Cognitive Neuropsychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pcgn20

The seductive allure of the brain: Dualism and lay perceptions of neuroscience

Gwendolyn Sandoboe & Iris Berent

To cite this article: Gwendolyn Sandoboe & Iris Berent (2021) The seductive allure of the brain: Dualism and lay perceptions of neuroscience, Cognitive Neuropsychology, 38:3, 205-230, DOI: 10.1080/02643294.2021.1976127

To link to this article: https://doi.org/10.1080/02643294.2021.1976127

COGNITIVE NEUROPSYCHOLOGY
2021, VOL. 38, NO. 3, 205–230
https://doi.org/10.1080/02643294.2021.1976127

Routledge
Taylor & Francis Group

The seductive allure of the brain: Dualism and lay perceptions of neuroscience

Gwendolyn Sandoboe and Iris Berent

Department of Psychology, Northeastern University

ABSTRACT
Laypeople prefer brain explanations of behavior (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008). We suggest that this preference arises from 'intuitive Dualism'. For the Dualist, mentalistic causation elicits a mind-body dissonance, as it suggests that the immaterial mind affects the body. Brain causation attributes behavior to the body, so it alleviates the dissonance, hence, preferred. We thus predict stronger brain preference for epistemic traits – those perceived as least material, even when no explanation is required. To test this prediction, participants diagnosed clinical conditions using matched brain- and behavioral tests. Experiments 1-2 showed that epistemic traits elicited stronger preference for brain tests. Experiment 3 confirmed that epistemic traits are perceived as immaterial. Experiment 4 showed that, the less material the trait seems, the stronger the surprise (possibly, dissonance) and brain preference. Results offer new insights into public perception of science, the role of intuitive Dualism, and the seductive allure of neuroscience.

ARTICLE HISTORY
Received 10 November 2020
Revised 27 June 2021
Accepted 30 August 2021

KEYWORDS
Seductive allure of neuroscience; dualism; causal reasoning; core knowledge; folk psychology

1. Introduction: The seductive allure of neuroscience

When laypeople evaluate explanations of behaviour, they tend to place undue weight on the brain. In the first systematic investigation of this phenomenon, Weisberg and colleagues (Weisberg et al., 2008) asked laypeople to rate their satisfaction with paired explanations of psychological phenomena (e.g., the "curse of knowledge"). One explanation detailed a cognitive mechanism (e.g., people mistakenly project their own knowledge onto others), without mentioning the brain; the matched neuroscience explanation was identical in all respects, except for the addition of irrelevant neuroscience details (e.g., "areas in the frontal lobe"). Participants rated neuroscience explanations to be more satisfying. In fact, laypeople preferred the explanations with neuroscience even when their logic was circular. Crucially, neuroscience experts (graduate, postdoctoral and faculty researchers) showed no such preference, confirming that the neuroscience explanations were no more informative. Accordingly, Weisberg et al. (2008) dubbed this phenomenon the "Seductive Allure of Neuroscience Explanations"; here we refer to it as the SAN effect.

Although the SAN effect has been replicated numerous times (Fernandez-Duque et al., 2015; Hopkins et al., 2016, 2019; Minahan & Siedlecki, 2016; Rhodes et al., 2014; Weisberg et al., 2015; Weisberg et al., 2018), the basis of this phenomenon remains unclear. In what follows, we review several existing explanations for the SAN effect. We next propose a novel account for this phenomenon and evaluate it in a series of experiments.

1.1 Some explanations of the seductive allure of neuroscience explanations

Why do people prefer brain-based explanations? One possibility is that people are simply captivated by conjuring the imagery of the brain. In line with this view, McCabe and Castel (2008) found that participants considered scientific articles more credible when the articles were accompanied by fMRI images compared to bar graphs, topographic maps, or no image. However, subsequent studies have failed to replicate this finding (Gruber & Dickerson, 2012; Hook & Farah, 2013; Michael et al., 2013; Schweitzer et al., 2013). It thus appears that the SAN effect does not merely arise from invoking the imagery of the brain.

A second set of explanations attribute the SAN effect to the complexity of neuroscience narratives. And indeed, in the SAN literature, some neuroscience

CONTACT Iris Berent i.berent@northestern.edu Department of Psychology, Northeastern University, 125 Nightingale Hall, 360 Huntington Ave. Boston MA 02115, United States ☎(617) 373 4033. (617) 373-8714
Supplemental data for this article can be accessed https://doi.org/10.1080/02643294.2021.1976127.

narratives are longer, and they introduce more complex technical jargon. But there is evidence that neither length nor complexity of jargon can explain the SAN effect. While participants do indeed consider longer explanations more satisfying (Weisberg et al., 2015), people prefer brain-based explanations even when brain explanations are matched to controls for length (Fernandez-Duque et al., 2015; Hopkins et al., 2016; Rhodes et al., 2014; Weisberg et al., 2015). Likewise, jargon is neither a necessary nor a sufficient condition for the SAN effect. The introduction of technical jargon (e.g., specific brain areas, fMRI) produces no preference above and beyond jargon-free explanations that simply invoke the brain, nor does the addition of jargon from fields other than neuroscience increase preference (Weisberg et al., 2015).

Complexity also does not seem to promote a preference because it distracts participants from logical inconsistencies or boosts their sense of comprehension. When presented with neuroscience information, participants were no more likely to ignore methodological flaws in research, nor did they report a stronger subjective sense of comprehension (Rhodes et al., 2014). Similarly, participants who showed a preference for neuroscience were no more likely to fall for other logical fallacies (as measured by cognitive reflection tests; Fernandez-Duque et al., 2015; Hopkins et al., 2016; Minahan & Siedlecki, 2016; Rhodes et al., 2014). Together, these results suggest that added complexity—be it length or technical jargon—does not fully explain the SAN effect.

On a third account, the SAN effect arises because readers possess prior beliefs on the specific topic at hand. When the narrative supports those beliefs, people might be less critical in their evaluation, and thus, less likely to detect its logical flaws (Rhodes et al., 2014). For example, a reader who believes that music improves studying may be more accepting of irrelevant neuroscience information in a narrative that advocate this view. This account would thus predict a stronger SAN effect only when participants' prior beliefs are supported by the article. However, this result was not found —there was no interaction between prior beliefs and the SAN effect (Rhodes et al., 2014).

A more promising explanation for the SAN effect appeals to a general preference for reductive explanations. Reductive explanations are ones that invoke information from a lower level of analysis. For example, a biological phenomenon, such as head bobbing behaviour in lizards during mating season, could be explained by appealing to chemistry: the lizards bob their heads due to heightened levels of testosterone in the bloodstream. Since neuroscience offers a lower-level analysis of behaviour, it is conceivable that the SAN effect could arise from a preference for reductive explanations in general. And indeed, laypeople demonstrably prefer reductive explanations: they favour physical explanations of chemistry, chemical explanations of biology, and biological explanations of neuroscience (Hopkins et al., 2016; Weisberg et al., 2018). Scientists show this preference as well (Hopkins et al., 2019). While the preference for brain-based explanations of behaviour is in line with reductionism, the preference for reductive psychological explanations (i.e., psychology —> neuroscience) is far greater than the preference for reduction at other levels of analysis (e.g., chemistry —> physics; Hopkins et al., 2016; Weisberg et al., 2018). In addition, when psychological explanations are reduced even further, by invoking biochemistry, physics, or mathematics, they are not preferred to the less reductive explanations that invoked neuroscience (Fernandez-Duque et al., 2015). Reductionism alone is thus insufficient to account for the SAN effect.

1.2 The role of Dualism

The discussion thus far suggests that people prefer psychological explanations that invoke the brain, and this preference is not fully explicable by complexity, past beliefs, or reductionism. To clarify, this is not to say that these explanations are inherently wrong; we do not wish to suggest that reductionism or simplicity are not virtues in scientific explanations. Instead, we submit that these explanations are *insufficient* to account for the SAN effect. The resulting question, then, is why the framing of psychological phenomena in reference to the brain acquires such special significance for laypeople. Intuitive Dualism presents an explanation for this preference.

Intuitive Dualism is the tacit belief that the mind is immaterial and separate from the material body (Bloom, 2014). A large literature shows that people indeed segregate bodies and minds. For example, when people are invited to imagine situations that perfectly duplicate a person's body, they intuit that physical (i.e., material) traits (e.g., a scar or a birthmark) are more likely to transfer to the duplicate than mental traits (e.g., a person's knowledge of their own name; Forstmann & Burgmer, 2015; Hood et al., 2012). Conversely, when people consider situations that transfer only the mind, such as in pre-life (Emmons & Kelemen, 2014), after-life (Bering, 2002; Bering & Bjorklund, 2004), or in mind switching scenarios (Cohen & Barrett, 2008; Hood et al., 2012), they conclude that mental traits are now the ones that are more likely to transfer.

This dissociation between mental and physical traits has been documented cross-culturally (Chudek et al., 2018; Cohen et al., 2011), even in a "mind opacity" soceity—a society whose members do not discuss or otherwise publicly advertise their mental states (Chudek et al., 2018; for discussion of mind opacity: Robbins & Rumsey, 2008). These results suggest that the belief that mind "stuff" is ethereal, distinct from the material body, is not strictly a Western phenomenon.

Indeed, the precursors of intuitive Dualism are seen in young infants. Young infants invoke different principles in reasoning about objects and agents. Infants (including newborns) seem to expect objects to interact according to principles of intuitive physics (Mascalzoni et al., 2013; Spelke et al., 1992; Spelke & Kinzler, 2007; Valenza & Bulf, 2011). For example, when newborns see one ball (A) launch towards another ball (B), they expect Ball B to launch only if it is immediately contacted by Ball A; if there is no immediate contact, infants are demonstrably surprised (Mascalzoni et al., 2013). Infants likewise expect objects to have continuity; once an object is launched, it will move on a connected path, rather than proceed in "jumps" (Spelke, Kestenbaum, et al., 1995). Infants, however, do not apply the same continuity requirement for agents (Kuhlmeier et al., 2004), nor do they expect agents to move only by contact (Spelke, Phillips, et al., 1995). Instead, infants seem to ascribe the actions of agents to their knowledge (Onishi & Baillargeon, 2005) and goals (Woodward, 1998, 2009). Thus, not only do humans seem to naturally contrast the "stuff of mind" and matter, but they further apply different causal principles in reasoning about these two substances. Infants interpret the causal interactions between objects by applying the laws of intuitive physics, but when they reason about agents, they suspend intuitive physics, and instead, appeal to theory of mind (Leslie, 1987; Spelke & Kinzler, 2007).[1]

Could intuitive Dualism, then, explain the allure of neuroscience? Weisberg et al. (2008) indeed considered this possibility, as Dualism "may illustrate a connection between the mind and the brain that people implicitly believe not to exist, or not to exist in such a strong way" (p.8). Similarly, Bloom (2014) noted that, when laypeople are captivated by neuroimaging findings, they "often seem fascinated by the mere fact that the brain is involved at all" (p.201). And according to Hook and Farah (2013), people are *surprised* by the correspondence between mental states (e.g., love) and the brain (p.1398). But the precise contribution of intuitive Dualism to the SAN effect remains elusive.

To assess the role of intuitive Dualism, past research invited participants to indicate their agreement with statements such as "the mind and brain are separate" (Fernandez-Duque et al., 2015; Hook & Farah, 2013). Results yielded no systematic association between these ratings and participants' preference for neuroscience explanations (Fernandez-Duque et al., 2015; Hook & Farah, 2013). While these results would seem to suggest that intuitive Dualism plays no role, this conclusion may not necessarily be warranted.

One source of concern is methodological. Intuitive Dualism, by hypothesis, is a tacit bias in human reasoning, whereas Dualist scales evaluate participants' explicit attitudes towards bodies and minds. These two manifestations of intuitive Dualism—the implicit and explicit—need not align. Indeed, many western adults know too well that cognition corresponds to brain activity, and when asked to respond to Dualist scales, their explicit ratings are typically neutral or leaning towards physicalism (Riekki et al., 2013; Stanovich, 1989). Yet past research has shown strong evidence for tacit Dualism in individuals who vehemently argue just the contrary (Bering, 2002). And as noted, evidence for implicit Dualism has been detected in adult members of societies that practice mind opacity (Chudek et al., 2018) and in children (Bering & Bjorklund, 2004; Chudek et al., 2018; Cohen et al., 2011; Emmons & Kelemen, 2014; Hood et al., 2012). Furthermore, Dualist scales assess how Dualist attitudes vary across individuals. But if intuitive Dualist attitudes are robust and universal (akin to having five fingers), then inter-personal variability would be minimal. If so, the absence of correlation between Dualism and behaviour would not necessarily show that Dualism plays no role in reasoning. For these reasons, the utility of explicit Dualist scales may be limited, methodologically; implicit methods are likely to offer a more sensitive gauge of the role of intuitive Dualist thinking.

