings seem indisputably felt. TOTs, in their folk understanding, have a strong phenomenal component. But confidence and feelings of knowing (despite the name) are

[16] Arango-Munoz (2014a) offers this: "E-feelings are phenomenal experiences that point towards mental capacities, processes, and dispositions of the subject, such as knowledge, ignorance, or uncertainty" (p. 158). We will soon see that every part of this definition is controversial.

less obviously phenomenal. They are typically operationalized as predictions or evaluations of one's own performance on an experimental task, and as such are indistinguishable from judgments lacking any phenomenal component. Some characterize epistemic feelings as emotions (Carruthers, 2017; de Sousa, 2009), but it is not at all clear that a subject's confidence rating on the twentieth forced-choice trial reporting the orientation of a Gabor grating is an emotional experience. Despite this general climate of disagreement and ambiguity, evaluativists agree on a significant and controversial claim about epistemic feelings, which I call core evaluativism. This claim is motivated in part by the Lima problem, to which I turn in the next section.

6.2.2. The Lima Problem

Dokic (2012) poses the Lima problem for transparency and direct access accounts of metacognition. Transparency theory (Byrne, 2005) holds that knowing whether I believe *p* is simply a matter of determining whether *p* is the case and then applying an "ascent routine." That is, when asked whether I *believe* that the cat is on the mat, I "look to the world" and employ whatever ability I would use to determine whether there *is* a cat on the mat. I then follow a simple rule: If there is a cat on the mat, conclude that I believe it. But if self-attributing knowledge that *p* is simply a matter of such an ascent routine, how can we explain the Lima example? In the Lima case I cannot report the first-order proposition, that Lima is the capital of Peru, and thus it would seem I cannot apply an ascent routine to draw any conclusions regarding my beliefs about it.

But transparency theory is more flexible than Dokic allows. The "world" to which transparency theorists look must certainly include questions asked of the subject. The real issue is how to account for the general accuracy of the resulting beliefs. The ascent routine could look something like: 'If asked *x* and *y* occurs, believe that you know the answer to *x*', where *y* is an FOK. Of course, if one conceives of an FOK as a mental state then the ascent routine fails, and it might seem obvious that an FOK is a mental state. It's a *feeling*, after all. But such tricky cases are the transparency theorist's bread and butter. All that is needed is a first-order characterization of y that picks out the same phenomenon. We can find a precedent in Byrne's ascent routine for thought - "If the inner voice speaks about *x*, believe that you are thinking about *x*" (2008, p. 117). Byrne characterizes the inner voice qualitatively, as "degraded" relative to outer voices (2008, p. 118). Including an epistemic feeling in the ascent routine is no different in principle. One might attempt a qualitative characterization of the FOK but general agreement on those qualities seems unlikely.[17] And a qualitative characterization of FOKs wouldn't *explain* the accuracy of *y*. Carruthers (2017) offers a first-order characterization of FOKs based on partial recall, and this could both serve in an ascent routine and potentially offer a non-evaluativist account of the accuracy of the resulting belief. A similar solution is suggested in the course of Dokic's argument against direct access solutions to the Lima Problem, to which I now turn.

While transparency theory holds that knowledge of our mental states is obtained through knowledge about the world, other theories posit a direct informational channel to one's own mind. A classic example is inner sense. Inner sense accounts of metacognition hold that we become aware of our own first-order mental states through a quasi-perceptual mechanism within

[17] Dokic claims that epistemic feelings register "internal physiological conditions and events" (2012, p. 307). Perrin, Michaelian, and Sant'Anna (2020) offer a phenomenological description of the feeling of remembering as one of "pastness, self, causality, and singularity." In either case the properties picked out are arguably non-mental.

the mind (e.g., Locke, 1689; Lycan, 1996). Lycan identifies the inner sense with attention and applies it to both propositional attitudes and sensory states. Carruthers (2011) opposes inner-sense views of propositional attitudes but does hold that we directly perceive mental imagery (e.g., subvocalization), which facilitates knowledge of propositional attitudes. Pitt (2004) holds that there is a proprietary phenomenology of propositional thought that makes it directly introspectable. The Lima problem here is that, whatever the mode of direct access, it's not clear how a direct informational channel could tell me *that* something is in my mind without telling me *what* it is. If I believe that I know the capital of Peru on the basis of directly perceiving that I have precisely *that* information, then I ought to be able to produce it.

But these considerations are not fatal for direct access. Dokic himself considers a solution but rejects it. It might be the case that, in the Lima example, the subject is directly aware of only part of the first-order belief, say 'The capital of Peru is ___.' Dokic argues that this strategy is incompatible with a common view of introspection.

> … introspection makes the subject aware of her own intentional mental states only by revealing their contents (see, e.g. Tye 2009). In other words, introspection is *fully transparent* with respect to the contents of the introspected states (whenever they have contents). The Direct Access Model denies that introspection is always transparent in this sense, since feelings of knowing are precisely introspective states about particular first-order memories, while their contents are only partially revealed to the subject. (2012, p. 306)

The argument seems to be that, if the only input to inner awareness is first-order mental content, then the introspected content can't have gaps. But the consequent only follows if one

equates transparency with exhaustivity. Consider outer perception. The view that perception is transparent with respect to its object is compatible with the view that perception is not exhaustive. I can perceive part of an object, or only some of its properties, such as its shape but not its color in dim light. Likewise, a direct informational channel might reveal only partial information about the content of my mind, even if it is transparent with respect to what it does reveal. A similar response might be available to the transparency theorist as well: If asked *x* and you produce a partial answer to *x*, then you know the answer to *x*.

This general type of solution to the Lima problem exists in the psychological literature as well. According to Brown and MacNeil (1966), TOTs occur when we access semantic information, the meaning of the word, in the absence of phonological information. Here again we access some first-order information directly, and that is enough to let us know that the information is there even if it doesn't allow us to report it. This type of account can easily be extended to FOKs as well. Nelson, Gerler, and Narens (1984) offer further possibilities compatible with direct access. It may be that "associative strength," a posited relation developed between the cue and the target, is the source of epistemic feelings (1984, pp. 295–296). When the strength of the association is above a certain threshold, recall occurs. When the strength is lower, recall fails but there is a FOK. When the strength is even lower, there is neither recall nor feeing of knowing.[18] A third possibility is that access is "multidimensional," in that various informational properties of the memory trace are accessed, but not the word form itself.[19] [20] The

[18] Whether associative strength is a direct access account may depend on how one conceives the metaphysics of the relation.

[19] The traditional claim that memory consists in a "trace" of an original event stored in memory is contested by those who take memory to be an entirely reconstructive process (e.g., Michaelian, 2016). But the issue here is whether memory is the informational source of epistemic feelings, not whether that source is reconstructed or stored.

[20] Hart (1965) and Nelson and Narens (1990) are also cited by evaluativists as the sort of psychological direct-access accounts they oppose (Arango-Muñoz, 2019; Proust, 2013). But Hart's claims about a memory "monitor" are too brief and general to characterize as direct access, and while Nelson and Narens sometimes seem to favor direct access (1990, p. 150) at other times they sound like evaluativists (1990, p. 158).

Lima problem, then, does not motivate abandoning transparency theory or direct access for evaluativism.

Epistemic feelings pose a different problem for theory-theoretical accounts of metacognition. Theory-theory (Gopnik, 1993) is the claim that we come to know our own mental states much as we do the mental states of others – by observing our behavior and deriving metacognitive conclusions using a theory of mind. Even if I don't recall the capital of Peru, I might infer, based on my knowledge of myself, my history, and my abilities, that I do know it. But if, in the case of epistemic feelings, the vehicle of higher-order information is a feeling, this appears incompatible with its being an inference from a theory of mind (Dokic, 2012, p. 304). Much depends here on how one characterizes feelings, of course. If we mean emotions, and emotions are conceived as cognitive states, then there is no problem. And however we conceive these feelings, it is a commonplace that feelings can result from propositional knowledge. The knowledge that a loved one has died (or an inference to that effect) will produce a feeling of sadness, so it might be that an epistemic feeling is the result of (possibly unconscious) inferences from a theory of mind. Theory-theoretic explanations of the Lima problem are also threatened by the possibility that animals lacking a theory of mind nonetheless demonstrate metacognitive abilities facilitated by epistemic feelings (Proust, 2013, Chapter 5). But the evidence for animal metacognition is controversial (see Carruthers, 2008 for a skeptical view), as is the claim that animals lack a theory of mind (see, e.g., Emery et al., 2004).

While the Lima problem is offered as a threat to some philosophical theories of metacognition, it is not fatal. Perhaps for these reasons, when evaluativists make such arguments, they invariably bolster them with empirical evidence for a more general claim about epistemic feelings. This claim is that epistemic feelings are components of a distinct metacognitive

mechanism that takes heuristics as its inputs. If this claim could be supported, it would rule out transparency, direct access, and theory-theoretical explanations of the Lima problem and instead solve it in a novel way. In the next section I will describe this claim in detail.

6.2.3. Evaluativism

The term 'evaluativism' is used by Proust (2013) to characterize her view of the function of epistemic feelings ('noetic' feelings in her parlance). Epistemic feelings, for Proust, are metacognitive in that they serve to *evaluate* a subject's own cognitive processes in terms of success or failure. Proust claims that epistemic feelings are the output of a phylogenetically primitive form of metacognition, which she calls 'procedural' metacognition. Procedural metacognition is distinct from forms of metacognition that employ propositional attitudes, instead operating in a nonconceptual format. This distinctive form of metacognition does not receive direct information from memory, but instead takes as its inputs simple cues ("heuristics") that merely tend to correlate with cognitive success or failure. A FOK, for Proust, tells me that recall is possible not because it receives direct information that the item is in memory, but because it receives information that the recall process is operating "fluently." Fluency will be discussed in detail in section 3, but the basic notion is intuitive enough. A fluent process, be it speaking in a foreign language, swinging a bat, or retrieving an item from memory, is one that runs smoothly and without difficulty. This, for Proust, is what epistemic feelings report. And insofar as fluent mnemonic processes tend to be successful ones, epistemic feelings can be said to report on the potential for successful recall, and thus indirectly on the presence of the item in memory. Other evaluativists, we will see, favor other heuristics, but in each case the functional

picture is similar. Call this claim *core evaluativism* - epistemic feelings are metacognitive but their direct inputs are not the cognitive states that they are "about." Rather, they take as inputs cues and heuristics that tend to correlate with the presence or absence of such states.

If the inputs to epistemic feelings are distinct from the mental content on which they report, then the Lima problem is solved. We know *that* we know without being to report *what* we know because the process doesn't involve access to the content in question. And if the inputs aren't propositional, this opens up an informational role for feelings. This solution to the problem also sits well with an influential current in philosophy and psychology. Dual process theories of cognition (J. Evans & Stanovich, 2013) hold that many of the mental abilities once thought to be governed by rational, epistemically grounded processes are often accomplished by more phylogenetically primitive systems using unconscious cues and heuristics that merely produce adaptive behaviors. These more primitive cognitive abilities are grouped together as "System 1" processes in contradistinction to the more sophisticated, rational "System 2" processes. Various heuristic biases that tend to trump logical and probabilistic reasoning in first-order decision-making have been demonstrated in humans and are taken as examples of System 1 processes (see J. Evans, 2003 for a review). Evaluativists identify procedural metacognition as a System 1 process, and if correct this is a novel and interesting extension of dual process theory into the realm of self-knowledge.

Core evaluativism, if true, would seem to rule out transparency, direct access, and theory-theoretic solutions to the Lima problem. It preempts any access to first order content as the basis of epistemic feelings and the metacognitive abilities they facilitate. This means that partial direct access or partial redeployment of first-order content cannot be the basis of these abilities. It also rules out theory-theoretic explanations insofar as these are identified with System 2.

