Philosophical Psychology

ISSN: 0951-5089 (Print) 1465-394X (Online) Journal homepage: www.tandfonline.com/journals/cphp20

Reality testing and metacognition

Nathaniel Greely

To cite this article: Nathaniel Greely (2025) Reality testing and metacognition, Philosophical Psychology, 38:6, 2465-2488, DOI: 10.1080/09515089.2024.2336588

To link to this article: https://doi.org/10.1080/09515089.2024.2336588

Published online: 09 Apr 2024.

Submit your article to this journal

Article views: 265

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cphp20

PHILOSOPHICAL PSYCHOLOGY
2025, VOL. 38, NO. 6, 2465–2488
https://doi.org/10.1080/09515089.2024.2336588

Routledge
Taylor & Francis Group

Check for updates

Reality testing and metacognition

Nathaniel Greely

Department of Philosophy, University of California San Diego, La Jolla, CA, USA

ABSTRACT

Reality testing is the process by which we distinguish our own perceptual states from imagination or episodic memory. I argue that reality testing is a metacognitive process. Since reality testing is also accomplished by creatures who lack mental state concepts, it follows that reality testing is a nonconceptual metacognitive process. I also provide prima facie evidence that reality testing is a necessary condition for prototypical cognitive states like belief. It follows that metacognition is phylogenetically and logically prior to cognition in many cases.

ARTICLE HISTORY
Received 22 September 2023
Accepted 22 March 2024

KEYWORDS
Metacognition; nonconceptual content; reality testing; cognitive ethology

1. Introduction

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

CONTACT Nathaniel Greely nategreely@icloud.com Department of Philosophy, University of California San Diego, 9500 Gilman Dr. #0119, La Jolla, CA 92093, USA

© 2024 Informa UK Limited, trading as Taylor & Francis Group

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 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; Johnson & Raye, 1981; Kim et al., 2016; McDannald & Schoenbaum, 2009).[1]

It is uncontroversial that the explicit form of reality testing that occurred on Night 2 is metacognitive. Beliefs that 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 explicit and implicit reality testing the same type of process?

Metacognition is taken by many philosophers and psychologists to be a relatively sophisticated type of mental state (e.g., Byrne, 2018; Carruthers, 2011; Gopnik, 1993).[2] This is because it is often 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. Some claim that to engage in metacognition about one's own mental states, one must first be able to think about other people's mental states, an ability termed *mindreading* (Gopnik, 1993). This latter view, taken to the extreme, is the "opacity of mind" thesis (Carruthers, 2011). 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 (e.g., Byrne, 2018) claim that metacognitive states are the result of mental operations applied to first order states – I ask myself whether the cat is on the mat, and if the answer is yes, I conclude that I believe the cat is on the mat. 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 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." The mechanisms underlying sensory metacognition have lately become an active area of study (See Rahnev, 2021 for a review). The central construct of this research is metacognitive confidence – higher-order judgments about the accuracy of one's perceptual states. This construct has been applied to creatures who arguably lack the mental state concepts required to engage in more sophisticated forms of metacognition – rats, corvids, and human infants among others. But dispute remains over whether the measures of confidence employed are truly metacognitive (See, e.g., Carruthers, 2008).

In this article I plan to bypass this dispute by focusing on reality testing. The metacognitive status of reality testing has often simply been assumed. In other cases, it hinges on the validity of confidence as a metacognitive measure (Schmack et al., 2021). But the claim that reality testing is a metacognitive process is not as obvious as it might seem. And it is too important to be left undefended or contingent on the validity of any particular metacognitive measure. I will offer a sustained philosophical defense of the claim that reality testing is a metacognitive process. If this claim is true, there are important implications. A complete argument for those implications requires more space, but I will offer prima facie evidence for two further claims. First, there is evidence that many creatures who lack mental state concepts perform reality testing. Thus, reality testing is a heretofore unrecognized form of *nonconceptual* metacognition. Second, extant accounts of the necessary conditions for basic cognitive states like belief implicitly assume a reality testing ability. Thus, metacognition is required for many basic forms of cognition.