Another obstacle to assess the role of intuitive Dualism is theoretical. To evaluate the contribution of intuitive Dualism to the SAN, one must detail how the mind/body divide promotes a preference for brain explanation. Existing proposals, however, do not specify this link. While it seems intuitively likely that laypeople are surprised by the involvement of the brain in psychological processes, it is unclear *why* this surprise would lead them to prefer a brain explanation that is superfluous or even flawed. Certainly, people do not invariably find all surprising explanations credible. For example, the heliocentric model of the solar system is arguably surprising, but it is an explanation that people have strongly resisted (Koestler & Butterfield, 1989). The resulting question, then, is how intuitive Dualism leads to the SAN effect. In what follows, we articulate such an account, and evaluate it in a manner that does not depend on the explicit endorsement of Dualism.

Our proposal further predicts that the preference for brain causes should arise generally, irrespective of whether people explicitly engage in explanations, or whether they merely engage in causal reasoning. Our experiments thus evaluate the role of brain causes in causal reasoning.

1.3 Could the SAN effect arise from a Dualist dissonance?

The present research is designed to evaluate a novel account for the relationship between intuitive Dualism and SAN (Berent, 2020). In this view, people prefer brain-based explanations because invoking the brain alleviates a cognitive dissonance, caused by intuitive Dualism. To explain why such a dissonance arises, we first ought to take a closer look at Dualism.

Earlier, we stated that intuitive Dualism is committed to the presumption that the "stuff of the mind" is distinct from matter—the stuff of bodies. In addition, when people (including young infants) reason about interactions among physical objects, they further invoke different ***causal*** principles relative to those they summon to explain the behaviour of agents. Young infants, recall, expect objects to follow the principles of intuitive physics (e.g., contact), whereas agents seem to be driven by their mental states (e.g., Hamlin et al., 2013; Kuhlmeier et al., 2004; Onishi & Baillargeon, 2005; Spelke et al., 1992; Woodward, 1998).

These distinct sets of intuitive causal principles continue to guide causal reasoning throughout life. Intuitive physics is known to guide adults' physical understanding, and indeed, it wreaks havoc in scientific physical reasoning (e.g., Mccloskey, 1983). Similarly, theory of mind continues to guide our intuitive reasoning about quotidian behaviour (Leslie et al., 2004). For example, we intuit that one might decide to extend one's arm towards the coffee mug because one believes there is still some coffee left inside. It is one's *belief* (a mental state) that causes one's decision (another mental state) to move one's arm.

But this neat Dualist division between mind and matter immediately runs into a major problem. The problem arises because, per theory of mind, mental states can cause ***behaviour*** to happen. It is one's belief about the contents of one's coffee mug (and one's desire for coffee) that caused one's arm to move (hereafter, *mentalistic* causation of behaviour). And per intuitive Dualism, one's arm is very much part of one's material body—it is a physical object, much like the launching balls.

Per intuitive physics, however, such interactions between mind and matter are impossible. Intuitive physics, recall, mandates that one physical object can only move by contact with another physical object. But per intuitive Dualism, mental states like "belief" and "desire" are immaterial. Mind, then, cannot possibly cause movement of a physical object, such as one's arm. Under Dualist reasoning, such mentalistic explanations should thus trigger a pressing mind–body dissonance (for illustration, see Figure 1a).

(1) The Dualist dissonance

  a. Per intuitive Dualism, the mind is immaterial, distinct from the material body.
  b. Per intuitive physics, one material object can only move by contact with another material object.
  c. Per theory of mind, mental states can cause the (material) body to move (i.e., behaviour), contrary to intuitive physics.
  d. The attribution of behaviour to mentalistic causes gives rise to a mind–body dissonance.

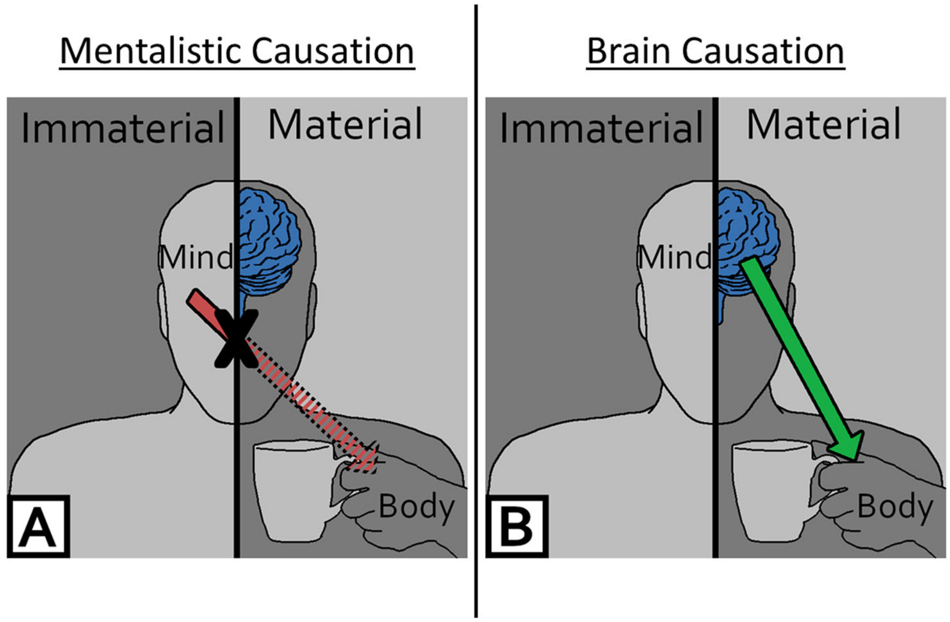

Figure 1. Visualization of the Dualist dissonance and its resolution by invoking brain causes of behaviour. (A) Per theory of mind, laypeople attribute behaviour (grabbing the mug) to mental states (e.g., the desire for coffee). Since, per Dualism, the mind is distinct from the body, and since (per intuitive physics), physical objects can only interact by contact, such mind-body interactions are problematic—they elicit a mind-body dissonance. By attributing the cause of behaviour to the brain (B), one situates the cause of behaviour within the body, so the Dualist dissonance is alleviated. For this reason, brain causes of behaviour are preferred.

In everyday situations, this dissonance is typically moot. This is because our theory of mind routinely attributes people's behaviour to mental states (e.g., we interpret people's actions in terms of their goals). In addition, people do not spontaneously entertain the possibility that behaviour arises from brain states. Since people crave to unveil the causes of events, including behaviour (Keil, 2006; Lombrozo, 2006), and since mentalistic, psychological causes of behaviour are not only pervasive but they are also the *only* internal causes of behaviour that are available to us intuitively, the dissonance they elicit is typically ignored.

Brain-based explanations, however, change this status quo. First, by presenting a choice between mentalistic and brain causes, brain-based explanations highlight the Dualist dissonance latent in the mentalistic view. Second, by suggesting an alternative cause of behaviour (i.e., *brain-causation*), brain-based explanations now offer a "way out" of this dissonance. If it is the material brain that commands the material hand to move, then it is now matter that causes movement in matter, and the mind–body dissonance is solved (see Figure 1b).[2]

By presenting a material cause of behaviour, then, brain-causation alleviates the mind–body dissonance inherent in mentalistic causation. Consequently, under intuitive Dualism, brain-based explanations are more satisfying, even if they are no more scientifically informative than the mentalistic explanations. It is the resolution of the Dualist dissonance that promotes the SAN effect.

This account makes two major testable predictions. First, we expect the seductive allure of neuroscience (SAN) to be a broader characteristic of causal reasoning that is not limited to explanations, specifically. Second, we expect the SAN effect to be further modulated by the type of psychological trait that is being explained—sensorimotor vs. epistemic. Below, we explain both predictions.

1.3.1 *The seductive allure of neuroscience shapes causal reasoning*

If the dissonance account is correct, then the seductive allure of the brain may be far broader than what had been previously assumed—a seductive allure of neuroscience, as opposed to strictly a seductive allure of neuroscience *explanations*. This is because, by this proposal, the SAN arises from the attribution of behaviour to mentalistic causes, and people infer such mentalistic causes spontaneously, even without engaging in explicit explanations; indeed, mentalistic causation seems to arise even in preverbal infants (e.g., Hamlin et al., 2007; Kiley Hamlin et al., 2013; Kuhlmeier et al., 2003). Since mentalistic causation is spontaneous, and it promotes the SAN (via the Dualist dissonance), we thus expect the SAN to arise in *all* forms reasoning about the causes of behaviour; the preference for neuroscience explanations is merely a special case.

This preference arises because people monitor the structure of their causal reasoning. Past research suggests that people spontaneously consider events that are linked to a cause (e.g., the bottle fell off the shelf) as more plausible relative to events that merely describe an attribute (e.g., the bottle is pretty; Connell, 2004; Connell & Keane, 2004; Emmons & Kelemen, 2004). We suggest that a similar evaluation takes place when people consider the causes of behaviour. Since brain causes alleviate the Dualist dissonance, we expect people to consider causal chains that invoke brain causes as better-formed, hence, preferred.

This preference can acquire multiple forms. When people engage in explanation (narrowly), the preference for brain causes would lead participants to consider brain explanations as more satisfying. Similarly, when the task merely promotes causal reasoning about an event/narrative, here, too, brain causation ought to be better-formed, so we expect people to consider such events/narratives as more plausible. For example, when participants are presented with a patient who suffers from a clinical condition, we expect people to consider the diagnosis more plausible when it is given by a brain- (relative to a behavioural) test. Similarly, participants should consider a brain test as a more appropriate means to diagnose the patient's condition (relative to a matched behavioural test). We summarize this logic in (2).

(2) The seductive allure of neuroscience (SAN) in causal reasoning
  a. When people reason about human behaviour, they tacitly evaluate the well-formedness of their reasoning.
  b. Well-formedness of causal reasoning is determined, in part, by intuitive Dualism: mind and matter cannot causally interact.
  c. Mentalistic causes violate intuitive Dualism (see (1)), so these causal chains are ill-formed relative to ones that attribute behaviour to brain causes. Consequently,
    i Brain explanations of behaviour should be considered more satisfying (relative to ones that invoke mentalistic causes).
    ii Scenarios that attribute behaviour to brain causes should be considered more plausible.

1.3.2 *The seductive allure of neuroscience should be stronger for epistemic traits*

A second prediction of our proposal has to do with which psychological traits (e.g., seeing, knowing) are most vulnerable to SAN. People typically consider psychological traits as causes of behaviour. For example, a behaviour, such as moving one's arm towards a cup, could be caused by psychological traits such as seeing the cup or knowing that the cup is on the table. And since these psychological traits are typically aligned with the mind, their effect on the body (behaviour) ought to elicit a dissonance, and in so doing, trigger the SAN effect—the stronger the dissonance, the stronger the SAN effect.

The strength of the dissonance, in turn, should depend on the perceived anchoring of psychological traits (the causes of behaviour) in the material body: the psychological traits that are perceived as most ethereal (i.e., as disembodied) are ones that should elicit the strongest mind–body dissonance.

Indeed, there are reasons to expect that psychological traits could differ in that regard. Sensory and motor capacities, like hearing, seeing, walking, and grasping, can be readily associated with specific body parts, so one would expect sensory and motor traits to be perceived as relatively material. But for abstract epistemic states, such as the knowledge of one's name, or the syntactic structure of language, the link to the material body is elusive.

In line with this possibility, past research has shown that epistemic states are thought to be more likely to transfer to the afterlife (Bering, 2002; Bering & Bjorklund, 2004) and to the minds of others (Chudek et al., 2018; Cohen et al., 2011). Moreover, when epistemic and sensorimotor traits were directly contrasted, people rated epistemic traits as reliably less likely to transfer to a body replica or be localized in the brain compared to sensory and motor traits, whereas sensory and motor traits were rated as less likely to transfer to the afterlife (Berent et al., in press). These results reflect tacit Dualist beliefs that sensorimotor traits are more materially embodied than abstract knowledge.