Proust is explicit about the functional claim:

> … the cues for cognitive success have nothing to do with the particular content of the words to learn, or with the intentional content of one's first-order thoughts. They are properties of processing, not properties of content. (Proust, 2013, p. 56)

Other researchers hold this view as well. Koriat developed much of the empirical support for evalutivism in his work from the 1970s to the present.

> Whereas information-based judgments entail deliberate, analytic inferences that rely on beliefs and memories, metacognitive feelings are mediated by the implicit application of nonanalytic heuristics. (Koriat, 2000, p. 128)

Similar expressions of the view can be found in Dokic (2012).

> The cues underlying noetic feelings are contingently but stably associated with epistemic states. This association holds in a normal (ecological) context, but it can be severed by psychologists, who can easily produce 'illusory' feelings of knowing … [epistemic] feelings have intentional contents beyond the body, but only in a derived way, through some kind of learning or association process. Such a process generates new heuristics, i.e. cognitive shortcuts that enable us to move spontaneously from our feelings to judgements concerning the task at hand. (Dokic, 2012, pp. 307–308)

Arango-Munoz also commits to core evaluativism.

> … for me (following the psychological tradition of metacognition research (e.g., Reder, 1987, 1996; Koriat 1993, 2000), the main function of low-level metacognition is to elicit E-feelings and control mental action based on cues and heuristics…In other words, the monitoring mechanism does not actually scan mental states. (Arango-Muñoz, 2014a, p. 150)

Evaluativists differ in their accounts of the representational format of epistemic feelings. Proust (2013) invokes a nonconceptual format borrowed from Strawson (1959) and Cussins (1992). Arango-Munoz (2014a) claims that, in addition to being nonconceptual, epistemic feelings are phenomenal states, whereas Proust claims they can be unconscious, which seems to preclude their being phenomenal. Koriat and Dokic are less explicit about the representational format, though Dokic does call them "experience-based" (2012, p. 304). But evaluativists converge on core evaluativism. The empirical basis of this claim comes largely from research on mnemonic (and some perceptual) epistemic feelings. Among these I have chosen FOKs, TOTs, and confidence, as they are most commonly cited, and I believe they offer the best case for evaluativism.[21] The most commonly cited heuristic inputs to these epistemic feelings are cue familiarity, recall of related information, and fluency. In the next section I will explain these heuristics.

6.3. Heuristics

[21] To take one example of the weaker evidence I will skip, Arango-Munoz (2019) cites work by Whittlesea and Williams (1998, 2001) in support of the claim that feelings of familiarity are causally sensitive to fluency. But Whittlesea and Williams operationalize feelings of familiarity as false alarms on a recognition task. False alarm rates are a first-order phenomenon, not a metacognitive measure.

Core evaulativism is the claim that epistemic feelings are not directly informed by the first-order mental content on which they report, but instead are sensitive to "cues" or "heuristics." What are these cues and heuristics? Among the most frequently cited heuristics are related information, cue familiarity, and fluency (Arango-Muñoz, 2019; Dokic, 2012; Koriat & Levy-Sadot, 2001; Proust, 2013).

Related information is typically invoked in the case of FOKs and TOTs. If the subject does not recall the word form sought in a memory task, but recalls some related details, this could be what triggers the epistemic feeling. So, while I might not recall the name of the capital of Peru, I might recall its location on the map, a dish I tasted on a recent visit, or its major exports, and on this basis feel that I know the answer. The cue familiarity heuristic takes the cue itself to be the source of epistemic feelings. If I were a subject in a memory study, I might be presented pairs of countries and their capitals: Canada: Ottawa, Ecuador: Quito, Peru: Lima. Later I might be asked to recall the capital based on the cue – the name of the country. Or I might be tested on general knowledge without prior training, and the cue might be "What is the capital of Peru?". In each case the claim is that a FOK would be based on whether the cue appears familiar to me, not on whether I have the answer in mind. What makes each of these factors heuristics is that they are thought to correlate with retrieval of the correct answer often enough that, for creatures like us, they provide a pretty good guide to whether the first-order content is in memory. They are fallible, of course, but a system employing these heuristics will likely be adaptive if direct access is time and energy consuming, or psychologically impossible.

The third heuristic is fluency. Fluency is the most difficult heuristic to characterize, as it is often defined and operationalized in different ways. Theorists have defined fluency as

"subjective ease" of processing (Duke et al., 2014), "speed and accuracy" of processing (Reber et al., 2002), or as a cluster of attributes comprised of "degree of activation," "speed," and "effort" (Winkielman et al., 2003). It is most frequently applied to confidence, feelings of familiarity, and judgments of learning (JOLs). JOLs are judgments that memorization has been successful, such that a given item can be recalled in the future. The idea behind the fluency heuristic is that when a stimulus is perceived by the subject as easier to see, easier to read, or in general easier to process, the subject predicts better performance on the relevant task. If a word pair like 'up-down' is processed more fluently than 'hammer-fish', subjects are more likely to report judgments of learning for the former (see Winkielman et al., 2003 for a review). "Regular" non-words like 'hension', which resemble words and are thus presumably processed more fluently than nonwords like 'jufict' are claimed to produce feelings of familiarity (Whittlesea & Williams, 1998). And subjects are more likely to report higher levels of confidence in their answers when the cognitive processes involved are more fluent (Finn & Tauber, 2015).

The phenomenon is tricky to operationalize, and as a result there are diverse ways of doing so. Some studies treat fluency itself as an epistemic feeling and measure it by subjective report. The epistemic feeling of fluency is then argued to be a causal factor in various judgments. For example, perceptual fluency appears to affect favorability ratings for various marketing materials (Graf et al., 2018). These studies are less relevant to evaluativism for our purposes, as they study the downstream effects of epistemic feelings, not their causal inputs. Some metacognitive studies operationalize fluency by manipulating the difficulty of tasks to be performed by subjects, for example the length of a list to be recalled (Schwarz et al., 1991), the regularity of a word to be memorized (Whittlesea & Williams, 1998), or in perceptual studies the clarity of the stimulus (Feustel et al., 1983; Kelley & Jacoby, 1998). In some studies fluency is

operationalized as reaction time (Benjamin et al., 1998), and often priming or other methods are used to manipulate reaction time (Whittlesea & Williams, 2001; Winkielman et al., 2003). Elsewhere priming alone is used to operationalize fluency, and it is not always clear whether this is meant as an implicit manipulation of reaction time (Jacoby, 1983). This is to say, in studies that invoke fluency it is often not clear what the heuristics are, what the epistemic feelings are, and whether the measures and manipulations employed are meant to be identified with heuristics, epistemic feelings, or causal antecedents or products of either.

To narrow the options, I will rely primarily on Proust's (2013) interpretation of fluency, since among evaluativists she places the most emphasis on it. Within her work fluency takes on many shades of meaning and I cannot do justice to that richness here. She variously describes fluency as an epistemic feeling (p. 58, 61, 102, 105), as an alternative epistemic norm to truth appropriate to nonconceptual content (p. 10, 125, 129-130, 137), as a heuristic on which other epistemic feelings are based (pp. 58-59, 62, 73, 105, 129), as the fundamental heuristic or epistemic feeling from which others develop in ontogeny or phylogeny (p. 73), as the output or operation of a mechanism known as an "adaptive accumulator" (p. 105, 129), and as a property of the neural assemblies that realize the adaptive accumulator (p. 130). All these various characterizations of fluency are not necessarily inconsistent, particularly if it is allowed that epistemic feelings can serve as heuristics for other epistemic feelings. While there is room for interpretation of her view, her argument focuses heavily on Vickers and Lee's (1998) adaptive accumulator model of decision making and on reaction time. These will be my focus in section 6.5.

6.4. The empirical argument for evaluativism

Evaluativists invoke a body of empirical research that purports to show that epistemic feelings are not causally sensitive to the target content but to heuristics. Three heuristics that evaluativists most frequently cite are cue familiarity, retrieval of related information, and fluency (e.g., Arango-Muñoz, 2019; Dokic, 2012; Koriat & Levy-Sadot, 2001; Proust, 2013). In this section I will consider the empirical evidence for each in turn and argue that in each case evaluativism is not supported. Other heuristics have been cited as potential inputs to epistemic feelings (e.g., study time as the heuristic for feelings of learning, Koriat & Ackerman, 2010), but space only permits discussion of the most prominent. In sections 4.1 and 4.2 I will critique the evidence for the cue familiarity and related information heuristics, respectively. In section 4.3. I will argue that evaluativism is ill-conceived, in that it fails to make a clear distinction between these first two heuristics and first-order content. In section 4.4. I will critique the fluency heuristic and argue that the relevant models are best interpreted as positing direct access to first-order content.

6.4.1. Cue familiarity

Reder and Ritter's (1992) study is cited ubiquitously by evaluativists in support of their view (e.g., Arango-Muñoz, 2013; Arango-Muñoz, 2019; Dokic, 2012; Koriat & Levy-Sadot, 2001; Proust, 2013). Although the study is rather old and, as we shall see, flawed, it plays an outsized role in evaluativist arguments. It is the only study cited by Dokic (2012) in support of the cue familiarity heuristic and is lauded by Koriat and Levy-Sadot as providing "remarkable support" for cue familiarity (2001, p. 35). Arango-Munoz (2013) also describes the study in detail as part of his argument.

Reder and Ritter (1992) trained subjects on a series of arithmetic problems that are difficult to work out mentally (e.g., 23×27). In experimental trials these same problems were presented, and subjects were offered a choice of answering strategy. They could choose to retrieve the answer from memory or to calculate the answer. If they chose retrieval, they were given less time (less than 1s) but more points for a correct answer produced within that time. If they chose to calculate, they were given more time (~15s) but fewer points for a correct answer within that time. The authors took a subject's choosing the retrieval option as an indication that they had experienced a FOK. FOKs did correlate with successful retrieval. The authors hypothesized, however, that the choice of the retrieval option was based on the familiarity of the question parts rather than the availability of the answer in memory. This was tested by examining trials in which question parts were the same as parts of previously seen problems, but in which the answers differed (e.g., 23×16, having previously seen 23×27 and 16×27). The authors found that choice of retrieval strategy showed a stronger correlation with previous exposure to the question *parts* than with correct responses. They concluded that cue familiarity, and not access to first-order content, drives FOKs.

While many researchers accept the results of this study uncritically (e.g., Hertzog et al., 2010; Hosey et al., 2009; Paynter et al., 2009; Walsh & Anderson, 2009), Koriat himself (1993) notes early on that many researchers commit the error of conflating what he calls "subjective" and "objective" properties of memory. A FOK, in the context of a typical memory study, is not understood merely as a feeling that the subject knows the contents of her memory. It is a feeling that she knows that the contents of her memory match the problem given in training. This is an artifact of how we often conceptualize memory. Memory, like perception and knowledge, is treated in such studies as a factive state. In order to remember I must not only be able to access a

mental representation, but that representation must also match the original stimulus. But this means that there are two ways a subject can fail on these tasks, and only one of them is a metacognitive failure.

Suppose that I am a subject in a similar study, and rather than being presented arithmetic problems and their answers I am presented with pairs of countries and their capitals. And suppose that, being a little mixed up after this barrage of information, I form the false memory (or belief if you like) that Quito is the capital of Peru. I may report a FOK but produce the wrong answer. In the studies in question this would be counted as a case in which FOKs do not correlate with metacognitive access to my first-order states. But this is not what happened in the proposed scenario. I did access my memory and my response was based on that content—the content just happened to be false. Furthermore, my FOK will correlate with the cue ('Peru'), but not for any deep reason. It will be an artifact of my having a false belief involving that cue stored in memory.

A similar analysis holds for the Reder and Ritter study. A subject who chooses the retrieval option, but answers incorrectly, may well have an answer in mind, it's simply incorrect. Many of us have experienced misleading TOTs. When we finally retrieve the name that was on the tip of the tongue, we find out that it was incorrect. And I may choose the retrieval option more frequently when I have already viewed parts of the answer because I incorrectly encoded problems I had already seen.