In section 2 I will introduce the phenomenon of reality testing. I will then argue that reality testing is a metacognitive process. However, there is no agreed definition of metacognition, so in section 3 I will survey the phenomena that, in practice, serve as the objects of metacognition research. In section 4 I will then construct a definition that is at least extensionally adequate – an improvement on extant definitions. In section 5 I will argue that reality testing satisfies this definition of metacognition. In section 6 I will consider Nelson and Narens (1990) foundational account of metacognition and argue that reality testing meets their criteria as well. In section 7 I will offer some evidence that reality testing is nonconceptual and necessary for cognitive states like belief.

2. Reality testing

Reality testing is the process by which a subject distinguishes her own perceptual states from imagination or episodic memory. The term originates in the psychoanalytic literature (Arlow, 1969), but has made its way into philosophy (Hohwy & Rosenberg, 2005), cognitive psychology

(Johnson & Raye, 1981), clinical psychology (Dagnall et al., 2018), and cognitive ethology (Kim et al., 2016; McDannald & Schoenbaum, 2009). The term has been used in 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. Reality testing on this construal refers only to the way we sort our own representations into two classes – perception and mental imagery. This sorting may not be entirely accurate. In the psychoanalytic view, 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. Many contemporary accounts of perception also posit that everyday perceptual states are partly constituted by background knowledge or internally generated imagery (e.g., De Lange et al., 2018; Grush, 2004; Lau et al., 2022). If my perception of a tree involves, unbeknownst to me, tree memories, this is a failure of reality testing, but of a harmless sort. The imagery involved in dreams is another such harmless failure. Reality testing is a pragmatic, adaptive ability to sort imagery into classes – "reality" for us needn't be reality writ large.

Some failures of reality testing, however, are not adaptive. 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, and recent studies of reality testing in rodents use a similar interpretation (Kim et al., 2016; McDannald & Schoenbaum, 2009).

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 up 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 – 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.

Reality testing is distinct from two other processes described in the literature on memory – reality monitoring and source monitoring (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. "Perceptual reality monitoring", on the other hand, is frequently used in a sense that is synonymous with "reality testing" (e.g., Dijkstra et al., 2022; Lau, 2019). The feeling of presence is another similar concept (Matthen, 2010). It is the feeling that an object is nearby and has been invoked as one way that perception differs from imagination. If so, the feeling of presence might be a criterion 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 (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 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, it 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.

3. The extension of '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 and Wellman (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. There is no single, canonical definition of metacognition, and the attempts offered in the seminal texts are inadequate.

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 frequently 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). J. 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 – by restricting the object of the metacognitive state or restricting the representational format of the higher- or lower-order state involved. An extensionally adequate 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 a definition of metacognition should cover. 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 Section 5 I will consider some additional conditions on metacognition proposed by Nelson and Narens (1990).

3.1. 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. 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.

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. Young children lack certain metacognitive abilities, as demonstrated by their failure at tasks like the "false belief task," suggesting that they lack a theory of mind (J. H. Flavell et al., 1970; Gopnik, 1993). 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.

3.2. 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. Non-propositional forms of representation might also explain intelligent behavior in creatures who lack language. These claims are controversial, but the point is 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.[3] 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?

Phenomena like tip-of-the-tongue states, feelings of knowing, and feelings of confidence have been objects of metacognition research from its inception (Hart, 1965; 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. de Sousa (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 a 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 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). 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 (2013) compares epistemic feelings to Gibsonian affordances (Gibson, 1979) 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-Muñoz (2011), Jerome Dokic (2012), and

Asher Koriat (2000), though they sometimes vary in their precise accounts of the representational format involved.

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 (1689) 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, then, two types 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. 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.

3.3. 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 therefore arguably non-propositional.

3.4. 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 is certainly a subject of contemporary metacognition research as well.

Theory-theorists apply their account of mindreading to self-directed cases 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. Inner-sense accounts of introspection posit a quasi-perceptual mechanism of access to the content of one's own mind, and evaluativism posits that some introspection is non-propositional.

3.5. The extension of metacognition research

We now have two axes that define a space of metacognitive research – one for representational format and one for the subject whose mental states are represented by the metacognitive state (see Fig. 1). Theory-theoretical accounts of mindreading occupy the propositional, other-directed corner of the space, while the theory-theoretical introspection occupies the propositional, self-directed corner along with transparency theory. Epistemic feelings and inner sense 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 Inner sense
Other-Directed	Theory-Theory	Simulation Theory Nonconceptual Mindreading

Figure 1. Areas of metacognition research.