So, when one recognizes that (material) sensorimotor traits (e.g., the capacity to move one's arm) can effect changes in matter (e.g., the lifting of a cup by one's arm), this effect of matter on matter is unremarkable, as it is consistent with intuitive physics—the mind–body dissonance is minimal. But when such material changes (e.g., the moving of one's body) arise from immaterial epistemic states (e.g., knowledge that there is a cup in the cupboard), intuitive Dualism should now cause one to experience a sharp dissonance.

This analysis of the SAN effect generates clear, testable predictions. First, if the SAN effect arises from a mind–body dissonance, then epistemic states should elicit a stronger dissonance, hence, a stronger SAN effect relative to sensorimotor traits. Second, the magnitude of the SAN effect should depend on perceived materiality of these traits. And finally, as noted (in the previous section), the SAN should arise even when the task does not require participants to engage in explanation, specifically. The present research tests these predictions.

1.4. The present research

To test the prediction that the seductive allure of neuroscience (SAN) is a property of causal reasoning, broadly (as opposed to explanations, specifically), in the following experiments, we gauge the SAN effect implicitly, without asking people to explicitly evaluate brain- and mentalistic explanations. Participants were invited to help a clinician choose between two methods of diagnosis for a psychological disorder—a brain test and a closely matched behavioural test. A preference for the brain test (relative to a behavioural test matched for its informative value) would demonstrate that people are seduced by the allure of neuroscience (i.e., a SAN effect) in causal reasoning.

To determine whether the SAN effect further depends on the type of psychological traits, we compared participants' preference for brain- and behavioural-based tests for two types of symptoms: epistemic and sensorimotor. Epistemic symptoms were designed to reflect the manipulation of mental representations. For example, theory of mind was considered an epistemic symptom because it invokes reasoning about another person's mental states. Sensorimotor symptoms, by contrast, directly appealed to sensory organs (e.g., auditory hyper-sensitivity) and bodily movements (e.g., motor coordination).

To test for the SAN effect, participants were presented with four pairs of vignettes, featuring four types of psychological disorders (neurodevelopmental, language, reading, and age-related degenerative disorders). Within each disorder pair, one vignette described a symptom related to an epistemic state (hereafter, an epistemic symptom), whereas its matched counterpart featured a sensorimotor symptom. For example, participants were presented with two symptoms of autism—an impairment in theory of mind (an epistemic symptom) and auditory hypersensitivity (a sensory symptom). Theory of mind is considered an epistemic state because it captures one's knowledge of another person's mental states. Auditory sensitivity, in contrast, patently corresponds to a bodily sensation, namely, hearing.

Each such symptom was evaluated by two matched tests— a brain test and a behavioural test. The behavioural test included methods such as looking time or response time, whereas the brain test gauged the brain's electric activity in a manner akin to EEG. Participants were told which test outcome is expected for typical individuals and which one would be indicative of a disorder. In each case, then, the test results only indicated whether participants' responses were typical or atypical—the brain test offered no additional information, such as the affected brain region or its connectivity.

With this information in mind, participants were asked to indicate their preference for the two tests. In Experiment 1, participants indicated which test provides a better screening for the disorder. In Experiment 2, participants were asked to rate how likely is it that an atypical test result indicates a disorder (using a 1–7 scale); participants made this judgment for both the behavioural and brain test.

Given the information presented to participants, both tests suggested a disorder (as participants were explicitly told that the pattern of results in each case was abnormal), and the two tests were strictly matched for their diagnostic value.

Accordingly, on scientific grounds, a positive outcome on the brain test is just as likely to suggest a disorder as a positive outcome on the behavioural test.

From the perspective of intuitive Dualism, however, these two tests differ markedly. The brain test explicitly references the body, as it measures changes in the (material) brain. The behavioural test, by contrast, does not explicitly invoke the body. And since, per theory of mind, people routinely interpret behaviour in mentalistic terms, the outcome of the behavioural test could thus be attributed to the mind- rather than the body. Indeed, past research using these same manipulations has consistently shown that people are more likely to associate the outcomes of the brain test with one's bodily essence relative to the outcomes of a matched behavioural test (Berent et al., 2020; Berent et al., in press; Berent & Platt, 2021). The behaviour test, then, elicits a Dualist dissonance, whereas the brain test alleviates this pressure (see (3)). Accordingly, we expected that participants should prefer the brain test over the behavioural test.

(3) How invoking brain causes of behaviour alleviates the Dualist dissonance:
  a. Mentalistic causes of behaviour elicit a Dualist dissonance.
    i Per theory of mind, the behavioural test suggests a mentalistic cause.
    ii Mentalistic causes of behaviour (changes in the body) elicit mind–body dissonance (see (1)).
  b. Brain causes attribute behaviour (changes in the body) to the material body (the brain).
  c. Brain causes alleviate the Dualist dissonance.

The Dualist dissonance account, however, further predicts that this preference for the brain test should be modulated by the type of symptom—epistemic vs. sensorimotor. As noted, laypeople perceive epistemic traits as disembodied (relative to sensorimotor traits; Berent et al., in press). We thus predict that epistemic symptoms (e.g., theory of mind) should elicit a stronger Dualist dissonance compared to the sensorimotor symptoms (e.g., auditory hypersensitivity), which, in turn, should result in a stronger preference for the brain test.

To further explain this prediction, consider again the above-mentioned symptoms of autism—auditory hypersensitivity and theory of mind (sensory and epistemic symptoms, respectively). The two vignettes make it clear that the epistemic and sensory symptoms have direct consequence for behaviour. For example, the failure of a patient with autism to appreciate another person's mental (epistemic) states should lead them to incorrectly predict where this person might look for a missing object (behaviour).

For the Dualist, such effects of epistemic states on behaviour present a sharp dissonance, as they demonstrate that the immaterial mind can effect change in the material body. The brain test alleviates the dissonance because it patently demonstrates that the abnormality in autism arises not from ethereal mind but from the material body (the pattern of EEG activity). Consequently, for such epistemic symptoms, we expect people to strongly prefer the brain test to the behavioural test (where no solution to the dissonance is presented). By contrast, for the sensorimotor symptom (e.g., auditory hypersensitivity), the dissonance is weaker, since the cause of the symptom is perceived to reside in the material body. So here, the preference for the brain test should correspondingly diminish.

(4) Why epistemic traits should elicit a stronger SAN effect.
  a. The SAN effect arises from a Dualist dissonance.
    i Brain tests confirm that the cause of the behaviour/symptoms lies within the body.
    ii People prefer to attribute behaviour to brain causes because in so doing, they alleviate the Dualist dissonance (in (1)).
    iii The magnitude of the SAN preference depends on the magnitude of the Dualist dissonance—the stronger the dissonance, the stronger the SAN effect.
  b. Epistemic traits elicit the strongest mind–body dissonance.
    i. The magnitude of the Dualist dissonance depends on the perceived immateriality of psychological traits—the less material a trait is, the more puzzling its effect on the material body (behaviour), hence, the stronger the dissonance.

   ii. Epistemic traits are perceived as immaterial, distinct from the body (Berent et al., in press)
   iii. Epistemic traits elicit the strongest Dualist dissonance.
 c. The preference for brain tests is strongest for epistemic traits.

Summarizing, then, if the preference for the brain test results from its capacity to alleviate the Dualist dissonance, then people should show stronger preference for the brain test for epistemic symptoms (where the dissonance is maximized) compared to sensorimotor ones (see (4)). And since the attribution of behaviour to brain causes should be considered better formed (as it alleviates the Dualist dissonance), we expect people to consider the entire diagnostic scenario more plausible (see 2c). Thus, when participants are informed that the patient has a clinical condition, people should consider the brain test as a better (i.e., more plausible) diagnostic test for that condition. Similarly, people should consider it more plausible that a positive diagnosis revealed via a brain test corresponds to the actual underlying condition, relative to a positive diagnosis detected by a behavioural test. Experiments 1 and 2 test these two predictions, respectively.

Experiments 3 and 4 were designed to test two auxiliary assumptions of the dissonance account. In this proposal, epistemic symptoms elicit a stronger SAN effect because they are perceived as relatively immaterial, so their demonstrable effect on behaviour elicits a stronger dissonance. Experiment 3 tested whether our epistemic symptoms were indeed viewed as less material than the sensorimotor ones. Experiment 4 gauged the dissonance elicited by epistemic and sensorimotor symptoms. We predict that epistemic symptoms should elicit a stronger dissonance, and this dissonance should depend on the perceived materiality of the symptom.

2. Part 1: Do laypeople prefer neuroscience more when reasoning about epistemic traits?

Experiments 1 and 2 each presented participants with a series of vignettes featuring epistemic and sensorimotor symptoms of a psychological disorder (e.g., for autism: theory of mind impairment vs. auditory hypersensitivity). In each case, the description of the symptom was followed by two tests for its presence, one behavioural (e.g., looking time) and one brain-based (e.g., spikes in brain response). Participants were asked to choose/rate the two tests and provide a brief explanation for their response. We hypothesized that laypeople would prefer brain tests to behavioural tests overall, and that the preference would be stronger when reasoning about epistemic symptoms relative to sensorimotor symptoms.

In Experiment 1, participants responded to a forced choice question ("which method is better"). By contrasting the two tests, we sought to determine whether participants will prefer the brain test even when the experimental task does not require that they evaluate a brain explanation explicitly. In Experiment 2, participants rated each test individually, on a 1–7 scale. Since this experiment does not elicit a comparison of the two tests, it allows us to determine whether the preference for the brain test emerges spontaneously. Thus, this second experiment presents a stronger test of the SAN effect. Our question, however, is not merely whether the SAN effect will emerge, but whether it depends on the symptom in question—epistemic or sensorimotor. Our dissonance account predicts that the preference for the brain test should be stronger for epistemic symptoms compared to sensorimotor symptoms.

Our account attributes this pattern to intuitive Dualism—a tacit bias that is potentially distinct from people's explicit beliefs about the mind/body divide. To address this distinction, and to compare our results with past research where Dualism was gauged explicitly (Fernandez-Duque et al., 2015; Hook & Farah, 2013), our experiments also assessed participants' explicit endorsement of Dualism. Since we hypothesize that SAN is caused by an intuitive Dualist bias, we do not predict a correlation between preference for the brain test and this explicit measure of Dualism. By contrast, we do expect the preference for the brain test to be linked to tacit Dualist beliefs, as gauged by the perceived materiality of the psychological symptoms in question. This latter prediction is evaluated in Part 2 (in Experiments 3-4).

2.1 Experiment 1: Preference for brain vs. behavioural test (forced choice)

2.1.1 Participants

48 Northeastern University students took part in this experiment in partial fulfillment of a course credit.

Most participants reported taking no advanced-level courses in psychology (87%), neuroscience (98%), or biology (69%). Sample size was determined based on a pilot study of Experiment 2 (rating responses), using an a-priori required sample size power analysis in GPOWER (Erdfelder et al., 1996). Results showed that a sample size of at least 12 participants would be required to obtain the moderate effect size for the hypothesized Symptom X Test interaction ($\eta^2_{partial}$ = .14) at a power of 0.8 and an alpha level of .05. We thus set the sample size for $N = 48$. This sample size was adopted in Experiments 1, 2 and 4; sample size for Experiment 3 was calculated separately and is described wherein.

Experiments 1–4 each employed a unique group of participants. All participants signed an informed consent, and the research protocol was approved by the IRB at Northeastern University.

2.1.2 Materials

The materials included four pairs of vignettes, describing symptoms of four types of psychological disorders: neurodevelopmental (Autism), language (Aphasias), reading (Alexia), and age-related degenerative (Alzheimer's, Parkinson's). Each vignette pair featured two symptoms of the disorder, one that reflected the manipulation of mental representations ("epistemic symptoms"), and one that involved the movements or sensations of the body ("sensorimotor symptoms"). The full list of vignettes is provided in Appendix A.

All vignettes had the same three-part format (see Figure 2). The first section describes the disorder and its symptoms—either epistemic or sensorimotor. The next section introduces a screening procedure, designed to test for the presence of the symptom. For example, theory of mind (an epistemic symptom of autism) is evaluated by a false belief task. The test features a character whose car keys are misplaced, unbeknownst to her. The viewer (i.e., patient) is asked to predict where the character will look for her keys. This section ends with a rationale, stating why the clinician believes this task will identify patients with the symptom, and thus the disorder. For example, since people with autism have difficulties reasoning about the minds of others, they are expected to be surprised to see the character search where she had left the keys, rather than where the viewer knows them to be.