Of course, the analogy between arithmetic problems and the Peru question limps, and so the evaluativist analysis of this study, as well as my counter-analysis, may feel awkward. Complex math problems are not the sort of thing we typically encode in memory for any length of time, and thus it is difficult to determine which explanation of the data is more plausible. But

even if one dismisses my analysis, the most that can be concluded is that in situations in which we are required to memorize things we don't typically tend to memorize, like difficult arithmetic problems, we might rely on cue familiarity. This falls far short of full-throated evaluativism.

Other studies of the cue familiarity heuristic suffer from similar problems. Metcalfe, Schwartz, and Joaquim (1993) use a paradigm from interference theory in which the word pairs to be memorized are provided alongside other word pairs designed to interfere with encoding and reduce memorability (e.g., the cue paired with a synonym of the target, the cue paired with a different word, or a word pair containing neither the cue or the target). Metcalfe et al. predict that if cue familiarity is the source of FOKs, then FOKs should occur in proportion to the frequency of appearance of the cue, not to memorability. The results of their study, even on its own terms, are rather ambiguous. In recognition tasks, subjects fail to perform in the way the interference paradigm traditionally predicts, and as a result, FOKs end up tracking recognition performance nearly as well as cue frequency. But the fundamental problem with the study is the same as in Reder and Ritter – the researchers identify direct access to memory with accurate performance on the recognition task. As we have seen, these are different things.

Other measures of direct access have been employed, but these are also insufficient to distinguish direct access from accuracy. There is room only to survey a few to demonstrate the variety of measures employed. Liu et al. (2007) manipulate target retrievability by repeating instructions to remember the target. Metcalfe and Finn (2008) manipulate target retrievability by providing multiple cues for a single target (high retrievability) or a single cue for multiple targets (low retrievability). While each of these manipulations might increase the likelihood of accurate recall, in each case the low retrievability option is consistent with incorrect encoding and direct

access to that incorrectly encoded content. Indeed, Metcalfe and Finn's low retrievability condition plausibly encourages incorrect encoding by pairing the cue with multiple targets.

Hanczakowski et al. (2013) recognize some of these difficulties and rather than attempt to measure or manipulate memorability simply measure the effect of priming the cues (a manipulation of cue familiarity) on FOKs regardless of objective performance. They show a significant effect of priming on FOK magnitude. This has the advantage of not identifying direct access with accuracy, but without *some* independent measure of direct access we can't compare the *relative* effects of the two variables, nor can we rule out the possibility that the manipulation of cue familiarity increased FOKs by increasing the accessibility of the target (or of content mistaken for the target).

There are many more studies involving cue familiarity than I can cover here. More recent studies often take evaluativism as given and instead focus on the relative contribution of various forms of cue familiarity and related information to FOKs, largely ignoring direct access as a contender.[22] Cue familiarity and partial information have been posited as responsible for different stages (Koriat & Levy-Sadot, 2001), different types (Liu et al., 2007), and different aspects (Isingrini et al., 2016) of FOKs. I will address some of these studies in the next section and argue that they do not offer methodological improvements.

6.4.2. Related information

[22] To take one example, Hertzog et al. examine varieties of the partial information heuristic, claiming that "All extant theories of FOKs reject the idea that individuals have direct access to information held in memory (2010, p. 772)."

Koriat (1993) is aware of the potential for ambiguity in the study of epistemic feelings. He notes a distinction between the objective and subjective factors involved in memory and agrees that ignoring the latter can cause researchers to draw hasty conclusions about the nature of epistemic feelings. He attempts an improved methodology and employs it in studies that purport to show that the related information heuristic is the causal factor to which FOKs are sensitive. I will describe and critique Koriat's seminal (1993) study as well as some more recent ones.

Koriat (1993) presents subjects with nonsensical four-letter stimuli (e.g., JKSD). Subjects are then asked to recall the letters, and the number of letters they recall in each case is recorded, along with reports of FOKs. It was found that the number of letters subjects were able to recall, whether correct or incorrect, correlated with FOKs. Although he is explicitly sensitive to the worries I describe in the last section, Koriat's study is not an improvement. The correlation he finds is consistent with the possibility that subjects sometimes incorrectly encode the letters to be recalled, report a FOK on the basis of the stored (incorrect) letters, and then report those letters.

Later studies do not improve matters. Koriat and Levy-Sadot (2001) manipulate the amount of related information by asking questions about categories with many recallable members (Who *composed* Swan Lake?) versus categories with fewer such members (Who *choreographed* Swan Lake?). The researchers suggest that larger categories should make direct access of the target less likely, since there are more competing items in memory. If so, a direct access account would predict fewer FOKs in such conditions (2001, p. 38), whereas the partial information heuristic would predict more FOKs, given the larger number of recallable items (e.g., composers). But of course, the reason that people know fewer choreographers is that they tend to be less famous, so it's equally plausible that a direct access account would predict more

FOKs in those conditions. Subjects are likely to know *some* composer, and even if it's not the correct one it may trigger an FOK.

More recent studies have expanded the partial information heuristic, positing an influence on epistemic feelings from a wide variety of factors, including emotional content of items to be remembered (Schwartz, 2010) or other contextual aspects of the encoding process (Hertzog et al., 2014), but none improve measures of direct access. Indeed, some of these studies are explicitly pluralistic, allowing for a degree of direct access (Schwartz, 2010). Others ignore direct access as a possibility altogether (Hertzog et al., 2014).

6.4.3 Heuristics or direct access?

But even if a better measure of direct access could be devised, there is a fundamental ambiguity in the related information heuristic. Koriat makes the point himself, noting that "the cues for the FOK are to be found in the very information that is activated or accessed during the course of the search-and-retrieval process (1993, p. 611).[23] When the information happens to be correct, we have normal recall. When the information is incorrect, it can still generate FOKs. But in either case the functional architecture is the same. The heuristic that evaluativists claim operates in lieu of access to the relevant first-order content turns out to be access to *fragments* of the relevant first-order content. Recall Koriat's (1993) study, in which the related information is simply a group of letters. If I am attempting to retrieve the capital of Peru from memory and retrieve 'Lim_', is this a heuristic or just three-fourths of the answer? I think we should say the latter, as this differs not at all from the inner-sense account that Dokic rejects. And if instead I retrieve

[23] Schwartz and Metcalfe (2011) also point out that the partial information heuristic is "compatible" with direct access.

"Qui_" and have a FOK, I have argued, and Koriat appears to agree, that to take this as evidence for evaluativism is to conflate the objective and subjective factors of memory.

To be clear, evaluativists do not deny that we access our memories. But they do claim that epistemic feelings are not informed by the target content. Whether the related information heuristic counts as a distinct source of information, then, will depend on one's theory of mental content. Consider the Lima example. Like Reder and Ritter's arithmetic problems, nonsense strings of letters are not entirely analogous to everyday cases of epistemic feelings. In typical cases of TOTs and FOKs, the partial information being recalled is not only letters in the name. I may not be able to recall the name of the capital of Peru, but I might be able to picture its location on a map, name some of its neighborhoods, and recall an NPR segment on its cuisine. If we compare these bits of recalled information with Brown and MacNeil's (1966) direct access account, on which subjects access an amodal "pure" meaning, there might appear to be a principled distinction between heuristic information and the target content. Perhaps the content proper is the meaning and everything else is heuristic information. This apparent foothold for evaluativists, however, becomes much more precarious under scrutiny, for related information may be *part* of the meaning. While Brown and MacNeil seem to imply a distinction between information in a phonological format and a "pure," amodal meaning, the actual studies evaluativists cite make no attempt to determine the format of the related information, nor do any evaluativists I know of define related information in terms of such format.[24] One could do so, but different empirical studies would be required to support that account.

[24] Carruthers (2011) makes such a distinction and claims that access to one's own propositional attitudes is indirect, but he is no evaluativist. He denies that epistemic feelings have metacognitive content and suggests a direct causal relationship between memory and action for feelings of knowing (2017).

And there are reasons evaluativists should want to avoid committing to such specifics. To pursue this sort of gambit the evaluativist must draw a clean line between the meaning of 'Lima' and even related *amodal* information. This amounts to an a priori rejection of any holistic account of mental content. The target content – the meaning of 'Lima' – must be considered distinct from content like 'is the capital of Peru'. Thus classical, definitional views of concepts must be rejected, inferential role accounts, theory-theoretical accounts, activation in a semantic network, and so on. Evaluativism becomes a branch of semantic atomism. It would no longer be a general account of the functional architecture of epistemic feelings. This, I take it, is not evaluativists' aim.

And this analysis can be applied to the cue familiarity heuristic. Cue familiarity is plausibly interpreted as a special case of related information since the cue is a bit of related information available to the subject. It is part of the definition, semantic network, or whatever one takes the content of 'Lima' to be, on a holistic view of content, that it is the capital of Peru. The Reder and Ritter study takes this to an extreme. The math problems presented are not integrated into a semantic network in the same way as knowledge of countries and their capitals, or even the basic multiplication tables. In this case the numbers used as the cues are the *only* relevant associated information available, at least for the non-mathematician. Other studies often use similarly artificial pairings of cues and answers (e.g., Liu et al., 2007). The cue familiarity hypothesis, then, is consistent with the claim that related information is the source of epistemic feelings, it's just that sometimes there is relatively little related information available—just the cue. Recent research points in this direction. Thomas et al. (2012) found that when participants were asked to focus on semantic properties of a cue during encoding, as opposed to "shallow" properties like the color of the type, FOK judgments were more accurate. Koriat and Bjork

(2005) find that JOLs are inflated when there is a strong but unexpected semantic association between cue and answer (e.g., 'find-seek' as opposed to 'find-lose').

If cue familiarity is just related information, and related information is just first-order content, then the fact that these so-called "heuristics" give us information about the content of our minds does not support evaluativism. Unless one adopts a very specific semantic theory, evaluativism collapses into a direct access account. There are influential atomic accounts of content (Dretske, 1981; Fodor, 1998; Millikan, 1987), but the heyday of such views appears to be in the past, and to yoke evaluativism to them is to sacrifice its status as a general account of epistemic feelings. It seems, then, that core evaluativism is not only unsupported by the empirical evidence, there is a fundamental problem in its very conception.

6.4.4. Fluency

While Dokic and Koriat favor cue familiarity and related information as the source of epistemic feelings, Proust has a different account. She holds that epistemic feelings reflect varying degrees of "fluency." In section 3 I noted the variety of interpretations of fluency in the empirical literature. Proust does an impressive job attempting to interpret and unify this unruly concept, but as I also noted in section 3, questions remain for her account. Here I will consider two of the most promising interpretations of Proust's fluency heuristic. The first conceives fluency as the operation or output of an adaptive accumulator (Vickers & Lee, 1998). I will argue that the adaptive accumulator model does not in fact support evaluativism, nor do more recent models of the same phenomena. Elsewhere Proust seems to describe fluency entirely in terms of a monitoring of reaction time (2013, pp. 129, 136). Therefore, it is also worth considering reaction

time itself as the relevant heuristic. I will argue that the empirical evidence does not support evaluativism on either interpretation of the fluency heuristic.

6.4.4.1 Fluency as operation of an adaptive accumulator

Proust clearly considers fluency, as the heuristic for feelings of confidence, to be intimately related to the operation of an "adaptive accumulator" as proposed by Vickers and Lee (1998). In this section I will consider whether this model supports an evaluativist account of metacognitive confidence.