4. Defining metacognition

I have characterized the extension of metacognition research along two axes – representational format and owner of the target state. While there may be other ways of characterizing it, 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", but I will adopt the broadly accepted claim that mental states, processes, or properties involve representations. This may rule out some, but certainly not all, neural states, processes, and properties. And obviously not all representations qualify as mental – this sentence, for example.

But not anything goes. By characterizing these states[5] as intentional I assume a higher bar than mere causation. Mental states cause and are caused by other mental states, but this by itself does not make them higher-order states any more than an eight ball in billiards is a higher-order ball. If causation were sufficient for intentionality, the distinction between cognition and metacognition could collapse. For example, I have noted that many accounts of perception posit that perceptual states involve internally generated imagery. On such a thin account of intentionality a first-order perceptual state that takes a memory as a causal input would itself become a metacognitive state. Some accounts of intentionality take a causal relation to be necessary for an intentional relation, but further conditions are typically added – for example that the mental state in question has the function of indicating the presence of the causal input (e.g., Millikan, 1984). 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 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. 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 is metacognitive.

5. Reality testing is metacognitive

In its propositional form, reality testing is clearly metacognitive. Recall the two episodes described in the introduction. 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 this unconscious, automatic form of reality testing could have been accomplished and will argue that all the plausible options qualify as forms of metacognition.

5.1. Propositional and non-propositional unconscious 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 metacognitive states (pace Carruthers, 2017).

But 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, for example, 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 usually 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.

So, if nonconceptual content is involved, perhaps it is 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 or 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 "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 – opportunities to engage with one's own mental content. 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.[6] If you feel that you 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 child to sort computer parts by color and shape. This would not require that the child have any conception of computers or that she conceive of the parts as computer parts. Nonetheless, you could say that the child represents computer parts in a de re sense. But if she 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 child. 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 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 (1739) "vividness," and thus support the metacognitive nature of reality testing (see also Dijkstra & Fleming, 2023). Other accounts involve propositional content and are thus irrelevant to the hypothesis of nonconceptual reality testing. But one account offered by Goldman (2006) 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 be conceived *as* mental.

Goldman's remarks on the subject are brief, so a bit of interpretation is needed. He offers an analogy with pain perception. Particular types of pain sensations implicate particular types of nociceptors, and he claims that on this basis subjects individuate pain types. There is 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 also how we distinguish beliefs from desires and other mental states – by introspecting the neural types involved. Goldman does not extend the view to reality testing, but if the view worked, he might.

Goldman's is a strange picture in that it makes levels of explanation intersect that are typically kept distinct. On the traditional view, causal

interactions at the neural level manifest as causal interactions between psychological states at a higher level. Psychological events, we hope, can be reduced to neural events, but 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 such events are typed. 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 only to intrinsic properties of the neuron, such as axon diameter or myelination.

Goldman's view, if extended to reality testing, would be at odds with current research on the neural basis of reality testing. A great deal of overlap has been shown between the brain areas involved in imagination and perception, and researchers typically look to functional differences between the two processes to explain our reality testing ability (see Dijkstra et al., 2022 for a review).

5.2. First-order and 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.[7]

I have argued for the metacognitive status of conscious propositional reality testing, unconscious propositional reality testing, and non-propositional reality testing. But why assume that reality testing is a mental process in the first place? One possible explanation 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 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 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 "architectural" view I am considering here posits that reality testing is solved by the hardware of the brain. We can often infer 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. 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 to flex 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. In section 7 I will describe studies that produce failures of reality testing in rats and mice. The variables manipulated in these studies are psychological variables. The

ways that reality testing fails suggest an explanation at the psychological level. And if this is the case, then reality testing is best categorized as 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 is another prominent account of metacognition that places more specific conditions on what counts as metacognition. In the next section I will argue that reality testing satisfies these conditions.

6. Nelson and Narens' 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 Roger Conant and Ross Ashby (1970), who point out that such 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" (Grush, 2004). 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 Nelson and Narens' monitoring function that gets around this problem. On such accounts, 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. I have argued against evaluativism in detail elsewhere (Greely, 2021). Regardless, the upshot is that the claim that the monitoring function of metacognition requires a model of first-order processes is untenable.

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 my 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 the greatest philosopher of all time, 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 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. Elsewhere I argue that reality testing involves fine-grained sensorimotor skills that could be interpreted as accomplishing a control function over the perceptual and imagination processes themselves (Greely, 2023). To take one example, visual saccades will result in different visual experiences 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.