The final section describes the two tests to measure the symptom, one behavioural and one brain based. The brain tests described recording brain spikes, akin to an EEG, while the behavioural tests describe a behavioural measure, such as looking time or response time. Critically, both tests evaluate the same construct. For example, in the

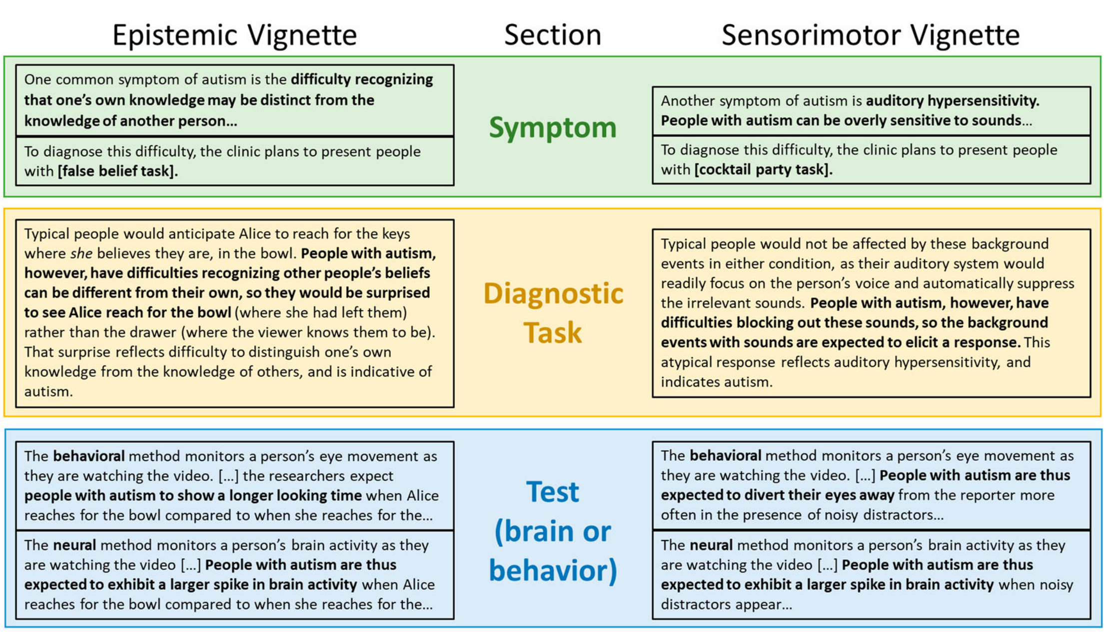

Figure 2. Overview of the vignette structure.

theory of mind vignette, both the brain and the behavioural test evaluate the patient's surprise (upon seeing the character search for her keys at the unexpected location). The behavioural test gauges surprise by monitoring the patient's eye movement at the display and participants are informed that "long looking times indicate surprise". The brain test gauges surprise by monitoring the patient's brain activity. Here participants are told that "there is a characteristic increase in the brain's activity (a "spike"), when a person is surprised".

For each test, participants were further informed what is the expected test outcome for typical individuals (e.g., longer looking time when the character searches in the new location) and what outcome is expected for people who suffer from the disorder (e.g., longer looking time when the character searches in the original location). The end of this section asks participants "Which method do you think would be best to screen people for [disorder]?"

The tasks featured in our vignettes were based on ones described in published scientific research (ERPs in theory of mind tasks: Sabbagh & Taylor, 2000; auditory ERPs: Pilling, 2009; phonological processing in dyslexia: Russeler et al., 2007; visual filtering in dyslexia: Roach & Hogben, 2007; ERPs in sentence processing: Thornhill & Van Petten, 2012; ERPs in response to unpredictable auditory stimuli: Squires et al., 1975; ERPs in semantic memory tasks: Castaneda et al., 1997; motor preparation deficits in Parkinson's: Fearon et al., 2017). However, the methods described in our vignettes were simplified to be more easily understood by a lay audience. For instance, instead of evaluating the distinctive characteristics of an ERP recording, the vignettes merely referred to a "spike in brain activity". Correspondingly, the matched behavioural test did not detail the complex dynamics of eye-movement. Instead, the behavioural test merely predicted that people with the disorder should exhibit longer (or shorter) "looking time". Additionally, our descriptions of the disorders altered some of the tasks reported in the literature so as to match the two tests for their outcomes. In particular, we designed the tests such that, for each symptom, both tests (behavioural and brain) identify the disorder by either a positive outcome (e.g., the presence of a spike/longer looking time) or a null outcome (e.g., the absence of a spike/no difference in looking time).

At the end of the experiment, participants were given two measures of explicit Dualist beliefs. The first, adapted from Forstmann et al. (2012), presented participants with seven diagrams representing the distance between the mind and the body as the distance between two circles. Participants were asked to "select the diagram that [they] believe best represents the relationship between the mind and the body". Participants responded to the "circle" Dualism question using a 1–7 scale representing the perceived distance between body and mind, with 7 being the most distant.

A second explicit test of Dualism (adapted from Fernandez-Duque et al., 2015) asked participants to express their explicit beliefs regarding the mind/body divide via a 1–5 scale. Items included Dualist statements such as "Each of us has a soul that is separate from the body" and physicalist statements such as "All of my conscious experience is the result of activity in my nervous system" (reverse-coded). Altogether, each participant responded to seven questions, and the mean was treated as their "questionnaire Dualism score". All Dualism items are presented in Appendix E.

2.1.3 Procedure

Participants were tested in the lab individually. After signing an informed written consent, each participant was seated at a computer and given the experiment via a Qualtrics survey. The survey first introduced the diagnostic scenario "we ask you to imagine that you are advising a clinic on the best methods to screen people who are suspected of having psychological disorders". Participants were then instructed to "carefully read the descriptions of the disorder in question, and then evaluate the effectiveness of each of the tests the clinic is considering." Participants indicated their response as a forced choice between the two test methods (Behaviour/Brain) and provided a brief explanation for their response. Each vignette appeared on a separate page, and participants could not revisit previously viewed vignettes. After responding to all eight vignettes, participants answered questions concerning their demographics and education. Finally, participants were given the "circle", followed by the "questionnaire" tests of their explicit Dualist beliefs. The entire experiment was completed in less than 30 min, on average.

2.1.4 Design

Each participant received all four vignette pairs in a 4 Disorder (Autism, Language, Reading, Aging) x 2 Symptom (Epistemic, Sensorimotor) within-subjects design. The order of the vignette pairs and the relative order of the two symptom vignettes were counterbalanced between subjects in a Latin square design. The order of appearance of the two tests within the vignette was also counterbalanced, between subjects.

2.1.5 Results and discussion

Experiment 1 collected three measures. Our primary interest is in participants' choice among the two tests. We also examined participants' written justifications of their responses, as well as their responses to the explicit Dualism scales. We analyse each of these three measures in turn.

a. *Test choice*. An inspection of the mean choice (see Figure 3) suggests that participants were more likely to choose the brain test for epistemic symptoms (0.63) than for sensorimotor symptoms (0.53).

To evaluate this observation, we submitted the results to a Generalized Linear Mixed Model (GLMM), with Symptom (epistemic/sensorimotor) as a fixed effect and Subject as a random effect. Disorder was excluded, as adding this factor to the model yielded a singular fit. There was a significant main effect of Symptom ($Z = -2.10$, $p = .04$, $OR = 1.57$), confirming that participants were more likely to prefer the brain test for epistemic symptoms compared to sensorimotor symptoms.

To further determine whether participants preferred the brain test, we next contrasted their mean choice against chance (.50) using a GLMMs with Subject as a random effect. These analyses were conducted separately for the epistemic and the sensorimotor symptoms. Results showed that the preference for the brain test was significantly higher than chance only for epistemic symptoms ($M = .63$, $Z = 3.46$, $p < .001$, $OR = 1.71$), but not for sensorimotor ones (proportion = .53, $Z = .72$, $p = .47$, $OR = 1.11$).

Thus, when participants were invited to diagnose a disorder, they reliably preferred a brain test over a behavioural test—a result in line with preference for brain-based explanations found in previous studies. Our results, however, show that this preference was reliable only when participants reasoned about epistemic, rather than sensorimotor symptoms. Additionally, epistemic symptoms yielded a reliably higher preference for the brain test.

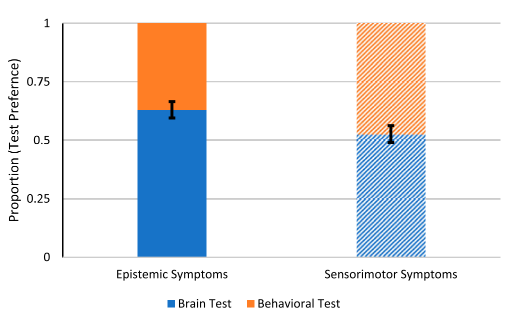

Figure 3. The choice proportion of the brain test for epistemic and sensorimotor symptoms. (Error bars represent standard errors of the proportions).

b. *Written justification*. We next examined participants' written justification of their choices. Of the total of 384 responses, 299 explicitly referenced a test (either the brain test, the behaviour test or both), either positively (e.g., *monitoring the activity in the brain seems to be a better way to truly determine if there is impairment*), or negatively (e.g., *autism occurs on a spectrum, so just measuring spikes may not account for less severe degrees of autism*).

An inspection of the mean proportion of positive responses (see Table 1) suggested that participants were most likely to comment on the test positively when a brain test was paired with epistemic symptoms. To evaluate the effects of Test and Symptom, we submitted the positive responses to a 2 Test x 2 Symptom generalized linear mixed effect model (GLMM). We found a significant effect of Test (Wald $\chi^2 = 7.65$, $p = .01$, $\varphi = .28$) as well as a reliable Test x Symptom interaction (Wald $\chi^2 = 7.65$, $p = .01$, $\varphi = .28$). The effect of Symptom was not significant (Wald $\chi^2 = .58$, $p = .45$).

Table 1. Participants' mean positive and negative comments on the brain and behavioural tests.

			Symptom	
Comment Type			Epistemic	Sensorimotor
Positive	Test	Brain	0.40	0.33
		Behavioural	0.22	0.33
Negative	Test	Brain	0.10	0.11
		Behavioural	0.29	0.27

To interpret the interaction, we next evaluated the effect of Test for each of the two symptoms. For epistemic symptoms, positive comments were more frequent for the brain- relative to the behavioural test (Wald $\chi^2 = 14.57$, $p < .001$, $\varphi = .74$); this, however, was not the case for the sensorimotor symptoms ($\Delta = 0$). Thus, not only were participants more likely to choose the brain test for epistemic symptoms, but they also expressed this preference in their positive comments.

Negative comments were overall less frequent (see Table 1). A 2 Test x 2 Symptom GLMM model only found a significant effect of Test (Wald $\chi^2 = 36.07$, $p < .001$, $\varphi = 1.30$), as participants were more likely to comment negatively on the behavioural- relative to the brain test. The effect of Symptom (Wald $\chi^2 < .001$, $p = .98$) and the interaction (Wald $\chi^2 = .28$, $p = .60$) were not significant.

c. *Explicit measures of Dualism.* The choice responses and (positive) comments considered thus far suggests that the preference for the brain test was stronger for epistemic symptoms. As noted, this preference could reflect participants' implicit bias towards Dualism, a position that is distinct from their explicit endorsement of the mind body divide. And indeed, when participants were asked to explicitly reason about the mind/body divide, their endorsement of Dualism was relatively low.

For the "circle" question, participants' average score was 2.06 (on a 1–7 scale). Similarly, their mean response on the Dualism "questionnaire" (M = 2.82) was not significantly different from 3, the value of the scale's "neutral" midpoint ($t = -1.38$, $p = .09$, $d = -.28$). As predicted, there were also no significant correlations between participants' preference for the brain test and their explicit Dualism, under either the circle ($r(46) = .11$, $p = .44$) or questionnaire ($r(46) = -.20$, $p = .17$) measures. Similarly, neither measure correlated with preference for the brain test for epistemic (circle: $r(46) = .09$, $p = .56$; questionnaire: $r(46) = -.13$, $p = .40$) and sensorimotor (circle: $r(46) = .10$, $p = .52$; questionnaire: $r(46) = -.20$, $p = .17$) symptoms.

Altogether, Experiment 1 shows that participants reliably prefer the brain test for symptoms that are epistemic. One limitation of these results, however, is that our procedure explicitly asked participants to contrast the two tests. The question thus arises whether participants would still show the same preference spontaneously, even when the task elicits no explicit comparison of the two tests.