Vickers and Lee's (1998) double accumulator model is grounded in signal detection theory (SDT) - one of a class of models that enriches the traditional SDT framework to account for dynamic aspects of reaction time and the metacognitive phenomenon of confidence. Space does not permit a detailed description of either traditional SDT models or their dynamic successors, but the fundamental point of relevance is that such models posit a noisy internal signal that serves as evidence about the state of the world and is the basis of our judgments about the world. Because the signal is noisy, there is no perfect correlation between first-order judgments and the world itself. Judgments are based on a variable criterion. If the signal strength, whether due to noise or the relevant state of the world, reaches that criterion a judgment is made. Vickers and Lee's adaptive accumulator also models the collection of evidence over time and confidence in first-order judgments. Their model of confidence is a "balance of evidence" account, on which the evidence for each of two options (SDT models are typically models of binary choice) is compared and confidence is based on the relative strength of the evidence in favor of the chosen option. Note that on this model confidence appears to be responsive to the

same evidence as the first-order judgment - a different operation is simply performed on that evidence. Vickers and Lee do not explicitly interpret their model in term of mental states and their intentional objects, but the most obvious interpretation appears to cut against evaluativism. Evaluativism claims that confidence and other epistemic feelings take as their inputs dynamic or other properties of the first-order cognitive *process*. But the model takes these inputs to be the same as inputs to the first-order process. Proust largely avoids this issue by concentrating on a further function of the model – the calibration of confidence ratings based on past accuracy (e.g., 2013, p. 99). But if confidence itself is a metacognitive epistemic feeling, then this calibration function of the accumulator is a *third*-order process, distinct from any heuristic input to the feeling of confidence itself. If Proust identifies fluency with the operation of Vickers and Lee's adaptive accumulator, the obvious interpretation of that model cuts against evaluativism.

More recent models appear to do a better job than Vickers and Lee's of predicting various phenomena related to confidence, but they also seem to defy an evaluativist interpretation. Instead of a balance of evidence model of confidence, Pleskac and colleagues (Pleskac & Busemeyer, 2010; Yu et al., 2015) posit that evidence continues to collect after the first-order judgment is made. Confidence will reflect the amount of evidence accumulated in favor of the first-order judgment in the interval between that judgment and the confidence rating. This improved model of confidence is no more amenable to evaluativist interpretation, as again confidence ratings are posited as responsive to the same evidence as first-order performance. Ratcliff and Starns offer a model of confidence on which "decision processes transform the strength of the match between a test item and memory to a confidence judgment (2013, p. 5)." The model explains confidence as the result of direct access to items in memory ("matching" of the memory and the cue). Grimaldi et al. model confidence judgments in Bayesian terms as the

width of the posterior probability distribution of evidence for a given decision (2015, p. 13). In each case first-order and metacognitive responses take the same evidence as inputs, they simply assess different properties or time-slices of that evidence. None of these models posit properties of the first-order process itself as the source of metacognitive responses.

If fluency is the operation of an adaptive accumulator, core evaluativism is ill-conceived, as the heuristic is not distinguished, even in principle, from direct access to first-order content.

6.4.4.2 Fluency as reaction time

But there may be another reading of Proust's account of fluency. She sometimes describes fluency as reaction time:

> Comparative fluency is the property, for a stimulus, of being processed more or less quickly and adequately, with respect to what is expected, in a kind of task (or in a control loop). (2013, p. 127)

> A temporal lag presents the subject with an error-signal: as compared with a normal behaviour, present activity is impaired. Here is the essential point: although the delay is a natural consequence of task difficulty, it becomes in addition a natural signal carrying information about a need to know what the affordance is. A plausible hypothesis therefore is that a temporal comparison between expected time for completion of the task and observed time, occurring as part of a given controlled activity (i.e. including a comparator), offers a key to making an affordance salient to the animal. (2013, p. 136)

Reaction time (RT) has been posited as a heuristic behind several epistemic feelings, including JOLs (Koriat & Ma'ayan, 2005) and confidence (Proust, 2013; Ratcliff, 1978; Volkman, 1934). It is less obvious how reaction time could inform TOTs and FOKs, since in these cases an answer is not yet retrieved. Unlike the other heuristics we have considered, however, reaction time does not appear to be subject to the criticism that it reduces to a direct access view. That is, if one's feeling of confidence in an answer is based on how quickly one retrieved it, this would be distinct from the first-order content. Instead it is a property of the cognitive process itself, and this sounds like evaluativism proper. The question is whether it's true that epistemic feelings are caused by reaction time. Here I will focus on confidence, since this is the subject of Proust's discussion.

Simple RT-based models of confidence exist which resemble Proust's description (Ratcliff, 1978; Volkman, 1934). On these accounts a longer RT means the process is not going well and should induce low confidence. The negative correlation does hold when there are no time limits on the response. But it has long been known that when responses are demanded quickly, this correlation no longer holds (Irwin et al., 1956). It seems unlikely, then, that RT is the only input to feelings of confidence. More recent models account for this by basing confidence ratings on the quality of the evidence, represented by "drift rate" – the speed with which the evidence approaches the decision criterion. When time is not limited, a slow response time is the result of a slow drift rate toward the first-order decision criterion. A model like Pleskac and Busemeyer's (2010) predicts low confidence in this case, as evidence will continue to collect slowly after the first-order decision. But when response time is set by the researcher, it

is no longer a function of drift rate and the negative correlation with confidence no longer holds.[25]

Other researchers claim that reaction time does have some causal influence on confidence, as studies suggest that manipulation of reaction time affects confidence when other variables are held constant (Kiani et al., 2014). These researchers conclude that reaction time may be a relevant factor *in addition to* first-order evidence in the production of feelings of confidence. But this is not support for core evaluativism, which denies direct access to first order content altogether.[26]

6.5. Epistemic Feelings and Feature Spaces

Evaluativists posit a distinct, primitive metacognitive process that produces epistemic feelings based on heuristic inputs rather than first-order content. I have argued that this is unmotivated, as there are viable transparency, direct-access, and theory-theoretical explanations of phenomena like the Lima problem. I then reviewed the empirical evidence that epistemic feelings take heuristics as their inputs. I concluded that the measures used in these studies do not distinguish heuristics from first-order content. I then argued that, as conceived, the heuristics *just are* first-order content. I conclude that the evaluativist account of epistemic feelings is unmotivated, unsupported, and ill-conceived. My argument thus far, then, has been a negative one. In what follows I will briefly discuss one aspect of Proust and Arango-Munoz's versions of evalautivism that remains viable, significant, and could form the basis of further research.

[25] The model does not posit that subjects access drift rate directly, which would implicitly involve a measure of RT. It posits different response criteria for different confidence levels. Higher quality evidence will tend to hit a higher confidence criterion in a given time interval.
[26] I thank an anonymous reviewer for pointing this out.

Part of the fascination of epistemic feelings, which gets lost when we focus on their functional architecture, is the notion that metacognition can occur in a non-propositional format. It seems to refute the familiar "thought about thought" definition of the phenomenon. But there is an implicit assumption in evaluativism that because nonconceptual content is understood as phylogenetically primitive it must also be informationally impoverished. Proust conceives the content as expressing something like "poor (excellent) *A*-ing affordance" where *A* is some cognitive ability (2013, p. 121). This conception of the content of epistemic feelings might be necessary if we insist that they operate in a system that can only process simple heuristics. But nonconceptual content, as standardly conceived, carries *more* information than conceptual content (e.g., Dretske, 1981). Cussins' (1992) account of nonconceptual content, on which Proust's is largely based, posits a nested series of behavioral dispositions that allows for a great deal of informational complexity and versatility (see Grush, 2000 for a similar account of spatial content). And crucially, Cussins holds that nonconceptual content exists on a continuum with conceptual content and that the relationship between the two is fluid, concepts frequently becoming unstable, devolving back into a nonconceptual format and refashioning into new concepts. If we abandon the dual-mechanism account, epistemic feelings might be conceived as one manifestation of a kind of nonconceptual metacognition that is deeply intertwined with and supports propositional metacognition. There may be other such manifestations. I propose that reality testing is one.

Arango-Munoz explicitly considers the possibility that conceptual metacognition is grounded in nonconceptual metacognition and that the two coexist in a single mechanism (2011, p. 78). He dismisses this account partially on the basis of the functional claims I dispute, adding the point that a one-mechanism account predicts "parallelism between judgments concerning the

self and others" (2011, p. 78). The argument is telegraphic, but the latter point appears to rely on the assumption that the one-mechanism account would be a form of simulation theory (Goldman, 2006). The Cussins-style account I have sketched is not, and it may help Arango-Munoz with another problem. Evidence that epistemic feelings are susceptible to conceptual priming drives him to claim that some epistemic feelings are conceptual (2014b), sacrificing that distinction between epistemic feelings and other forms of metacognition. If the conceptual and nonconceptual are continuous and part of the same system, then one needn't posit two kinds of epistemic feelings to explain the data. That epistemic feelings are a form of nonconceptual metacognition constitutes a significant claim, even if there is no evidence that they reveal the existence of a distinct mechanism. Evidence for nonconceptual metacognition could come from studies of animal cognition, but as I have briefly noted, the methods are controversial, and more work must be done. The situation is trickier in human subjects if we assume that conceptual and nonconceptual metacognition often intermingle. I have argued that the measures employed in studies of epistemic feelings are flawed. A different theoretical perspective may improve our methods.

Proust describes epistemic feelings as feature-placing content, borrowing from Cussins and Strawson, but employs feature-placing content in a different way than I have. She describes the content in question as indicating the presence of a remembering affordance as an evaluation of that affordance along a single "good/bad" dimension. In Grush's terminology, the information to which epistemic feelings respond on the evaluativist account would be that provided by a single manifold, as it varies along one dimension, be that reaction time, cue familiarity, or volume of associated information. One can, to some degree, define 'affordance' however one likes, and it is unsurprising that whatever explanatory variables evaluativists invoke will be

limited, since it is essential to the nature of the empirical work cited to limit the number of relevant variables. But remember that epistemic feelings are conscious, phenomenal states, and as such we must rely on the phenomenological reports of experimental subjects. There is a real danger of distorting those reports by imposing limits on what can be reported. A one-dimensional survey will elicit a one-dimensional response. Therefore, it is worth considering a more complete, unconstrained phenomenal description of an epistemic feeling in order to determine whether such distortion is taking place. I will consider a tip-of-the-tongue experience, since these are the most undeniably conscious and phenomenal.

Most of us have had the experience of a song being stuck in one's head. The tune will appear in imagination unexpectedly and unbidden. Sometimes one cannot identify the song that appears, and a tip-of-the-tongue experience will ensue. I will recount one actual episode of this type. While walking my dogs, two bars of descending vocal "aahs" popped into my head. A tip-of-the-tongue feeling ensued shortly after as I struggled to identify the song the bars were from. One thing to note off the bat, however, is that it was not clear precisely what information I was seeking. I was seeking to identify the song, but there are a number of ways to do that. It was only after the tip-of-the-tongue experience was "satisfied," so to speak, and ceased that I knew what I was after. That is, at the moment that the feeling commenced, it might have been the name of the song, or the name of the artist, or the rest of the music, or the lyrics, or some portion of the music or lyrics that would have satisfied the feeling.

Soon after I experienced the two bars, other content became accessible. Insofar as it did not cause the tip-of-the-tongue experience to cease, however, it was not the content I was after. I was certain that the song was first released in the 1970s and that the production was clever, but not offensively slick. None of this content was contained in the bars themselves. The notes and

some elements of the production were presented, but they were not sufficient for me to determine the "slickness." This content, I felt, was distinct. None of this information satisfied the epistemic feeling. After several deliberate repetitions of the two bars in imagination, the song continued to the guitar solo. But the feeling did not cease. Several more repetitions of the two bars and the solo were required, at which point things began to happen rather quickly. Within the span of 1-2 seconds, I felt that I knew the artist, though the name did not appear. Shortly thereafter I recalled the name of the artist. Somewhere in this 1-2 seconds the timbre of his voice and some additional lyrics appeared, though I cannot say in what order. The tip-of-the-tongue feeling appeared to me to be satisfied after grasping the artist, but before I could recall his name (Paul Simon, "Run That Body Down").