7. Reality testing is nonconceptual and necessary for propositional thought

I have argued that reality testing is a metacognitive ability. For some, this might seem obvious, but it should be argued for systematically. The argument is worth making because of its implications. First, if reality testing is accomplished by creatures that lack mental state concepts, then reality testing exists in a nonconceptual form, which entails that nonconceptual metacognition exists. Second, if we accept any of the most prominent accounts of the necessary conditions for conceptual mental content, reality testing is necessary for conceptual content. This entails that metacognition is logically prior to basic cognitive states like belief. In what follows I will offer reason to believe that these claims are plausible enough to merit further investigation.

Any creature that possesses faculties of perception and episodic memory or imagination must be able to perform reality testing if it is to live long. There is evidence that rodents, corvids, and human infants possess these faculties. Here I will confine my comments to rodents. Reality testing is studied explicitly in rats and mice, where failures of reality testing are induced by classical conditioning and then recovered by the rodents over time. A tone is paired with a flavor which is subsequently paired with nausea (in the absence of the tone). Later, the tone itself produces aversive behavior associated with an unpleasant flavor. This cannot be explained by direct association, as the tone was never paired with nausea, and so a hallucinatory flavor is posited (Holland, 1990). This interpretation is supported by the fact that the effect is accentuated by drugs and gene knockouts associated with schizophrenia (Kim et al., 2016) and eliminated by drugs which control schizophrenia (Fry et al., 2020). The effect goes away on its own eventually. This is interpreted as an induced failure of reality testing and a subsequent return of the ability. There is little evidence that rats possess mental state concepts. Studies have suggested evidence of empathy in rats – a tendency to

free conspecifics that are confined. But physical confinement is a physical situation and requires no attribution of mental states to see that it is bad.

A strain of philosophical scholarship beginning with Strawson (1959) and Evans (1985) considers the necessary conditions for the possession of conceptual content. Similar accounts are offered by Campbell (1994), O'Keefe (1994), Cussins (1992), Grush (2000), and Bermúdez (1998). Burge (2010) offers a very different one. The accounts vary, but the general idea is that in order for a creature to possess a concept like "red", it must possess representations with the type of compositional structure that allows "red" to be applied to any arbitrary object, which in turn requires a conception of objects - *objectivity*. Objectivity requires an ability to conceive of things existing independently of our perceiving them. Strawson and Bermudez argue that this requires the ability, upon re-encountering a previously perceived object, to identify it as numerically identical to one previously encountered. Cussins, O' Keefe, Campbell, and Grush argue that objectivity requires an ability to correlate perceived objects with an allocentric map in the imagination. In each case, the ability requires perception and episodic memory or imagination, and thus, I would argue, a reality testing ability. Burge disputes the relatively sophisticated nature of these accounts and instead argues that objectivity occurs in the early stages of visual processing. Nonetheless, his account invokes the participation of episodic memory and thus implicitly invokes reality testing. The upshot is that, if any of these accounts are correct, reality testing is a necessary condition for conceptual content. Thus, metacognition is prior to cognitive states like belief.

8. Conclusion

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 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. I then argued that reality testing satisfies this definition. I considered first-order or architectural accounts of reality testing and argued that they are insufficient to explain the phenomena. I considered some additional conditions on metacognition offered by Nelson and Narens' influential account and argued that reality testing meets these as well. Finally, I briefly previewed arguments that reality testing is both nonconceptual and necessary for conceptual content. Any attempt to understand cognition proper, understood as mental states like belief, requires an understanding of metacognition.

Notes

1. I did not coin the term "reality testing". To "test" reality may connote something conscious and sophisticated, but reality testing is often unconscious and, I argue, phylogenetically primitive.
2. The term "metacognition" has many, often incompatible, uses in the literature. Some researchers reserve the term "metacognition" for mental states employed in metamemory, such as feelings of knowing (e.g., Nelson & Narens, 1990). Proust (2013) claims that such processes are not "sophisticated" but are in fact entirely nonconceptual. Part of the goal of this essay will be to regiment the terminology in a more felicitous way. But that dispute is merely notational.
3. "Nonconceptual" is more commonly used – more on this shortly.
4. Proust employs idiosyncratic terminology, reserving the term "metacognition" for epistemic feelings and referring to propositional metacognition as "metarepresentation". This is unhelpful, as nonconceptual mental states are also representations.
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.
6. Gibson (1979), who coined the term "affordance", used it in an essentially behaviorist account of perception. Strawson offers a more general kind of non-propositional psychological explanation, which requires representation (see Bermúdez, 2007 for an argument to this effect).
7. There are methods of determining whether the informational inputs involved in a metacognitive task differ from those in first-order versions of the task. This has been confirmed for some perceptual tasks (Wokke et al., 2017). But such methods have not, to my knowledge, been extended to reality testing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

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

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

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

Bermúdez, J. L. (2007). *Thinking without words*. Oxford University Press.