To address this question, Experiment 2 asked participants to evaluate each test individually (on a 1–7 scale). After reading the vignette, participants were informed of the actual test outcome for the patient, which invariably suggested a disorder. With this information in mind, participants were asked to evaluate how likely is it that the patient has the disorder given the results of the brain and behavioural tests. The design and materials were otherwise identical to Experiment 1. We predict a stronger preference for the brain (over the behavioural) test for epistemic (relative to sensorimotor) symptoms.

2.2 Experiment 2: Preference for brain vs. behavioural test (rating)

2.2.1 Participants

48 Northeastern University students took part in this experiment in partial fulfillment of a course credit. Most participants reported taking no advanced level courses in psychology (79%), neuroscience (98%), or biology (63%).

2.2.2 Materials and Procedure

The vignettes were the same as those used in Experiment 1, with the exception that participants rated (on 1–7 scale) how effectively each test would evaluate the given symptom (rather than make a forced choice, as in Experiment 1) . Specifically, participants were asked how likely the patient would be to have the disorder if they demonstrated a positive test result. For example, the theory of mind vignette asked, "If a person was tested using the [behavioural/neural] method and showed a longer looking time when Alice reaches for the bowl, how likely is it that they have autism?". The response scale was as follows: 1 = Not at all Likely, 2 = Slightly Likely, 3 = Somewhat Likely, 4 = Moderately Likely, 5 = Very Likely, 6 = Extremely Likely, 7 = Absolutely Certain (Appendix B). Participants were also invited to offer a brief explanation for their response. The procedure was exactly the same as Experiment 1.

2.2.3 Design

Each participant received all four vignette pairs in a 4 Disorder (Autism, Language, Reading, Aging) x 2 Symptom (Epistemic, Sensorimotor) x 2 Test (Behavioural, Brain) within-subjects design. The order of the vignettes was counterbalanced in a Latin square design between subjects. In addition, the order of appearance of the two tests within the vignette (and the corresponding rating questions at the end) was also counterbalanced, between subjects.

2.2.4 Results and discussion

a. *Rating data.* An inspection of the mean ratings (Figure 4) suggests that overall, participants rated the brain test higher than the behavioural test. However, the preference for the brain test was stronger for epistemic symptoms ($\Delta = .31$) than for sensorimotor ones ($\Delta = .15$).

We evaluated these results using a 4 Disorder (autism/language/reading/aging) x 2 Symptom (epistemic/sensorimotor) x 2 Test (brain/behavioural) ANOVA. This analysis yielded a reliable main effect of Test ($F(1, 383) = 16.21$, $p < .001$, $\eta^2_{partial} = .26$), as the brain test ($M = 4.26$) was rated higher than the behavioural test ($M = 4.03$). The main effect of Symptom was non-significant ($F(1, 383) = 2.82$, $p = .10$). Critically, the interaction of Symptom and Test was significant ($F(1,191) = 4.22$, $p = .046$, $\eta^2_{partial} = .08$).

A Fisher's LSD test showed that for epistemic symptoms, participants rated the brain test significantly higher than the behavioural test ($\Delta = .31$, $t(47) = 3.51$, $p < .001$, $d = .36$). In contrast, when the symptoms were sensorimotor, the preference for the brain test was weaker ($\Delta = .15$) and only marginally significant ($t(47) = 1.97$, $p = .06$, $d = .20$).

Returning to the ANOVA results, we also found a significant effect of Disorder ($F(3, 189) = 26.08$, $p < .001$, $\eta^2_{partial} = .36$); as the language vignettes ($M = 3.09$) were rated lower than all other vignettes (autism ($M = 4.27$; $t(141) = 3.36$, $p < .001$, $d = .34$); reading ($M = 4.68$; $t(141) = 3.36$, $p < .001$, $d = .34$); aging ($M = 4.53$; $t(141) = 3.36$, $p < .001$, $d = .34$)). The effect of Disorder did not further interact with Symptom ($F(3, 93) = 2.13$, $p = .100$), but the interaction of Disorder and Test was significant ($F(3, 93) = 2.70$, $p = .05$, $\eta^2_{partial} = .05$). Tukey HSD showed that the brain test was rated higher than behavioural test only for the autism vignette ($\Delta M_{brain\text{-}behavioural} = .43$; $t(141) = 3.36$, $p < .001$, $d = .49$) but not for language ($\Delta M_{brain\text{-}behavioural} = .16$; $t(141) = .35$, $p = .73$, $d = .05$), aging ($\Delta_{brain\text{-}behavioural} = .06$; $t(141) < .01$, $p = .10$, $d < .001$) or reading ($\Delta M_{brain\text{-}behavioural} = .26$; $t(141) = 1.60$, $p = .11$, $d = .23$). Critically, these effects were not further modulated by symptom type, as the three-way interaction was not significant ($F(3, 45) = 1.81$, $p = .15$).

Summarizing, the rating results of Experiments 2 converge with those of Experiment 1 to suggest

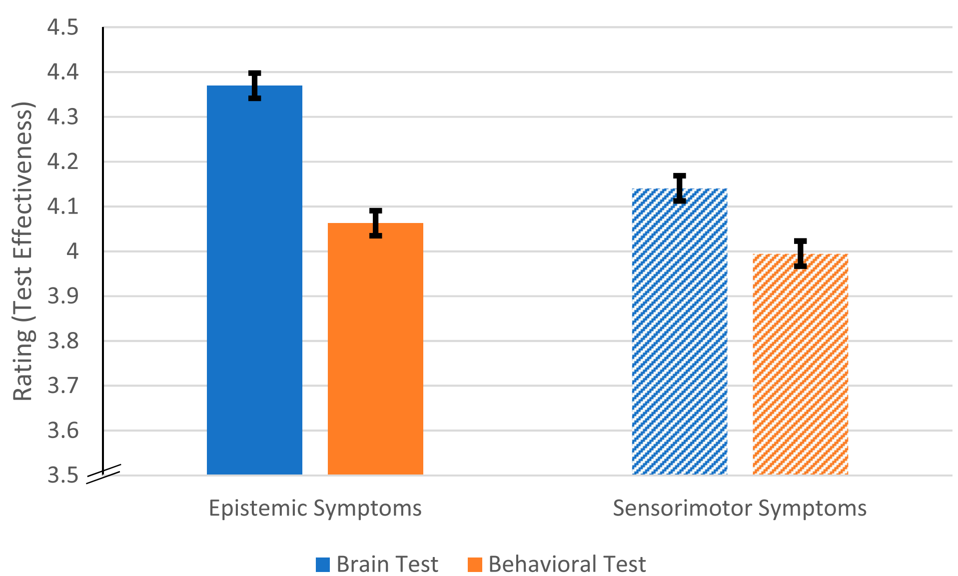

Figure 4. Preference for the brain and behavioural tests for epistemic and sensorimotor symptoms. (Error bars represent 95% confidence intervals for the difference between the means).

that the preference for the brain test is stronger for epistemic symptoms compared to sensorimotor symptoms. This result is remarkable, as the test outcomes were strictly matched, and participants in this experiment were not required to compare the two tests. Nonetheless, participants spontaneously preferred the brain test to the behavioural test.

b. *Written justification*. We next turned to examine participants' written justifications of their rating responses. An inspection of the mean proportion of positive responses (see Table 2) suggested that participants were most likely to offer positive comments when the brain test was paired with epistemic symptoms.

Table 2. Participants' mean positive and negative comments on the brain and behavioural tests.

			Symptom	
Comment type			Epistemic	Sensorimotor
Positive	Test	Brain	0.12	0.10
		Behavioural	0.03	0.08
Negative	Test	Brain	0.08	0.09
		Behavioural	0.22	0.22

A 2 Test x 2 Symptom GLMM for the positive responses yielded a reliable effect of Test (Wald $\chi^2 = 8.20$, $p = .004$, $\varphi = .30$) and a reliable Test x Symptom interaction (Wald $\chi^2 = 3.96$, $p = .05$, $\varphi = .14$); the effect of Symptom was not significant (Wald $\chi^2 = 1.603$, $p = .21$). A follow up analysis found that the brain test was associated with more positive comments for epistemic symptoms (Wald $\chi^2 = 9.36$, $p = .002$, $\varphi = .478$) but not for sensorimotor ones (Wald $\chi^2 = .51$, $p = .47$). Thus, as in Experiment 1, participants' positive comments mirrored the rating results. Not only did epistemic symptoms yield stronger preference for the brain test in rating, but they also elicited more positive comments for the brain- (relative to the behavioural) test.

Also, in line with Experiment 1, in Experiment 2, negative comments were overall infrequent. The 2 Test x 2 Symptom GLMM only found a significant effect of Test (Wald $\chi^2 = 25.114$, $p < .001$, $\varphi = .91$), as the brain test was less likely to elicit negative comments relative to the behavioural test. The effects of Symptom (Wald $\chi^2 = .09$, $p = .77$) and the interaction (Wald $\chi^2 = .09$, $p = .77$) were not significant.

c. *Explicit measures of Dualism*. As in Experiment 1, there was no evidence of Dualism in participants' explicit responses to the Dualism scales. For the "circle" question, participants' average score was 2.22 (on a 1–7 scale). Their response to the "questionnaire"(M = 2.82) was not significantly different from the scale's neutral midpoint (3- "Neither agree nor disagree", t = −1.03, $p = .15$, d = −.21). There were also no significant correlations between the average difference in rating between the brain or behavioural test and either the circle (r(46) = .19, $p = .20$) or questionnaire (r(46) = .01, $p = .95$) measure of explicit Dualism. Similarly, neither measure correlated with the difference in rating across epistemic (circle: r(46) = .07, $p = .63$; questionnaire: r(46) = .13, $p = .39$) or sensorimotor (circle: r(46) = .24, $p = .10$; questionnaire: r(46) = −.11, $p = .44$) symptoms. These conclusions converge with past research (Fernandez-Duque et al., 2015; Hook & Farah, 2013), suggesting that the preference for the brain is not due to explicit beliefs in Dualism. In what follows, we examine whether this attitude could arise from an implicit Dualist bias.

3. Part 2: Does the preference for the brain test arise from an implicit Dualist dissonance?

Why do people prefer the brain test? We suggest that this preference arises from an implicit Dualist dissonance between mind and the body. The dissonance arises because people spontaneously attribute the behaviour (changes to the body) of agents to their mental states (see Figure 1). For example, upon reading the "theory of mind" vignette, people might presume that it is the patient's beliefs about the protagonist that caused the patient to make a certain response (e.g., to indicate where the protagonist might look for the missing keys).

Per intuitive Dualism, mental states are immaterial, distinct from the body. Per intuitive physics, however, only material bodies can interact, so such effects of the immaterial mind on matter are impossible. Consequently, a Dualist dissonance ensues (see (1)). And since epistemic states (e.g., theory of mind) are perceived as particularly immaterial (more immaterial than sensations and actions), the dissonance produced by epistemic states is particularly acute (see (4)).

The brain test alleviates the dissonance by suggesting that the cause of the patient's behaviour effectively lies not in the realm of the ethereal mind but rather in matter—in the body (see (2)), and for this reason, people prefer to attribute behaviour to brain- relative to behavioural causes (which do not explicitly reference the body). Critically, since epistemic traits are expected to produce a stronger dissonance, they are expected to elicit a stronger preference for the brain test compared to sensorimotor traits (see (3)).

Summarizing, we assume that (a) epistemic traits are perceived as less material than sensory traits; and that (b) immaterial traits elicit a stronger dissonance, which in turn, is linked to a stronger preference for the brain test. Experiments 3–4 test each of these two predictions, in turn.

Experiment 3 tests whether participants view the epistemic traits (corresponding to the symptoms used in the previous experiments) to be less material than the traits corresponding to sensorimotor symptoms. To assess materiality, we invite participants to predict whether a given trait corresponds to a specific brain region. We hypothesized that epistemic traits would be perceived as less material; hence, they should be rated as less likely to be instantiated in the brain (compared to sensorimotor traits).

Experiment 4 evaluates the dissonance generated by epistemic and sensorimotor traits. We hypothesized that the less material the trait, the less likely it should be to "show up" in the brain. Accordingly, upon learning that a psychological trait *can* affect the brain, participants should experience a mind–body dissonance, and this dissonance, in turn, should elicit surprise. The surprise response can thus be used to gauge the underlying dissonance—its putative cause.

To gauge the mind–body dissonance, Experiment 4 invited participants to indicate how surprising it is that the traits in question would be localized in the brain (i.e., the body). If the preference for the brain is due to the Dualist dissonance, then paradoxically, the less material the trait (i.e., the less likely it is to be perceived as localized in the brain), the stronger the expected preference for the brain test.