I have described the tip-of-the-tongue experience in as exhaustive detail as I could recall so as to compare it to the phenomenon as studied in the lab. The methodology of the studies cited by evaluativists cannot capture this sort of richness of detail. The satisfier of the feeling is decided by the researchers before either they or the subject are in a position to know it. Nor does the methodology capture the gradual discovery of more related content as steps toward discovering the satisfier of the feeling. Nor are they designed to allow for the possibility that the satisfier could be semantic content without lexical content, as it seemed to be in my case. There is good reason to design studies in such ways. Variables need to be controlled so that we can isolate a variable of interest. My dispute is with the conclusions evaluativist philosophers draw from such studies—a picture of metacognitive content that is limited to these few, isolated variables.

This richer description of the tip-of-the-tongue feeling suggests a groping through a space of associated content, and this space, I propose, is a feature space. The feeling is something like

waking up at a random location and attempting to find one's way home. Though disorienting, one can find one's bearings by locating nearby landmarks. The feature space of a song is different than the feature spaces that define perception and imagination or episodic memory. It is more like the feature space that defines a mug. But it is not obviously associated with motor contingencies. The contingencies appear to be almost entirely "sensoricognitive." Having heard the aahs, I predict the guitar solo. Interestingly, I also predict certain features of the production, and a feeling of the decade in which it was produced. Much of this, as initially experienced, is nonconceptual. The music itself is not conceptual, as I lack the proper musical concepts to describe the melody. And as experienced, even the content for which I did have concepts arrived before the words could be found. Perhaps this was semantic, conceptual content without the lexical component. But it might also have been described as nonconceptual – the two are indistinguishable phenomenologically in this case. As the cognitive trails of this feature space are explored, at some point there is a sudden shift of perspective as one locates oneself on the map. All of the associated content comes in one rush as the experience snaps into place as part of a complete, chunked concept. This, I propose, matches better the experience of a tip-of-the tongue experience than Proust's one-dimensional characterization of the feeling as the detection of a remembering affordance.

7. Conclusion

I have argued that reality testing is a form of nonconceptual metacognition. Reality testing is metacognitive because it is a mental state or process that takes another mental state or process as its intentional object. Reality testing also qualifies as metacognitive according to more specific

criteria offered by Nelson and Narens and Schwitzgebel. Reality testing is nonconceptual because creatures that lack mental state concepts are able to perform reality testing. These creatures include rats, mice, corvids, and human infants. I have also argued that reality testing is a necessary condition for the possession of concepts. I have not offered a demonstrative argument for this claim, but I have surveyed a representative sample of accounts of conceptual content possession and have argued that each either implicitly or explicitly requires reality testing. I then argued that no extant accounts of metacognition can account for nonconceptual reality testing. I offered my own account of reality testing as a sensorimotor skill. Finally, I argue that the only other account of nonconceptual metacognition, Joelle Proust's evaluativism, is unmotivated, unsupported, and ill-conceived. I then provided provisional evidence that the phenomena evaluativism is meant to explain can be explained by my own sensorimotor account of nonconceptual metacognition.

I believe that my account of reality testing could be developed to explain other phenomena. We do not only distinguish between perception and imagination or episodic memory. We also distinguish among other attitudes that we take to a given content. This ability is likely nonconceptual as well. Determining that one is entertaining a belief on the basis of one's other beliefs suggests an infinite regress, whereas an explanation of that ability in terms of nonconceptual content would not. Though an explicit argument remains to be made, it is my contention that nonconceptual metacognition is at the core of many of the everyday mental capacities that we commonly call 'cognition'.

Chapter 6, in part, contains material that has been published. The dissertation author is the sole author of this paper.

References

Aizawa, K., & Gillett, C. (2009). The (Multiple) Realization of Psychological and other Properties in the Sciences. *Mind & Language*, *24*(2), 181–208.

Allen, C. (1999). Animal Concepts Revisited: The Use of Self-Monitoring as an Empirical Approach. *Erkenntnis*, *51*(1), 33–40.

Allen, M. (2019). *Mental imagery: Eyes open and shut*.

Arango-Muñoz, S. (2011). Two Levels of Metacognition. *Philosophia*, *39*(1), 71–82. https://doi.org/10.1007/s11406-010-9279-0

Arango-Muñoz, S. (2013). Scaffolded Memory and Metacognitive Feelings. *Review of Philosophy and Psychology*, *4*(1), 135–152. https://doi.org/10.1007/s13164-012-0124-1

Arango-Muñoz, S. (2014a). Metacognitive Feelings, Self-Ascriptions, and Mental Actions. *Philosophical Inquiries*, *2*(1), 154–162. https://doi.org/10.1007/s11406-010-9279-0

Arango-Muñoz, S. (2014b). The nature of epistemic feelings. *Philosophical Psychology*, *27*(2), 193–211. https://doi.org/10.1080/09515089.2012.732002

Arango-Muñoz, S. (2019). Cognitive phenomenology and metacognitive feelings. *Mind & Language*, *34*(2), 247–262. https://doi.org/10.1111/mila.12215

Arlow, J. A. (1969). Fantasy, Memory, and Reality Testing. *The Psychoanalytic Quarterly*, *38*(1), 28–51. https://doi.org/10.1080/21674086.1969.11926480

Babb, S., & Crystal, J. D. (2006). Episodic-like memory in the rat. *Current Biology*, *16*, 1317–1321.

Barr, R., Marrott, H., & Rovee-Collier, C. (2003). The role of sensory preconditioning in memory retrieval by preverbal infants. *Animal Learning & Behavior*, *31*(2), 111–123. https://doi.org/10.3758/BF03195974

Bartal, I. B.-A., Decety, J., & Mason, P. (2011). Empathy and Pro-Social Behavior in Rats. *Science*, *334*(6061), 1427–1430. https://doi.org/10.1126/science.1210789

Beck, J. (2012). The Generality Constraint and the Structure of Thought. *Mind*, *121*(483), 563–600. https://doi.org/10.1093/mind/fzs077

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The Mismeasure of Memory: When Retrieval Fluency Is Misleading as a Metamnemonic Index. *Journal of Experimenal Psychology: General*, *127*(1), 55–68.

Bentall, R. P. (1990). The illusion of reality: A review and integration of psychological research on hallucinations. *Psychological Bulletin*, *107*(1), 82–95.

Bermúdez, J. L. (1998). *The paradox of self-consciousness*. MIT Press.

Bermúdez, J. L. (2007). *Thinking Without Words*. Oxford University Press.

Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception & Psychophysics*, *64*(5), 844–854. https://doi.org/10.3758/BF03194750

Bick, A., & Kinsbourne, M. (1987). Auditory hallucinations and subvocal speech in schizophrenic patients. *American Journal of Psychiatry*, *144*(2), 222–225.

Blaisdell, A. P. (2019). Mental imagery in animals: Learning, memory, and decision-making in the face of missing information. *Learning & Behavior*, *47*(3), 193–216. https://doi.org/10.3758/s13420-019-00386-5

Boller, K. (1997). Preexposure effects on infant learning and memory. *Developmental Psychobiology*, *31*(2), 93–105. https://doi.org/10.1002/(SICI)1098-2302(199709)31:2<93::AID-DEV2>3.0.CO;2-O

Borst, G., & Kosslyn, S. M. (2008). Visual mental imagery and visual perception: Structural equivalence revealed by scanning processes. *Memory & Cognition*, *36*(4), 849–862. https://doi.org/10.3758/MC.36.4.849

Borst, G., Kosslyn, S. M., & Denis, M. (2006). Different cognitive processes in two image-scanning paradigms. *Memory & Cognition*, *34*(3), 475–490. https://doi.org/10.3758/BF03193572

Brandt, S. A., & Stark, L. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*(1), 27–38.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Brown, A. S. (1991). A review of the tip of the tongue experience. *Psychological Bulletin*, *109*(2), 204–223. https://doi.org/10.1016/B978-012370509-9.00142-X

Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 325–337. https://doi.org/10.1016/S0022-5371(66)80040-3

Buettner, A. (2001). Observation of the Swallowing Process by Application of Videofluoroscopy and Real-time Magnetic Resonance Imaging—Consequences for Retronasal Aroma Stimulation. *Chemical Senses*, *26*(9), 1211–1219. https://doi.org/10.1093/chemse/26.9.1211

Bugnyar, T., Reber, S. A., & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications*, *7*(1), 10506. https://doi.org/10.1038/ncomms10506

Burdach, K. J., & Doty, R. L. (1987). The Effects of Mouth Movements, Swallowing, and Spitting on Retronasal Odor Perception. *Psychology and Behavior*, *41*, 353–356.

Burge, T. (2010). *Origins of objectivity*. Oxford University Press.

Bushnell, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, *7*, 3–15.

Bussey, T. J., Clea Warburton, E., Aggleton, J. P., & Muir, J. L. (1998). Fornix Lesions Can Facilitate Acquisition of the Transverse Patterning Task: A Challenge for "Configural" Theories of Hippocampal Function. *The Journal of Neuroscience*, *18*(4), 1622–1631. https://doi.org/10.1523/JNEUROSCI.18-04-01622.1998

Butler, S., Caron, A., & Brooks, R. (2000). Infant Understanding of the Referential Nature of Looking. *Journal of Cognition and Development*, *1*(4), 359–377.

Butterworth, G., & Cochcran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioral Development*, *3*, 253–272.

Byrne, A. (2005). Introspection. *Philosophical Topics*, *33*(1), 79–104.

Byrne, A. (2008). Knowing That I Am Thinking. In A. Hatzimoysis (Ed.), *Self-Knowledge* (pp. 105–124). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199590728.003.0006

Byrne, A. (2018). *Transparency and self-knowledge*. Oxford University Press.

Campanella, J., & Rovee-Collier, C. (2005). Latent Learning and Deferred Imitation at 3 Months. *Infancy*, *7*(3), 243–262. https://doi.org/10.1207/s15327078in0703_2

Campbell, J. (1994). Objects and objectivity. *Proceedings of the British Academy*, *83*, 3–20.

Campbell, J. (1995). *Past, Space, and Self*. MIT Press.

Carruthers, P. (2008). Meta-cognition in Animals: A Skeptical Look. *Mind & Language*, *23*(1), 58–89. https://doi.org/10.1111/j.1468-0017.2007.00329.x

Carruthers, P. (2011). *The opacity of mid*. Oxford University Press.

Carruthers, P. (2017). Are epistemic emotions metacognitive? *Philosophical Psychology*, *30*(1–2), 58–78. https://doi.org/10.1080/09515089.2016.1262536

Chambers, K. C. (2018). Conditioned taste aversions. *World Journal of Otorhinolaryngology - Head and Neck Surgery*, *4*(1), 92–100. https://doi.org/10.1016/j.wjorl.2018.02.003

Charles, L., Van Opstel, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage*, *73*, 80–94.

Chater, N., & Heyes, C. (1994). Animal Concepts: Content and Discontent. *Mind & Language*, *9*(3), 209–246. https://doi.org/10.1111/j.1468-0017.1994.tb00224.x

Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language*, *35*(1), 26–58.

Churchland, P. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, *78*(2), 67–90.