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

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

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

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 mind*. 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

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

Cussins, A. (1992). Content, embodiment and objectivity: The theory of cognitive trails. *Mind*, *101*(404), 651–688. https://doi.org/10.1093/mind/101.404.651

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

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

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

Dijkstra, N., & Fleming, S. M. (2023). Subjective signal strength distinguishes reality from imagination. *Nature Communications*, *14*(1), 1627. https://doi.org/10.1038/s41467-023-37322-1

Dijkstra, N., Kok, P., & Fleming, S. M. (2022). Perceptual reality monitoring: Neural mechanisms dissociating imagination from reality. *Neuroscience & Biobehavioral Reviews*, *135*, 104557. https://doi.org/10.1016/j.neubiorev.2022.104557

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.

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.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

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

Fry, B. R., Russell, N., Gifford, R., Robles, C. F., Manning, C. E., Sawa, A., . . . 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

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

Greely, N. (2021). Epistemic feelings, metacognition, and the Lima problem. *Synthese*, *199* (3–4), 6803–6825. https://doi.org/10.1007/s11229-021-03094-8

Greely, N. (2023). *Nonconceptual metacognition* [Doctoral Dissertation, University of California San Diego]. Proquest dissertations and Theses Global.

Grush, R. (2000). Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind*, *1*(1), 59–92. https://doi.org/10.1023/A:1010039705798

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*(3), 377–396. https://doi.org/10.1017/S0140525X04000093

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

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. 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

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

Johnson, M. K., & Raye, C. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85. https://doi.org/10.1037/0033-295X.88.1.67

Kim, H., Koh, H.-Y., & Choi, D.-S. (2016). Impaired reality testing in mice lacking phospholipase Cβ1: Observed by persistent representation-mediated taste aversion. *Public Library of Science One*, *11*(1), e0146376. https://doi.org/10.1371/journal.pone.0146376

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

Lau, H. (2019). Consciousness, metacognition, & perceptual reality monitoring [Preprint]. PsyArXiv.

Lau, H., Michel, M., LeDoux, J. E., & Fleming, S. M. (2022). The mnemonic basis of subjective experience. *Nature Reviews Psychology*, *1*(8), 479–488. https://doi.org/10.1038/s44159-022-00068-6

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

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

Matthen, M. (2010). Two visual systems and the feeling of presence. In N. Gangopadhyay, M. Madary, & S. Finn (Eds.), *Perception, action, and consciousness* (pp. 107–124). Oxford University Press.

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

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.

Millikan, R. (1984). *Language, thought, and other biological categories*. MIT Press.

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

Nelson, T. (1992). *Metacognition: Core readings*. Allyn & Bacon.

Nelson, T., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 26, pp. 125–173). Elsevier.

O'Keefe, J. (1994). Cognitive maps, time, and causality. *Proceedings of the British Academy*, *83*, 35–45.

Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness* (First ed.). Oxford University Press.

Rahnev, D. (2021). Visual metacognition: Measures, models, and neural correlates. *American Psychologist*, *76*(9), 1445–1453. https://doi.org/10.1037/amp0000937

Schmack, K., Bosc, M., Ott, T., Sturgill, J. F., & Kepecs, A. (2021). Striatal dopamine mediates hallucination-like perception in mice. *Science*, *372*(6537), eabf4740. https://doi.org/10.1126/science.abf4740

Strawson, P. F. (1959). *Individuals: An essay in descriptive metaphysics*. Routledge.

Wokke, M. E., Cleeremans, A., & Ridderinkhof, K. R. (2017). Sure I'm sure: Prefrontal oscillations support metacognitive monitoring of decision making. *The Journal of Neuroscience*, *37*(4), 781–789. https://doi.org/10.1523/JNEUROSCI.1612-16.2016