Both experiments also included the explicit Dualism measures, discussed previously. In so doing, we sought to contrast the putative effect of intuitive tacit Dualism with that of participants' explicit beliefs in Dualism.

3.1 Experiment 3: The perceived materiality of epistemic vs. sensorimotor traits

To evaluate the perceived materiality of epistemic and sensorimotor traits, Experiment 3 asked participants to indicate how likely it is that the traits are localized in the brain. We chose this measure because laypeople typically identify brain instantiation with a specific brain area—few people are aware that cognition can be distributed across the brain. Participants, here, were invited to advise a neuroscientist in selecting an appropriate topic for his neuroimaging research. For this research to succeed, the scientist ought to target traits that are likely to be localized in a specific area in the brain. To help the scientist choose the appropriate target for neuroimaging, participants were asked to indicate how likely it is that any given trait corresponds to a specific region in the brain. The traits in question corresponded to the four pairs of symptoms—epistemic and sensorimotor—from previous experiments. If epistemic traits are perceived to be relatively immaterial, then participants should consider them as less likely to correspond to a specific region of the brain relative to sensorimotor traits.

3.1.1 Participants

24 participants were recruited from the undergraduate subject pool at Northeastern University to take part in this experiment for course credit. The majority reported no prior advanced level courses in psychology (83%), neuroscience (96%), or biology (66%).

The sample size was determined based on a pilot study of this experiment. An a-priori power analysis (GPOWER, Erdfelder et al., 1996) showed that a sample size of at least 6 participants would be required to obtain the large effect size for the hypothesized main effect of Trait ($\eta^2_{partial} = .48$) at a power of .8 and an alpha level of .05. We thus set the sample size for N=24.

3.1.2 Materials and procedure

The experiment invited participants to advise a neuroscientist in planning a neuroimaging study. The scientist seeks to determine which trait to image. This, participants are told, is a difficult question, as "it is unclear whether all human traits are equally likely to activate a specific region in the brain. While several human traits have been associated with specific brain regions, others simply don't have a clear physical location in the brain. If the scientist were to target one of the latter traits, his experiment would yield no clear results". Against this backdrop, participants were asked to rate each of the traits based on how likely it was to be associated with a specific brain region.

As in the previous experiments, the materials included four pairs of vignettes, each pair featuring an epistemic and a sensorimotor trait. These trait pairs were the same as those used in the previous set of vignettes, but they were presented as characteristics of typical individuals (as the planned neuroimaging study concerned normal brain functioning).

Each vignette began by introducing a trait, followed by a procedure designed to gauge whether a given brain region is linked to that trait. The traits and the procedures were similar to the symptoms in previous experiments (e.g., theory of mind, for autism, in Experiments 1–2). Finally, participants were given a rationale explaining how brain activation in response to the procedure (e.g., brain response in the true vs. false belief conditions) is linked to the trait in question (e.g., theory of mind).

After each vignette, participants rated how likely it is that there is a brain region that is mainly responsible for the trait in question (1 = Not at all Likely, 2 = Slightly Likely, 3 = Somewhat Likely, 4 = Moderately Likely, 5 = Very Likely, 6 = Extremely Likely, 7 = Absolutely Certain). All vignettes are provided in Appendix C.

Participants were tested individually in our lab. After signing informed written consent, each participant was seated at a computer and given the experiment via a Qualtrics survey. Each vignette appeared on a separate page, and participants could not revisit previous vignettes. After all eight vignettes, the participants answered demographic questions about their education and responded to the explicit Dualism questionnaires described in Experiment 1.

3.1.4 Design

Each participant received all four vignette pairs in a 4 Disorder (Autism, Language, Reading, Aging) x 2 Trait (Epistemic, Sensorimotor) within-subjects design. The order of the vignettes was counterbalanced in a Latin square design between subjects.

3.1.5 Results and discussion

An inspection of the means suggested that epistemic traits were rated as less likely to be localized in the brain than sensorimotor traits (see Figure 5).

These results were evaluated by a 4 Disorder (autism/language/reading/aging) x 2 Trait (epistemic/sensorimotor) repeated-measure ANOVA . The significant main effect of Trait ($F(1,95) = 4.91$, $p = .04$, $\eta^2_{partial} = .18$) confirmed that the epistemic traits were indeed considered as less likely to correspond to a specific brain region than the sensorimotor traits.

There was also a significant main effect of Disorder ($F(3,45) = 3.71$, $p = .02$, $\eta^2_{partial} = .14$). A Tukey's HSD test indicated that the aging vignettes ($M = 4.73$) were rated significantly higher than autism vignettes ($M = 3.77$; $t(69) = 2.47$, $p = .02$, $d = .70$). The Disorder x Trait interaction was not significant ($F(3,21) = 1.28$, $p = .29$).

The finding that epistemic traits are considered as less likely to be localized in the brain is in line with the possibility that epistemic traits are relatively immaterial. One concern, however, is that the "brain localization" may not necessarily reflect the perception of traits as immaterial. Rather, participants might perceive traits to be distributed over large brain networks, in line with modern neuroscience (Bressler & Menon, 2010).

There are several reasons to believe this is unlikely. First, most of our participants reported no advanced knowledge of neuroscience. Second, if, guided by state-of-the-art neuroscience, people should view brain functions as distributed, then they should also reject the localization of sensorimotor traits, not specifically epistemic ones (Bressler & Menon, 2010). Third, past research (with the same population) found that judgments of brain localization converge with body-replication—another test of materiality (see also Forstmann & Burgmer, 2015). People considered epistemic traits as less likely to be targeted by these body-affecting manipulations —both the brain localization and body replication (Berent et al., in press). But when asked to evaluate the propensity of the traits to persist in the afterlife (after the demise of the body), epistemic traits were now considered as most likely to persist (Berent et al., in press). This double dissociation demonstrates that brain localization and body replication target the same construct (the material body), as distinct from the mind (in line with Dualism).

To put this concern to rest, in a partial replication of Experiment 3, we asked another group of participants ($N = 24$) to indicate whether the traits in question are likely to activate the brain (without reference to any specific localization; the vignettes were otherwise

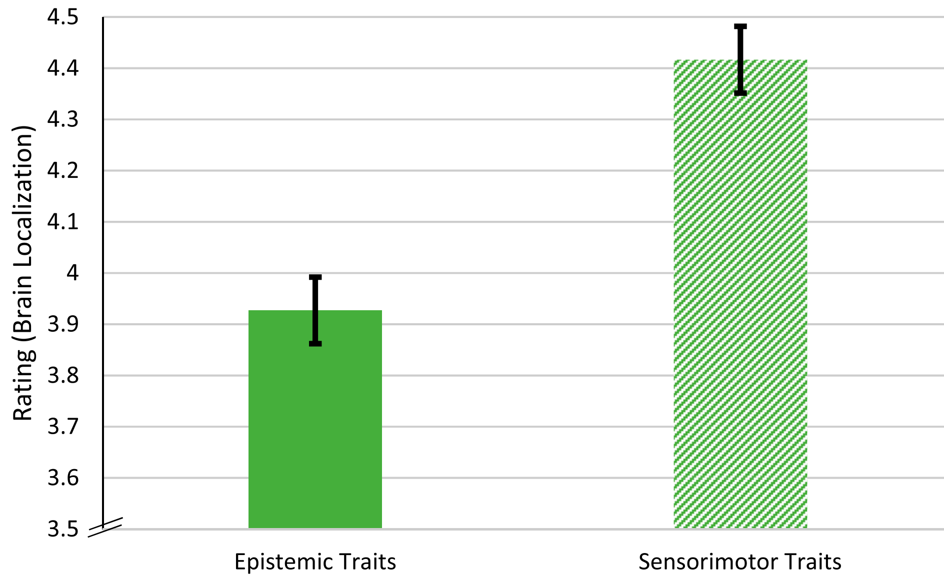

Figure 5. Difference in perceived materiality between epistemic and sensorimotor traits. (Error bars represent 95% confidence intervals for the difference between the means).

identical to the ones in Experiment 3). The effect of Trait was, once again, significant ($F(1, 95) = 4.16$, $p = .05$, $\eta^2_{partial} = .15$), irrespective of Disorder (for the interaction: ($F(3, 21) = 0.23$, $p = .86$)). Moreover, a comparison of the original Experiment 3 and its replication (using a 2 Experiment x 4 Disorder x 2 Trait ANOVA) yielded a reliable effect of Trait ($F(1, 383) = 7.39$, $p = .01$, $\eta^2_{partial} = .24$). The effect of Experiment was not significant ($F(1, 383) = 1.01$, $p = .32$) nor did it interact with Trait ($F(1, 191) = 0.01$, $p = .91$).

With all likelihood, then, brain localization is a valid measure of perceived materiality, not "blobology" per se. If so, the rating of epistemic traits as ones that are less likely to be localized in the brain indicates they are perceived as immaterial (relative to sensorimotor traits).

The Dualist dissonance account asserts that it is this perceived immateriality that leads participants to prefer brain-based diagnoses for epistemic symptoms (as outlined in Part 1). Briefly, we suggest that the effect of such immaterial traits on behaviour (changes in the material body) triggers a stronger dissonance, which is alleviated by the brain test. We thus predict that the traits perceived as least material should elicit the strongest preference for the brain test. Paradoxically, then, people should favour the brain diagnosis the most for symptoms/traits that they consider least susceptible to neuroimaging (i.e., for traits they consider least likely to be localized in the brain).

To evaluate the predicted (directional) relationship between the preference for the brain test and putative materiality, we correlated the putative materiality of the eight traits (as assessed by the mean rating of their potential to be localized in the brain, in Experiment 3) with the preference for the brain test (the difference rating for the brain and behavioural tests, in Experiment 2) using a directional Spearman's rho correlation . The correlation was negative and significant ($\rho = -.65$, $p = .04$, one-tailed; Figure 6).

While our results suggest that the preference for the brain test is linked to participants' implicit Dualism (as gauged by their perception of materiality), we found no such association with the explicit measures of Dualism. Participants' average score on the "circle" question was 2.0 on the 1–7 scale, and their means on the "questionnaire" scale ($M = 2.82$) were not significantly different from the scale's "neutral" midpoint (3- "Neither agree nor disagree", $t = -.95$, $p = .18$, $d = -.27$). There was no significant correlation between the average materiality rating and either the circle ($r(22) = .05$, $p = .80$) or questionnaire ($r(22) = .08$, $p = .71$) measure of explicit Dualism. Similarly, neither measure correlated with materiality across epistemic (circle: $r(22) = .08$, $p = .71$; questionnaire: $r(22) = .09$, $p = .67$) or sensorimotor (circle: $r(22) = .00$, $p = 1.00$; questionnaire: $r(22) = .03$, $p = .88$) traits.

Summarizing, Experiment 3 showed that participants rate epistemic traits as less likely to be localized in the brain, suggesting they are viewed as relatively

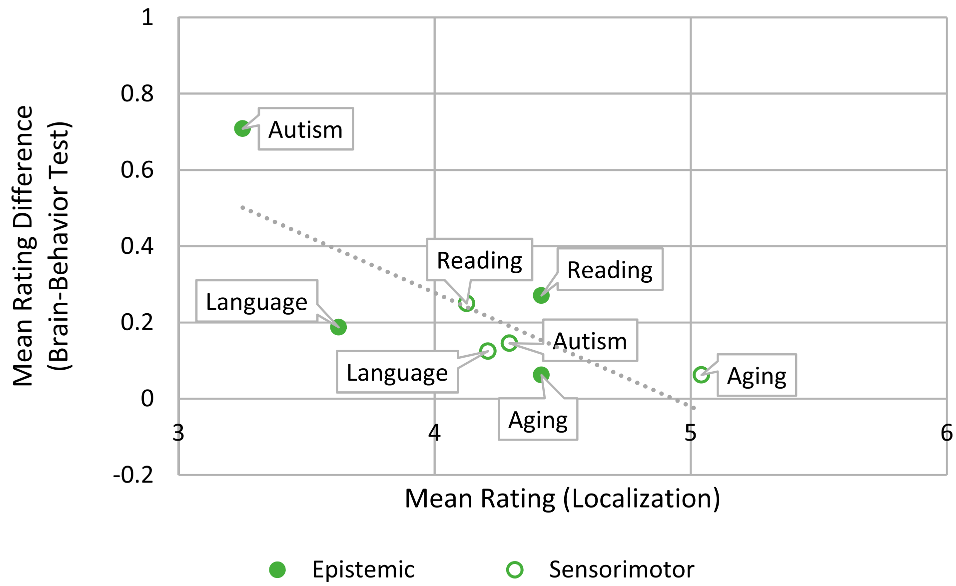

Figure 6. Relationship between perceived materiality and magnitude of preference for brain tests.

immaterial. Moreover, as the putative materiality of a trait decreased, there was an increase in the preference for brain- over the behavioural test. The Dualist dissonance account predicts that the perceived immateriality of epistemic traits should conflict with their capacity to affect behaviour (i.e., to effect change in the material body), and consequently, a Dualist dissonance should ensue. Experiment 4 tests for this dissonance and evaluates its contribution for the preference for neuroscience.