Clark, A. (2015). *Surfing Uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*(6699), 272–274. https://doi.org/10.1038/26216

Clayton, N. S., Yu, K. S., & Dickinson, A. (2003). Interacting cache memories: Evidence for flexible memory use by Western scrub-jays (Aphelocoma californica). *Journal of*

Experimental Psychology: Animal Behavior Processes, *29*(1), 14–22. https://doi.org/10.1037/0097-7403.29.1.14

Cole, L. E. (1939). A Comparison of the Factors of Practice and Knowledge of Experimental Procedure in Conditioning the Eyelid Response of Human Subjects. *The Journal of General Psychology*, *20*(2), 349–373. https://doi.org/10.1080/00221309.1939.9710016

Conant, R. C., & Ross Ashby, W. (1970). Every good regulator of a system must be a model of that system †. *International Journal of Systems Science*, *1*(2), 89–97. https://doi.org/10.1080/00207727008920220

Corlett, P. R., Honey, G. D., & Fletcher, P. C. (2007). From prediction error to psychosis: Ketamine as a pharmacological model of delusions. *Journal of Psychopharmacology*, *21*(3), 238–252. https://doi.org/10.1177/0269881107077716

Csibra, G. (2000). Gamma Oscillations and Object Processing in the Infant Brain. *Science*, *290*(5496), 1582–1585. https://doi.org/10.1126/science.290.5496.1582

Cussins, A. (1992). Content, Embodiment and Objectivity: The Theory of Cognitive Trails. *Mind*, *101*(404), 651–686.

Dagnall, N., Denovan, A., Parker, A., Drinkwater, K., & Walsh, R. S. (2018). Confirmatory Factor Analysis of the Inventory of Personality Organization-Reality Testing Subscale. *Frontiers in Psychology*, *9*, 1116. https://doi.org/10.3389/fpsyg.2018.01116

Davidson, D. (1975). *Inquiries into truth and interpretation*. Oxford University Press.

de Sousa, R. (2009). *Epistemic Feelings*. *7*(2), 139–161.

DeCasper, A., & Fifer, W. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

Delamater, A. R., LoLordo, V. M., & Berridge, K. C. (1986). Control of Fluid Palatability by Exteroceptive Pavlovian Signals. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*(2), 143–152.

Dellantonio, S., & Pastore, L. (2019). How Can You Be Sure? Epistemic Feelings as a Monitoring System for Cognitive Contents. In Á. Nepomuceno-Fernández, L. Magnani, F. J. Salguero-Lamillar, C. Barés-Gómez, & M. Fontaine (Eds.), *Model-Based Reasoning in Science and Technology* (Vol. 49, pp. 407–426). Springer International Publishing. https://doi.org/10.1007/978-3-030-32722-4_23

DeWitt, L. A., & Samuel, A. G. (1986). Perceptual restoration of music. *The Journal of the Acoustical Society of America*, *80*(S1), S110–S110. https://doi.org/10.1121/1.2023559

Dokic, J. (2012). Seeds of self-knowledge: Noetic feelings and metacognition. In J. Brandl, J. Perner, & J. Proust (Eds.), *The Foundations of Metacognition* (pp. 302–321). Oxford University Press.

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, *14*(1), 37–40. https://doi.org/10.3758/BF03198614

Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press.

Duke, D., Fiacconi, C. M., & Kohler, S. (2014). Parallel effects of processing fluency and positive affect on familiarity-based recognition decisions for faces. *Frontiers in Psychology*, *5*, 1–11. https://doi.org/10.3389/fpsyg.2014.00328

Dummett, M. (1993). *Origins of Analytic Philosophy*. Harvard University Press.

Elliott, R., McKenna, P., Robbins, T., & Sahakian, B. (1995). Neuropsychological evidence for frontostriatal dysfunction in schizophrenia. *Psychological Medicine*, *25*(3), 619–630.

Ellson, D. G. (1941). Hallucinations produced by sensory conditioning. *Journal of Experimental Psychology, 28*(1), 1–20. https://doi.org/10.1037/h0054167

Emery, N. J., Dally, J. M., & Clayton, N. S. (2004). Western scrub-jays (Aphelocoma californica) use cognitive strategies to protect their caches from thieving conspecifics. *Animal Cognition*, *7*(1), 37–43. https://doi.org/10.1007/s10071-003-0178-7

Evans, G. (1982). *The Varieties of Reference*. Oxford University Press.

Evans, G. (1985). Things without the mind: A commentary upon Chapter 2 of Strawson's Individuals. In *Collected Papers* (pp. 249–290). Clarendon Press.

Evans, J. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Evans, J., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Fagan, J. (1973). Infants' delayed recognition memory and forgetting. *Journal of Experimental Child Psychology*, *16*, 424–450.

Fagan, J. F. (1970). Memory in the infant. *Journal of Experimental Child Psychology*, *9*(2), 217–226. https://doi.org/10.1016/0022-0965(70)90087-1

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and Lexical Contributions to the Repetition Effect in Word Identification. *Journal of Experimenal Psychology: General*, *112*(3), 309–346.

Fields, P. E. (1932). Studies in concept formation I: The development of the concept of triangularity by the white rat. *Comparative Psychology Monographs*, *9*, 1–70.

Finke, R., & Pinker, S. (1983). Directional scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(3), 398–410.

Finn, B., & Tauber, S. K. (2015). When Confidence Is Not a Signal of Knowing: How Students' Experiences and Beliefs About Processing Fluency Can Lead to Miscalibrated Confidence. *Educational Psychology Review*, *27*(4), 567–586. https://doi.org/10.1007/s10648-015-9313-7

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911.

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, *1*(4), 324–340. https://doi.org/10.1016/0010-0285(70)90019-8

Flavell, J. H., Miller, P. H., & Miller, S. A. (2002). *Cognitive development* (4th ed). Prentice Hall.

Flavell, J., & Wellman, H. (1975). *Metamemory*. 83rd Meeting of the American Psychological Association, Chicago, Il.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the Rat. *Current Biology*, *17*(6), 551–555. https://doi.org/10.1016/j.cub.2007.01.061

Franco, F., & Butterworth, G. (1996). Pointing and social awareness: Declaring and requesting in the second year. *Journal of Child Language*, *23*(2), 307–336. https://doi.org/10.1017/S0305000900008813

Friedman, O., & Leslie, A. M. (2007). The conceptual underpinnings of pretense: Pretending is not 'behaving-as-if.' *Cognition, 105*(1), 103–124. https://doi.org/10.1016/j.cognition.2006.09.007

Fry, B. R., Russell, N., Gifford, R., Robles, C. F., Manning, C. E., Sawa, A., Niwa, M., & Johnson, A. W. (2020). Assessing reality testing in mice through dopamine-dependent associatively evoked processing of absent gustatory stimuli. *Schizophrenia Bulletin*, *46*(1), 54–67. https://doi.org/10.1093/schbul/sbz043

Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology*, *11*(3), 199–217. https://doi.org/10.1353/ppp.2004.0063

Garvey, C. R. (1933). A study of conditioned respiratory changes. *Journal of Experimental Psychology*, *16*(4), 471–503.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton, Mifflin, and Co.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* Oxford University Press.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*(1), 1–14. https://doi.org/10.1017/S0140525X00028636

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, *1*(2), 158–171. https://doi.org/10.1111/j.1468-0017.1986.tb00324.x

Graf, L. K. M., Mayer, S., & Landwehr, J. R. (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology*, *28*(3), 393–411. https://doi.org/10.1002/jcpy.1021

Green, M. F., & Kinsbourne, M. (1989). Auditory hallucinations in schizophrenia: Does humming help? *Biological Psychiatry*, *25*(5), 633–635. https://doi.org/10.1016/0006-3223(89)90225-4

Griffin, J. D., & Fletcher, P. C. (2017). Predictive processing, source monitoring, and psychosis. *Annual Review of Clinical Psychology*, *13*(1), 265–289. https://doi.org/10.1146/annurev-clinpsy-032816-045145

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews*, *55*, 88–97. https://doi.org/10.1016/j.neubiorev.2015.04.006

Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological Reviews*, *90*(3), 983–1012. https://doi.org/10.1152/physrev.00026.2009

Gruber, R., Schiestl, M., Boeckle, M., Frohnwieser, A., Miller, R., Gray, R. D., Clayton, N. S., & Taylor, A. H. (2019). New Caledonian crows Use mental representations to solve metatool problems. *Current Biology*, *29*(4), 686-692.e3. https://doi.org/10.1016/j.cub.2019.01.008

Grush, R. (2000). Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind*, *1*, 59–92.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*, 377–442.

Grush, R. (2007). Skill theory v2.0: Dispositions, emulation, and spatial perception. *Synthese*, *159*, 389–416.

Gurtner, L. M., Bischof, W. F., & Mast, F. W. (2019). Recurrence quantification analysis of eye movements during mental imagery. *Journal of Vision, 19*(1), 17. https://doi.org/10.1167/19.1.17

Gurtner, L. M., Hartmann, M., & Mast, F. W. (2021). Eye movements during visual imagery and perception show spatial correspondence but have unique temporal signatures. *Cognition, 210*, 104597. https://doi.org/10.1016/j.cognition.2021.104597

Hall, W. C. (2017). What You Don't Know Can Hurt You: The Risk of Language Deprivation by Impairing Sign Language Development in Deaf Children. *Maternal and Child Health Journal*, *21*(5), 961–965. https://doi.org/10.1007/s10995-017-2287-y

Hamm, J., Matheson, W., & Honig, W. (1997). Mental rotation in pigeons (Columba livia)? *Journal of Comparative Psychology*, *111*(1), 76–81.

Hanczakowski, M., Pasek, T., Zawadzka, K., & Mazzoni, G. (2013). Cue familiarity and 'don't know' responding in episodic memory tasks. *Journal of Memory and Language*, *69*(3), 368–383. https://doi.org/10.1016/j.jml.2013.04.005

Hart, J. (1965). *Recall, Recognition, and the Memory-Monitoring Process [Doctoral Dissertation, Stanford University]*.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216. https://doi.org/10.1037/h0022263

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1263–1271. https://doi.org/10.1098/rstb.2008.0296

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review*, *24*(1), 33–73. https://doi.org/10.1016/j.dr.2003.09.007

Hayne, H., & Imuta, K. (2011). Episodic memory in 3- and 4-year-old children. *Developmental Psychobiology*, *53*(3), 317–322. https://doi.org/10.1002/dev.20527

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*(4), 285–302.

Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition*, *38*(6), 771–784. https://doi.org/10.3758/MC.38.6.771

Hertzog, C., Fulton, E. K., Sinclair, S. M., & Dunlosky, J. (2014). Recalled aspects of original encoding strategies influence episodic feelings of knowing. *Memory & Cognition*, *42*(1), 126–140. https://doi.org/10.3758/s13421-013-0348-z

Heyes, C. (2015). Animal mindreading: What's the problem? *Psychonomic Bulletin & Review*, *22*(2), 313–327. https://doi.org/10.3758/s13423-014-0704-4

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, *21*(1), 101–114. https://doi.org/10.1017/S0140525X98000703

Heyes, C. M., Dawson, G., & Nokes, T. (1992). Imitation in rats: Initial responding and transfer evidence. *The Quarterly Journal of Experimental Psychology*, *45B*(3), 229–240.

Hofman, P. M., Van Riswick, J. G. A., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience*, *1*(5), 417–421. https://doi.org/10.1038/1633

Hohwy, J., & Rosenberg, R. (2005). Unusual experiences, reality testing and delusions of alien control. *Mind and Language*, *20*(2), 141–162. https://doi.org/10.1111/j.0268-1064.2005.00280.x

Holland, P. (1998). Amount of training affects associatively-activated event representation. *Neuropharmacology*, *37*(4–5), 461–469. https://doi.org/10.1016/S0028-3908(98)00038-0

Holland, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition*, *37*(1–2), 105–131. https://doi.org/10.1016/0010-0277(90)90020-K

Holland, P. C. (2005). Amount of training effects in representationmediated food aversion learning: No evidence of a role for associability changes. *Learning & Behavior*, *33*(4), 464–478. https://doi.org/10.3758/BF03193185

Hollard, V., & Delius, J. (1982). Rotational invariance in visual pattern recognition by pigeons and humans. *Science*, *18*(4574), 804–806.