3.2 Experiment 4: Can neuroscience elicit mind–body dissonance?

To gauge the Dualist dissonance, we presented participants with evidence that epistemic and sensorimotor traits are, in fact, localized in the brain and asked them to indicate their reaction. The "cover" story invited participants to help a New York Times editor choose "scoops" for the Science section. Newsworthy stories, they were told, were ones that readers of the papers will find surprising—ones that the general public would find unexpected. Participants were presented with short descriptions of recent scientific discoveries, indicating that various psychological traits—epistemic or sensorimotor—are instantiated in the brain. The traits were similar to the ones in previous experiments. Participants rated their level at surprise at the discovery that a trait is instantiated in the brain. Surprise, in turn, was our gauge of the Dualist dissonance—we expected stronger dissonance to elicit higher surprise rating.

This measure, however, is limited inasmuch that the participants (students taking an introductory Psychology course) have learned that all psychological traits are instantiated in the brain. Accordingly, these participants are unlikely to consider any of these findings truly unexpected. Our interest, then, was not at the absolute level of awe but rather at the relative level of surprise in response to the various traits. In particular, we asked (a) whether the level of surprise would depend on the perceived the materiality of the trait; and (b) whether surprise would further correlate with the preference for the brain test.

The Dualist dissonance account predicts that, upon learning that traits perceived as disembodied can affect behaviour (as these vignettes patently demonstrate), participants should experience a Dualist dissonance, which, in turn, should promote their preference for the brain test (see (3)-(4)).

We thus made two predictions. First, the preference for the brain test is expected to be linked to the level of dissonance. Accordingly, the stronger the dissonance (i.e., surprise at the instantiation of a trait in the brain), the stronger the preference for the brain test. Second, dissonance is expected to result from the perceived immateriality of the trait. We thus expect that, the less material the trait, the stronger the dissonance (i.e., surprise) associated with its instantiation in the brain.

3.2.1 Participants

48 participants were recruited from the undergraduate subject pool at Northeastern University to take part in this experiment for course credit. The majority reported no prior advanced level courses in psychology (81%), neuroscience (100%), or biology (69%).

3.2.2 Materials and procedure

The experiment invited participants to help a New York Times editor select a noteworthy story for the Science section—one that "the general audience would not have otherwise expected".

As in the previous experiments, the materials included four pairs of vignettes, each pair featuring an epistemic and a sensorimotor trait; the traits were the ones from previous experiments, except that they were presented as characteristic of typical individuals (as in Experiment 3).

Each vignette first introduced the trait, and described a behavioural test used in its evaluation in past research (similar to the methods described in previous experiments). Participants were next presented with the current study, where the behavioural test was shown to activate a specific brain region. Participants were asked to rate (on a 1–7 scale) how surprising the finding is (1 = Completely obvious, 2 = Obvious, 3 = Predictable, 4 = Unremarkable, 5 = Noteworthy, 6 = Surprising, 7 = Extremely surprising). All vignettes are included in Appendix D.

Participants were tested individually in our lab. After signing informed written consent, each participant was seated at a computer and given the experiment via a Qualtrics survey, featuring the vignettes described above. Each vignette appeared on a separate page, and participants could not revisit previous vignettes. After responding to all eight vignettes, participants were presented with questions about their demographic and education and given the explicit Dualism questionnaires described in Experiment 1.

3.2.4 Design

Each participant received all four vignette pairs in a 4 Disorder (Autism, Language, Reading, Aging) x 2 Trait (Epistemic, Sensorimotor) within-subjects design. The order of the vignettes was counterbalanced in a Latin square design between subjects.

3.2.5 Results and discussion

Overall, participants were not very surprised to learn that psychological traits are instantiated in the brain. The mean "surprise" rating (M = 3.88) did not differ significantly from 4, which corresponds to the scale's "unremarkable" midpoint ($t(47) = .81$, $p = .42$). Accordingly, the 4 Disorder x 2 Trait ANOVA did not yield a significant main effect of Trait ($F(1,191) = .85$, $p = .36$), nor did Trait interact with Disorder ($F(3,45) = 2.08$, $p = .11$). The only significant effect was of Disorder ($F(3, 93) = 6.15$, $p < .001$, $\eta^2_{partial} = .12$), as the mean rating of the aging vignettes (M = 3.33) was significantly lower than for autism (4.17; $t(141) = 3.36$, $p < .001$, d = .57), reading (3.95; $t(141) = 2.32$, $p = .02$, d = .38), and language (4.08; $t(141) = 3.15$, $p = .002$, d = .54) vignettes.

Although participants showed little surprise to learn that psychological traits are instantiated in the brain, it is still possible that the level of surprise would vary across individual traits, and it would be associated with the preference for the brain test. To evaluate this possibility, we correlated the surprise rating for each trait with the preference for the brain test for that trait (gauged by the difference between mean rating for the brain and behavioural test in Experiment 2) using a Spearman's rho correlation. The correlation between surprise and the preference for the brain test was strong, significant, and positive ($\rho = .78$, $p = .023$; Figure 7). As participants were more surprised to learn that a trait is instantiated in the brain, they paradoxically preferred the brain as a diagnostic method for the trait in question.

The significant association between surprise (a gauge of their putative dissonance) and the preference for the brain test is in line with the hypothesis that the preference for the brain test is due to a dissonance, generated by epistemic traits. To further determine whether the dissonance is linked to tacit Dualist beliefs about the materiality of body and mind, we next examined the association between the surprise rating of each trait (a measure of dissonance) and its perceived materiality rating (in Experiment 3). The correlation was significant and negative ($\rho = -.71$, $p = .05$; Figure 8). Thus, as participants were less likely to perceive a trait as material, they were more surprised to learn that is instantiated in the brain.

Summarizing, participants in Experiment 4 were not particularly surprised to learn that epistemic

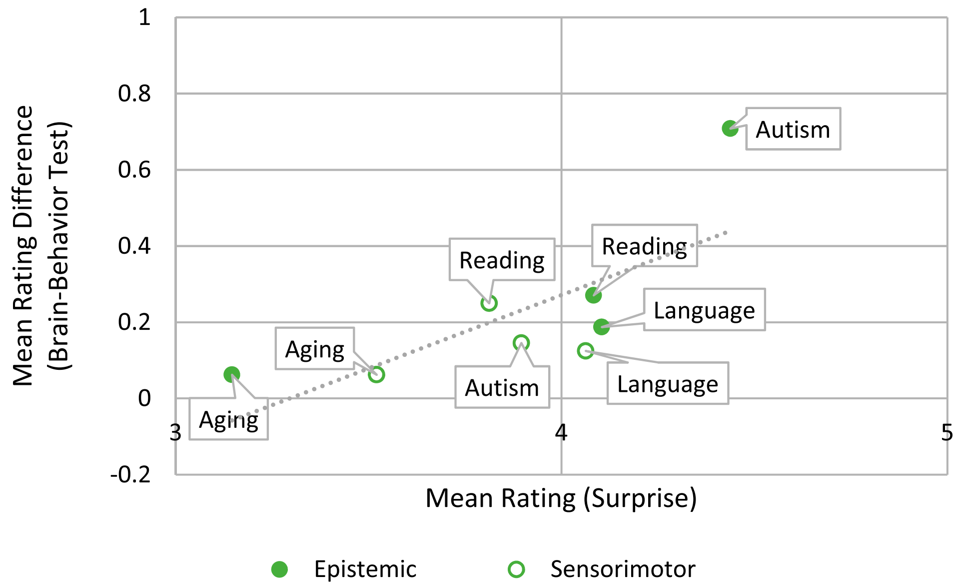

Figure 7. Relationship between reported dissonance and magnitude of preference for brain tests, for the eight traits used in this set of experiments.

traits are instantiated in the brain, possibly, because, as psychology students, they were well aware of the findings from cognitive neuroscience. Nonetheless, we did find a significant correlation between dissonance and the preference for the brain test (in Experiment 2). Thus, the more surprised participants were to discover that a given trait is instantiated in the brain, the more likely they were to prefer a brain-based diagnostic test for the trait in question. Additionally, dissonance (as gauged by surprise) also correlated with materiality ratings (in Experiment 3): the less material the trait, the stronger the dissonance.

The first finding— the positive association between the perceived materiality of a trait and the surprise at its neural instantiation—is rather expected. If a trait is unlikely to correspond to a brain region, then it would be surprising indeed if the traits were detected by an imaging study. In contrast, the negative association between the preference for the brain test and surprise is utterly paradoxical. We see no reason to prefer a brain test for traits that are unlikely to be localized in the brain. While this paradoxical association is striking, it is entirely predicted by the logic of Dualist

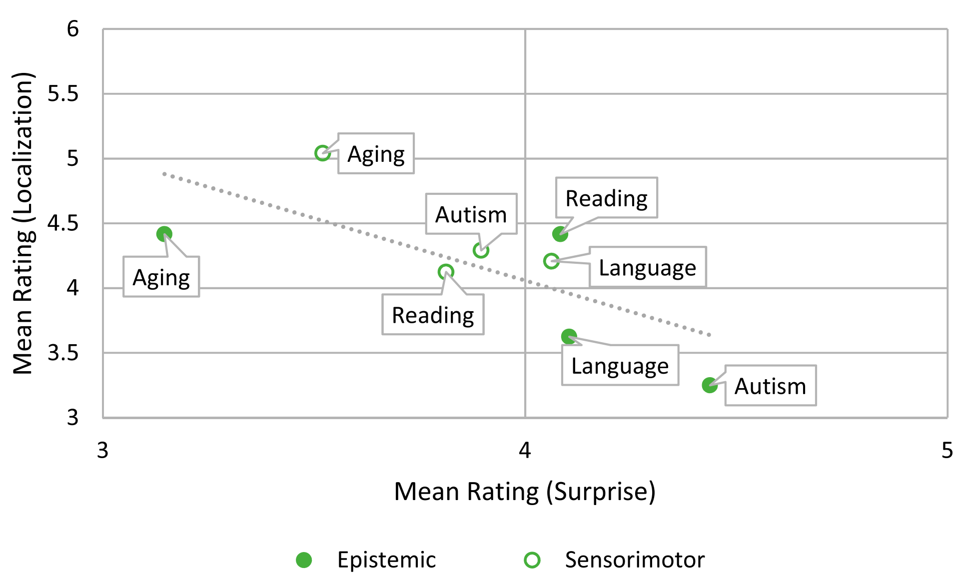

Figure 8. Relationship between reported dissonance and perceived materiality, for the eight traits used in this set of experiments.

intuition. Since immaterial psychological traits are a source of dissonance, it would be rational to seek to alleviate the dissonance by seeking to anchor the trait in the material brain. Thus, implicit Dualism offers an explanation for the paradoxical preference for a brain test for traits perceived to be difficult to image as well as for their overall preference for brain causes of behaviour.

As in previous experiments, we found no evidence that the preference for the brain test is explained by explicit beliefs in Dualism. For the "circle" scale, participants' average score was 1.85, and their "questionnaire" means were not significantly different from 3 (Neither agree nor disagree), the midpoint of that scale ($M = 2.87$, $t = -.09$, $p = .19$, $d = -.02$). There was no significant correlation between the average dissonance rating and either the circle ($r(46) = -.02$, $p = .92$) or questionnaire ($r(46) = -.08$, $p = .61$) measure of explicit Dualism. Similarly, neither measure correlated with dissonance across epistemic (circle: $r(46) = .05$, $p = .74$; questionnaire: $r(46) = .02$, $p = .90$) or sensorimotor (circle: $r(46) = -.08$, $p = .60$; questionnaire: $r(46) = -.16$, $p = .28$) traits. Altogether, then, our results offer no evidence that the preference for the brain is linked to explicit Dualist beliefs about body and mind. In contrast, our results suggest that the brain preference could be linked to perceived dissonance, prompted by implicit beliefs in Dualism.