Holsanova, J., Hedberg, B., & Nilsson, N. (1999). Visual and verbal focus patterns when describing pictures. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current oculomotor research* (pp. 303–304). Springer.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059–3071. https://doi.org/10.1121/1.2188377

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, *9*(2), 131–134.

Hosey, L. A., Peynircioğlu, Z. F., & Rabinovitz, B. E. (2009). Feeling of knowing for names in response to faces. *Acta Psychologica*, *130*(3), 214–224. https://doi.org/10.1016/j.actpsy.2008.12.007

Howells, T. H. (1944). The experimental development of color-tone synesthesia. *Journal of Experimental Psychology*, *34*(2), 87–103. https://doi.org/10.1037/h0054424

Hume, D. (1739). *A treatise of human nature*. Penguin Classics.

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Disentangling the effects of spatial cues on selection and formation of auditory objects. *The Journal of the Acoustical Society of America*, *124*(4), 2224–2235. https://doi.org/10.1121/1.2973185

Irwin, F. W., Smith, W. A. S., & Mayfield, J. F. (1956). Tests of two theories of decision in an "expanded judgment" situation. *Journal of Experimental Psychology*, *51*(4), 261–268. https://doi.org/10.1037/h0041911

Isingrini, M., Sacher, M., Perrotin, A., Taconnat, L., Souchay, C., Stoehr, H., & Bouazzaoui, B. (2016). Episodic feeling-of-knowing relies on noncriterial recollection and familiarity: Evidence using an online remember-know procedure. *Consciousness and Cognition*, *41*, 31–40. https://doi.org/10.1016/j.concog.2016.01.011

Jachs, B., Blanco, M. J., Grantham-Hill, S., & Soto, D. (2015). On the independence of visual awareness and metacognition: A signal detection theoretic analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 269–276. https://doi.org/10.1037/xhp0000026

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experiental Psychology: Learning, Memory, and Cognition*, *9*(1), 21–38.

Johnson, M. K., & Raye, C. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development*, *74*(1), 94–108. https://doi.org/10.1111/1467-8624.00523

Jozet-Alves, C., Bertin, M., & Clayton, N. S. (2013). Evidence of episodic-like memory in cuttlefish. *Current Biology*, *23*(23), R1033–R1035. https://doi.org/10.1016/j.cub.2013.10.021

Kagan, J., & Hamburg, M. (1981). The enhancement of memory in the first year. *The Journal of Genetic Psychology*, *138*(1), 3–14. https://doi.org/10.1080/00221325.1981.10532837

Keddy-Hector, A., Allen, C., & Friend, T. (1999). *Cognition in domestic pigs: Relational concepts and error recognition*. Unpublished Manuscript.

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, *98*(2–3), 127–140. https://doi.org/10.1016/S0001-6918(97)00039-5

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*(4), 483–524. https://doi.org/10.1016/0010-0285(83)90017-8

Kerfoot, E. C., Agarwal, I., Lee, H. J., & Holland, P. C. (2007). Control of appetitive and aversive taste-reactivity responses by an auditory conditioned stimulus in a devaluation task: A FOS and behavioral analysis. *Learning and Memory*, *14*(9), 581–589. https://doi.org/10.1101/lm.627007

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015

Kim, H., & Koh, H.-Y. (2016). Impaired reality testing in mice lacking phospholipase Cβ1: Observed by persistent representation-mediated taste aversion. *PLOS ONE*, *11*(1), e0146376. https://doi.org/10.1371/journal.pone.0146376

Koh, M. T., Ahrens, P. S., & Gallagher, M. (2018). A greater tendency for representation mediated learning in a ketamine mouse model of schizophrenia. *Behavioral Neuroscience*, *132*(2), 106–113. https://doi.org/10.1037/bne0000238

Kononowicz, T. W., van Wassenhove, V., & Doyère, V. (2022). Rodents monitor their error in self-generated duration on a single trial basis. *Proceedings of the National Academy of Sciences, 119*(9), e2108850119. https://doi.org/10.1073/pnas.2108850119

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimenal Psychology: General*, *126*(4), 349–370.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, *9*(2), 149–171. https://doi.org/10.1006/ccog.2000.0433

Koriat, A. (2007). Metacognition and consciousness. In P. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press.

Koriat, A., & Ackerman, R. (2010). Metacognition and mindreading: Judgments of learning for Self and Other during self-paced study. *Consciousness and Cognition*, *19*(1), 251–264. https://doi.org/10.1016/j.concog.2009.12.010

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 187–194. https://doi.org/10.1037/0278-7393.31.2.187

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 34–53. https://doi.org/10.1037/0278-7393.27.1.34

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*(4), 478–492. https://doi.org/10.1016/j.jml.2005.01.001

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 117–135). Psychology Press. https://doi.org/10.4324/9780203805503.ch7

Kosslyn, S. M. (1980). *Image and mind*. Harvard University Press.

Kosslyn, S. M. (1991). Visual mental images in the brain. *Proceedings of the American Philosophical Society*, *135*(4), 524–532.

Kosslyn, S. M., Ball, T. M., & Brian Reiser. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(1), 47–60.

Krystal, J., Karper, L., Seibyl, J., Freeman, G., Delaney, R., Bremner, J. D., Heninger, G., Bowers Jr., M., & Charney, D. (1994). Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans: Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. *Archives of General Psychiatry*, *51*(3), 199–214.

LaBerge, S., Baird, B., & Zimbardo, P. G. (2018). Smooth tracking of visual targets distinguishes lucid REM sleep dreaming and waking perception from imagination. *Nature Communications*, *9*(1), 3298. https://doi.org/10.1038/s41467-018-05547-0

Lee, D., & Kalmus, H. (1980). The optic flow field: The foundation of vision [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *290*(1038), 169–179.

Lenzenweger, M. F., Clarkin, J. F., Kernberg, O. F., & Foelsch, P. A. (2001). The Inventory of Personality Organization: Psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and self-domains in a nonclinical sample. *Psychological Assessment*, *13*(4), 577–591. https://doi.org/10.1037/1040-3590.13.4.577

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Lillard, A., Pinkham, A. M., & Smith, E. (2010). Pretend play and cognitive development. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (1st ed., pp. 285–311). Wiley. https://doi.org/10.1002/9781444325485.ch11

Liu, Y., Su, Y., Xu, G., & Chan, R. C. K. (2007). Two dissociable aspects of feeling-of-knowing: Knowing that you know and knowing that you do not know. *Quarterly Journal of Experimental Psychology*, *60*(5), 672–680. https://doi.org/10.1080/17470210601184039

Locke, J. (1689). *An essay concerning human understanding*. Oxford University Press.

Lotto, A., & Holt, L. (2011). Psychology of auditory perception. *WIREs Cognitive Science*, *2*(5), 479–489. https://doi.org/10.1002/wcs.123

Lycan, W. (1996). *Consciousness and experience*. MIT Press.

Mack, A., & Rock, I. (1998). *Inattentional Blindness*. MIT Press.

Mandik, P. (1998). Objectivity without space. *Electronic Journal of Analytic Philosophy*, *6*.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412. https://doi.org/10.3758/BF03204884

Matthen, M. (2010). Two visual systems and the feeling of presence. In *Perception, Action, and Consciousness* (pp. 107–124). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199551118.001.0001

Matussek, P. (1987). Studies in delusional perception. In J. Cutting & M. Shepherd (Eds.), *The clinical roots of the schizophrenia concept: Translations of seminal European contributions on schizophrenia* (pp. 89–103). Cambridge University Press.

Mayer-Gross, W., & Stein, J. (1928). Pathologie der Wahrnehmung. Psychopathologie und Klinik der Trugwahrnehmungen. In O. Bumke (Ed.), *Handbuch der Geisteskrankheiten* (pp. 352–507). Springer.

McDannald, M. A., Whitt, J. P., Calhoon, G. G., Piantadosi, P. T., Karlsson, R.-M., O'Donnell, P., & Schoenbaum, G. (2011). Impaired reality testing in an animal model of schizophrenia. *Biological Psychiatry*, *70*(12), 1122–1126. https://doi.org/10.1016/j.biopsych.2011.06.014

McDannald, M., & Schoenbaum, G. (2009). Toward a model of impaired reality testing in rats. *Schizophrenia Bulletin*, *35*(4), 664–667. https://doi.org/10.1093/schbul/sbp050

McDowell, J. (1994). *Mind and World*. Harvard University Press.

McFarlane, A. (1975). Olfaction in the development of social preferences in the human neonate. *Ciba Foundation Symposium*, *33*, 103–117.

McGinn, C. (2006). *Mindsight: Image, dream, meaning* (1. paperback ed). Harvard Univ. Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.

Meltzoff, A. (1988). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development*, *59*(1), 217–225.

Melzack, R., & Wall, P. D. (1988). *The challenge of pain*. Basic Books.

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1084–1097. https://doi.org/10.1037/a0012580

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851–861. https://doi.org/10.1037/0278-7393.19.4.851

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition: Knowing about knowing*. MIT Press.

Michaelian, K. (2016). *Mental time travel: Episodic memory and our knowledge of the personal past*. MIT Press.

Millar, B. (2021). Towards a sensorimotor approach to flavour and smell. *Mind & Language, 36*(2), 221–240. https://doi.org/10.1111/mila.12275

Millikan, R. (1987). *Language, thought, and other biological categories*. MIT Press.

Moore, C., & Corkum. (1994). Social understanding at the end of the first year of life. *Developmental Review*, *14*, 349–371.

Morgan, C. (1894). *An introductino to comparative psychology*. Walter Scott.

Morissette, P., Ricard, M., & Décarie, T. G. (1995). Joint visual attention and pointing in infancy: A longitudinal study of comprehension. *British Journal of Developmental Psychology*, *13*(2), 163–175. https://doi.org/10.1111/j.2044-835X.1995.tb00671.x

Morris, R. G. M. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation*, *12*(2), 239–260. https://doi.org/10.1016/0023-9690(81)90020-5

Mullally, S. L., & Maguire, E. A. (2014). Learning to remember: The early ontogeny of episodic memory. *Developmental Cognitive Neuroscience, 9*, 12–29. https://doi.org/10.1016/j.dcn.2013.12.006

Nakano, T., & Kitazawa, S. (2017). Development of long-term event memory in preverbal infants: An eye-tracking study. *Scientific Reports, 7*(1), 44086. https://doi.org/10.1038/srep44086

Nanay, B. (2015). Perceptual content and the content of mental imagery. *Philosophical Studies, 172*(7), 1723–1736. https://doi.org/10.1007/s11098-014-0392-y

Neisser, U. (1976). *Cognition and Reality*. W. H. Freeman.

Neisser, U. (1978). Anticipations, images, and introspection. *Cognition, 6*(2), 169–174. https://doi.org/10.1016/0010-0277(78)90021-5

Neiworth, J. J., & Rilling, M. E. (1987). A method for studying imagery in animals. *Journal of Experimental Psychology: Animal Behavior Processes, 13*(3), 203–214.

Nelson, T. (1992). *Metacognition: Core readings*. Allyn & Bacon.

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113*(2), 282–300.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). Elsevier. https://doi.org/10.1016/S0079-7421(08)60053-5

Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford University Press.

Nordmark, J. (1970). Time and frequency analysis. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (pp. 57–83). Academic Press.

O'Keefe, J. (1994). Cognitive maps, time, and causality. *Proceedings of the British Academy, 83*, 35–45.

Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica*, *124*(1), 106–128. https://doi.org/10.1016/j.actpsy.2006.09.009

O'Regan, J. K., & Noe, A. (2001). What It Is like to see: A sensorimotor theory of perceptual experience. *Synthese*, *129*(1), 79–103.

Pahl, M., Zhu, H., Pix, W., Tautz, J., & Zhang, S. (2007). Circadian timed episodic-like memory – a bee knows what to do when,and also where. *Journal of Experimental Biology*, *210*(20), 3559–3567. https://doi.org/10.1242/jeb.005488

Panoz-Brown, D., Corbin, H. E., Dalecki, S. J., Gentry, M., Brotheridge, S., Sluka, C. M., Wu, J.-E., & Crystal, J. D. (2016). Rats remember items in context using episodic memory. *Current Biology*, *26*(20), 2821–2826. https://doi.org/10.1016/j.cub.2016.08.023

Panoz-Brown, D., Iyer, V., Carey, L. M., Sluka, C. M., Rajic, G., Kestenman, J., Gentry, M., Brotheridge, S., Somekh, I., Corbin, H. E., Tucker, K. G., Almeida, B., Hex, S. B., Garcia, K. D., Hohmann, A. G., & Crystal, J. D. (2018). Replay of episodic memories in the rat. *Current Biology*, *28*(10), 1628-1634.e7. https://doi.org/10.1016/j.cub.2018.04.006

Panton, K. R., Badcock, D. R., & Badcock, J. C. (2016). A metaanalysis of perceptual organization in schizophrenia, schizotypy, and other high-risk groups based on variants of the embedded figures task. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00237

Paynter, C. A., Reder, L. M., & Kieffaber, P. D. (2009). Knowing we know before we know: ERP correlates of initial feeling-of-knowing. *Neuropsychologia*, *47*(3), 796–803. https://doi.org/10.1016/j.neuropsychologia.2008.12.009

Pearce, J. M. (1989). The acquisition of an artificial category by pigeons. *The Quarterly Journal of Experimental Psychology*, *41*(4), 381–406.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *86*(6), 21.

Perky, C. W. (1910). An experimental study of imagination. *The American Journal of Psychology*, *21*(3), 422. https://doi.org/10.2307/1413350

Perrin, D., Michaelian, K., & Sant'Anna, A. (2020). The phenomenology of remembering Is an epistemic feeling. *Frontiers in Psychology*, *11*, 1531. https://doi.org/10.3389/fpsyg.2020.01531

Perris, E. E., Myers, N. A., & Clifton, R. (1990). Long-term memory for a single infancy experience. *Child Development*, *61*(6), 1796–1807.

Piaget, J. (1952). *The origins of intelligence in children*. International Universities Press.

Piaget, J. (1962). *Play, dreams, and imitation in childhood*. Norton.

Pienkos, E., Giersch, A., Hansen, M., Humpston, C., McCarthy-Jones, S., Mishara, A., Nelson, B., Park, S., Raballo, A., Sharma, R., Thomas, N., & Rosen, C. (2019). Hallucinations beyond voices: A conceptual review of the phenomenology of altered perception in psychosis. *Schizophrenia Bulletin*, *45*(Supplement_1), S67–S77. https://doi.org/10.1093/schbul/sby057

Pitt, D. (2004). The phenomenology of cognition or What is it like to think that P? *Philosophy and Phenomenological Research*, *69*(1), 1–36.

Pleskac, T., & Busemeyer, J. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 864–901.

Polger, T., & Shapiro, L. (2016). *The multiple realization book*. Oxford University Press.

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science, 357*(6351), 596–600. https://doi.org/10.1126/science.aan3458

Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness* (First edition). Oxford University Press.

Quinn, P. C., & Intraub, H. (2007). Perceiving "outside the box" occurs early in development: Evidence for boundary extension in three- to seven-month-old infants. *Child Development*, *78*(1), 324–334. https://doi.org/10.1111/j.1467-8624.2007.01000.x

Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature, 445*(7130), 919–921. https://doi.org/10.1038/nature05575

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review*, *120*(3), 697–719. https://doi.org/10.1037/a0033152

Reber, R., Fazendeiro, T. A., & Winkielman, P. (2002). Processing fluency as the source of experiences at the fringe of consciousness. *Psyche*, *8*(10), 1–21.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experiental Psychology: Learning, Memory, and Cognition*, *18*(3), 435–451.

Reeves, A. (1980). Visual imagery in backward masking. *Perception & Psychophysics*, *28*(2), 118–124. https://doi.org/10.3758/BF03204336

Roitblat, H. L. (1980). Codes and coding processes in pigeon short-term memory. *Animal Learning & Behavior*, *8*(3), 341–351. https://doi.org/10.3758/BF03199615

Rovee-Collier, C. (1997). Dissociations in infant memory: Rethinking the development of implicit and explicit memory. *Psychological Review*, *104*(3), 467–498.

Russell, J., & Thompson, D. (2003). Memory development in the second year: For events or locations? *Cognition*, *87*(3), B97–B105. https://doi.org/10.1016/S0010-0277(02)00238-X

Sartre, J.-P. (1966). *The psychology of imagination*. Washington Square Press.

Scarpelli, S., Bartolacci, C., D'Atri, A., Gorgoni, M., & De Gennaro, L. (2019). The functional role of dreaming in emotional processes. *Frontiers in Psychology*, *10*, 1–16. https://doi.org/10.3389/fpsyg.2019.00459

Schacter, D. L., & Moscovitch, M. (1984). Infants, amnesics, and dissociable memory systems. In M. Moscovitch (Ed.), *Infant Memory* (pp. 173–216). Springer US. https://doi.org/10.1007/978-1-4615-9364-5_8

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Schwartz, B. L. (2010). The effects of emotion on tip-of-the-tongue states. *Psychonomic Bulletin & Review*, *17*(1), 82–87. https://doi.org/10.3758/PBR.17.1.82

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition*, *39*(5), 737–749. https://doi.org/10.3758/s13421-010-0066-8

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*(2), 195–202.

Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Nous*, *36*(2), 249–275.

Schwitzgebel, E. (2014). Introspection. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

Segal, S. J., & Fusella, V. (1970). Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology, 83*(3, Pt.1), 458–464. https://doi.org/10.1037/h0028840

Shaw, R. C., & Clayton, N. S. (2013). Careful cachers and prying pilferers: Eurasian jays (Garrulus glandarius) limit auditory information available to competitors. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1752), 20122238. https://doi.org/10.1098/rspb.2012.2238

Shergill, S. S., Murray, R. M., & McGuire, P. K. (1998). Auditory hallucinations: A review of psychological treatments. *Schizophrenia Research*, *32*(3), 137–150. https://doi.org/10.1016/S0920-9964(98)00052-8

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology*, *7*, 1–8. https://doi.org/10.3389/fpsyg.2016.00218

Simons, D., & Levin, D. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.

Skov-Rackette, S. I., Miller, N. Y., & Shettleworth, S. J. (2006). What-where-when memory in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(4), 345–358. https://doi.org/10.1037/0097-7403.32.4.345

Smith, J. D. (2007). Studies of uncertainty monitoring and metacognition in animals and humans. In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 242–271). Oxford University Press.

Smitha, B., Thakar, J., & Watve, M. (1999). Do bee eaters have theory of mind? *Current Science, 76*(4), 574–577.

Snyder, J. S., Carter, O. L., Lee, S.-K., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 1007–1016. https://doi.org/10.1037/0096-1523.34.4.1007

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry, 84*(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

Sterzer, P., Mishara, A. L., Voss, M., & Heinz, A. (2016). Thought insertion as a self-disturbance: An integration of predictive coding and phenomenological approaches. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00502

Strawson, P. F. (1959). *Individuals: An essay in descriptive metaphysics*. Routledge.

Strawson, P. F. (1982). Imagination and perception. In R. Walker (Ed.), *Kant on pure reason* (pp. 82–99). Oxford University Press.

Theeuwes, J., Belopolsky, A., & Olivers, C. N. L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica, 132*(2), 106–114. https://doi.org/10.1016/j.actpsy.2009.01.005

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2012). An analysis of the determinants of the feeling of knowing. *Consciousness and Cognition, 21*(4), 1681–1694. https://doi.org/10.1016/j.concog.2012.09.005

Thomas, N. J. T. (1999). Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cognitive Science*, *23*(2), 207–245.

Todorović, D. (2010). Context effects in visual perception and their explanations. *Review of Psychology*, *17*(1), 17–32.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208.

Tomasello, M. (2023). Social cognition and metacognition in great apes: A theory. *Animal Cognition*, *26*(1), 25–35. https://doi.org/10.1007/s10071-022-01662-0

Tulving, E. (1972). Episodic and semantic memory. In *Organization of memory* (pp. 381–402). Academic Press.

Tustin, K., & Hayne, H. (2010). Defining the boundary: Age-related changes in childhood amnesia. *Developmental Psychology*, *46*(5), 1049–1061. https://doi.org/10.1037/a0020105

Van Noorden, L. (1975). *Temporal coherence in the perception of tone sequences*. Eindhoven University of Technology.

Vickers, D., & Lee, M. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator model. *Nonlinear Dynamics, Psychology, and Life Sciences*, *2*(3), 169–193.

Vinckier, F., Gaillard, R., Palminteri, S., Rigoux, L., Salvador, A., Fornito, A., Adapa, R., Krebs, M. O., Pessiglione, M., & Fletcher, P. C. (2016). Confidence and psychosis: A neuro-computational account of contingency learning disruption by NMDA blockade. *Molecular Psychiatry*, *21*(7), 946–955. https://doi.org/10.1038/mp.2015.73

Volkman, J. (1934). The relation of the time of judgment to the certainty of judgment. *Psychological Bulletin*, *31*, 672–673.

von Helmholtz, H. (1924). *Helmholtz's treatise on physiological optics*. Optical Society of America.

Waldmann, M. R., Schmid, M., Wong, J., & Blaisdell, A. P. (2012). Rats distinguish between absence of events and lack of evidence in contingency learning. *Animal Cognition*, *15*(5), 979–990. https://doi.org/10.1007/s10071-012-0524-8

Waller, Schweitzer, Brunton, & Knudson. (2012). A century of imagery research: Reflections on Cheves Perky's contribution to our understanding of mental imagery. *The American Journal of Psychology*, *125*(3), 291. https://doi.org/10.5406/amerjpsyc.125.3.0291

Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive Psychology*, *58*(3), 416–440. https://doi.org/10.1016/j.cogpsych.2008.09.003

Watanabe, A. (2018). Exploring the bird mind: A review of episodic memory and metacognition studies of western scrub-jays. *Japanese Journal of Animal Psychology*, *68*(1), 57–65. https://doi.org/10.2502/janip.68.1.4

Whishaw, I. (1991). Latent learning in a swimming pool place task for rats: Evidence for the use of associative and not cognitive mapping processes. *The Quarterly Journal of Experimental Psychology*, *43B*(1), 83–103.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica*, *98*(2–3), 141–165. https://doi.org/10.1016/S0001-6918(97)00040-1

Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings of familiarity. *Journal of Experiental Psychology: Learning, Memory, and Cognition*, *27*(1), 3–13.

Williams, D. (2020). Predictive coding and thought. *Synthese*, *197*(4), 1749–1775. https://doi.org/10.1007/s11229-018-1768-x

Winer, G., & Cottrell, J. (2004). The odd belief that rays exit the eye during vision. In D. Levin (Ed.), *Thinking and Seeing: Visual Metacognition in Adults and Children* (pp. 97–119). MIT Press.

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Erlbaum.

Woods, A. T., Poliakoff, E., Lloyd, D. M., Dijksterhuis, G. B., & Thomas, A. (2010). Flavor expectation: The effect of assuming homogeneity on drink perception. *Chemosensory Perception*, *3*(3–4), 174–181. https://doi.org/10.1007/s12078-010-9080-2

Woolley, J. D., & Van Reet, J. (2006). Effects of context on judgments concerning the reality status of novel entities. *Child Development*, *77*(6), 1778–1793. https://doi.org/10.1111/j.1467-8624.2006.00973.x

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, *144*(2), 489–510. https://doi.org/10.1037/xge0000062