4. General discussion

This research was designed to evaluate a novel explanation for the Seductive Allure of Neuroscience (SAN). Past research has shown that people exhibit a paradoxical preference for brain explanations, even when these explanations are, in fact, circular (e.g., Weisberg et al., 2008). Here, we explore the possibility that the SAN effect arises from a Dualist dissonance (Berent, 2020).

This account assumes that laypeople are tacit Dualists—they perceive the mind as immaterial, distinct from the material body (Bloom, 2004). Everyday behaviour, however, shows that agents can control their behaviour—they command changes in their physical body. And per intuitive theory of mind, people ascribe this control to agent's (immaterial) mind (e.g., their beliefs and goals). Per intuitive physics, however, only material objects can interact. Laypeople thus struggle to understand how the immaterial mind can affect behaviour (changes to the material body, such as moving one's arm); a mind–body dissonance thus ensues. Neuroscience information alleviates the dissonance by suggesting that behaviour is caused by the (material) brain. Since causation is now strictly material (brain —> behaviour), brain causes are preferred. Thus, the SAN effect arises from tacit Dualism.

The dissonance account generates two unique predictions. First, since Dualist dissonance (the cause of the SAN preference) arises from the tendency to attribute behaviour to mentalistic causes, and since (per theory of mind), people presume these mentalistic causes spontaneously, even when they are not engaged in explicit explanation, we expect the dissonance, and thus the preference for neuroscience, to arise in causal reasoning about behaviour (as opposed to solely in the context of explaining it). The seductive allure, then, is of neuroscience, generally, not of neuroscience explanations, narrowly.

Second, we predict that the strength of the SAN effect should depend on the strength of the dissonance. The dissonance, in turn, should depend on the perceived materiality of the psychological traits in question. Psychological traits that are perceived as devoid of material instantiation in the body should engender a larger dissonance, hence, a stronger preference for brain-based explanations. Given that past research has shown that traits which invoke epistemic states (e.g., knowledge that "objects are cohesive") are perceived as less material than sensorimotor traits (e.g., grasping with one's hands; Berent et al., in press), we thus predicted that epistemic traits should generate a larger dissonance, hence, a stronger SAN effect.

Experiments 1–2 examined whether the preference for brain-based tests would be stronger for epistemic traits relative to sensorimotor ones. Experiments 3–4 investigated whether the preference for the brain test corresponds to the perceived materiality of the trait and the strength of mind–body dissonance. The results were largely consistent with the Dualist dissonance hypothesis.

To examine the SAN in causal reasoning, in Experiments 1–2, we invited participants to compare brain- and behavioural tests of a clinical condition (e.g., autism). Since the two tests were strictly matched for their informative value, a preference for the brain test would suggest a seductive allure of neuroscience. These two tests, in turn, were applied to two types of clinical conditions, affecting either epistemic (e.g., a theory of mind deficit) or sensorimotor (e.g., auditory hypersensitivity) traits.

We reasoned that, if people prefer to attribute behaviour to brain causes (possibly, because such causal chains are better formed, see (2)), then they should consider scenarios that are linked to brain causes as more plausible. Thus, when participants are told that a patient suffers from a disorder, they should be more likely to choose a brain- (over a behavioural) diagnostic test for the disorder (in Experiment 1). Similarly, they should rate the brain test as a better diagnostic for that condition relative to a behavioural test (in Experiment 2).

Results from the two experiments converged. In both cases, the rating results showed a larger preference for the brain test for epistemic- relative to sensorimotor symptoms. Additionally, participants were more likely to positively comment on the brain- (relative to the behavioural test) when the brain test was paired with epistemic symptoms. The generality of this preference across procedures, irrespective of whether participants contrasted the two tests explicitly (in Experiment 1) or implicitly (in Experiment 2), demonstrates that the SAN preference is robust, and it emerges spontaneously.

Experiments 3–4 further suggest that this preference arises from a Dualist dissonance, prompted by the perceived instantiation of psychological traits in the material body. Experiment 3 evaluated the perceived materiality of epistemic and sensorimotor traits. To this end, we invited participants to judge whether the traits in question are localized in the brain. Results showed that epistemic traits are perceived as less material (i.e., less likely to be localized in the brain) relative to sensorimotor traits. Moreover, this putative materiality of a trait correlated negatively with the preference for the brain test, in Experiment 2.

Experiment 4 evaluated whether epistemic traits are associated with a stronger mind–body dissonance. Here, participants indicated how surprising it is that a given trait activates the brain; we considered this surprise as the putative dissonance. Given that our participants—college students in Introduction to Psychology—are well aware that all psychological traits affect the brain, we did not expect these participants to be overtly surprised by such findings, and indeed, they were not. Still, the level of surprise (i.e., dissonance) was negatively associated with the perceived materiality of the trait and it was positively associated with the strength of the SAN effect (i.e., the preference for the brain- relative to the behavioural test, in Experiment 2). This latter association is paradoxical: the less likely were people to perceive a trait as neutrally instantiated, the stronger was their preference for a brain test. In other words, people considered a brain test *most* appropriate for the traits they considered *least* likely to "show up" in the brain. This result makes it clear that the SAN preference is based on intuitive, rather than deliberate reasoning . Indeed, while our findings offer various indications that the SAN effect depends on a tacit mind–body dissonance, none of our results were modulated by explicit beliefs in Dualism.

Our results offer several contributions to the SAN literature. First, we have demonstrated that the SAN bias arises in causal reasoning implicitly and spontaneously, even when the task does not explicitly ask participants to evaluate the soundness of an explanation (as in Weisberg et al., 2008). Second, our findings are the first to show that the strength of the SAN effect varies for epistemic and sensorimotor traits based on their perceived materiality: epistemic (i.e., relatively immaterial) traits show a stronger SAN effect than the (more material) sensorimotor characteristics. Third, the preference for the brain test emerged even when brain and behavioural tests were strictly matched for their informative value. In particular, unlike in previous research (Fernandez-Duque et al., 2015; Hopkins et al., 2016; Rhodes et al., 2014; Weisberg et al., 2008; Weisberg et al., 2015), brain tests in the present investigation could not have possibly offered additional information (e.g., the localization of a brain function) over the behavioural test, as the test only showed whether or not the manipulation of interest activated the brain. As such, our results confirm that SAN is an implicit bias.

Our results converge with the past findings of Fernandez-Duque et al. (2015) and Hook and Farah (2013) to suggest that SAN is unaffected by explicit beliefs in Dualism. Remarkably, in our experiments, the explicit and implicit beliefs in Dualism dissociate. While we found no evidence that the SAN effect depends on explicit beliefs in Dualism, our results offer multiple indications that link the SAN effect to a tacit Dualist bias. To the best of our knowledge, our results are the first to link the SAN bias to tacit Dualism.

In addition, our results contribute to the literature on core knowledge and its effects on scientific reasoning. Previous research has shown that human infants possess core knowledge of the physical and social worlds (Spelke & Kinzler, 2007) that shapes science understanding in adulthood, even in individuals who have received formal training (Shtulman, 2015). For example, Mccloskey (1983) found that high school and college students incorrectly expect that a ball whirled on a string which is cut loose will follow a curved trajectory, due to the core principle of Continuity (Spelke, 1994). The account presented here has shown how these principles could further pair with intuitive Dualism to elicit the SAN effect.

Finally, our present results converge with ongoing research, showing how intuitive Dualism and Essentialism can combine to systematically bias our intuitive understanding of the mind (Berent, 2020; Berent et al., 2019; Berent et al., 2020; Berent et al., in press; Berent & Platt, 2021; Berent & Platt, 2021). These results call for caution in the evaluation of cognitive and neuroscience explanations of behaviour by laypeople and scientists alike. We hope that by identifying these biases, these findings could help promote science literacy.

Declarations of interest

None.

Notes

1. How these two early systems of core knowledge—about intuitive physics and the minds of others— relate to intuitive Dualism is unknown. One possibility is that Dualism is simply the conjunction of these two core knowledge systems. Another possibility is that the Dualist bias is distinct from these early systems. To err on side of caution, here, we adopt the latter hypothesis; we thus discuss Dualism as distinct from these two core knowledge systems. But we note the possibility that they could indeed be one and the same.
2. In the account outlined here, brain causes of behavior are preferred because they *replace* the mentalistic causes, supplied to us by theory of mind. But it is possible that the original mentalistic causation is not entirely eliminated but only *demoted*; the brain-cause is preferred because it is now the more immediate (efficient) cause of behavior. Our present results do not allow us to adjudicate between these possibilities. We are indebted to an anonymous reviewer who pointed the second possibility to us.

Disclosure statement

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

References

Berent, I. (2020). *The blind storyteller: How we reason about innateness*. Oxford University Press US.

Berent, I., Feldman-Barrett, L. F., & Platt, M. (2020). Essentialist biases in reasoning about emotions. *Frontiers In Psychology: Cognitive Science*, *11*(23). https://doi.org/10.3389/fpsyg.2020.562666

Berent, I., & Platt, M. (2021). The true "me"—mind or body? *Journal of Experimental Social Psychology*, *93*, 104100. https://doi.org/10.1016/j.jesp.2020.104100

Berent, I., & Platt, M. (2021). Laypeople's misconceptions towards psychiatric disorders: Brain disorders are presumed innate. *Cognitive Science*. e12970.

Berent, I., Platt, M., & Sandoboe, G. M. (2019). People's intuitions about innateness. *Open Mind*, *3*, 101–114. https://doi.org/10.1162/opmi_a_00029

Berent, I., Platt, M., & Sandoboe, G. M. (in press). Empiricism is natural: It arises from dualism and essentialism. *Oxford Studies in Experimental Philosophy*.

Bering, J. M. (2002). *Intuitive Conceptions of Dead Agents' Minds: The Natural Foundations of Afterlife Beliefs as Phenomenological Boundary*, *2*(4), 263. https://doi.org/10.1163/15685370260441008

Bering, J. M., & Bjorklund, D. F. (2004). The natural emergence of reasoning about the afterlife as a developmental regularity. *Developmental Psychology*, *40*(2), 217–233. https://doi.org/10.1037/0012-1649.40.2.217

Bloom, P. (2014). *Descartes baby: How the science of child development explains what makes us human*. Basic Books.

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), 277–290. https://doi.org/10.1016/j.tics.2010.04.004

Castaneda, M., Ostrosky-Solis, F., Perez, M., Bobes, M. A., & Rangel, L. E. (1997). ERP assessment of semantic memory in Alzheimer's disease. *International Journal of Psychophysiology*, *27*(3), 201–214. https://doi.org/10.1016/S0167-8760(97)00064-0

Chudek, M., Mcnamara, R. A., Birch, S., Bloom, P., & Henrich, J. (2018). Do minds switch bodies? Dualist interpretations across ages and societies. *Religion, Brain & Behavior*, *8*(4), 354–368. https://doi.org/10.1080/2153599X.2017.1377757

Cohen, E., & Barrett, J. (2008). When minds migrate: Conceptualizing spirit possession. *Journal of Cognition and*

Culture, *8*(1-2), 23–48. https://doi.org/10.1163/156770908X289198

Cohen, E., Burdett, E., Knight, N., & Barrett, J. (2011). Cross-cultural similarities and differences in person-body reasoning: Experimental evidence from the United Kingdom and Brazilian Amazon. *Cognitive Science*, *35*(7), 1282–1304. https://doi.org/10.1111/j.1551-6709.2011.01172.x

Connell, L. (2004). Making the implausible plausible. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*(26). https://escholarship.org/uc/item/1017s268

Connell, L., & Keane, M. T. (2004). What plausibly affects plausibility? Concept coherence and distributional word coherence as factors influencing plausibility judgments. *Memory & Cognition*, *32*(2), 185–197. https://doi.org/10.3758/BF03196851

Emmons, N. A., & Kelemen, D. (2014). The development of children's prelife reasoning: Evidence from two cultures. *Child Development*, *85*(4), 1617–1633. https://doi.org/10.1111/cdev.12220

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Fearon, C., Butler, J., Waechter, S., Killane, I., Newman, L., Munteanu, T., … Lynch, T. (2017). EEG shows subclinical motor preparation deficits in Parkinson's disease patients with freezing of gait during locomotion [abstract]. *Movement Disorders*, *32*.

Fernandez-Duque, D., Evans, J., Christian, C., & Hodges, S. D. (2015). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*, *27*(5), 926–944. https://doi.org/10.1162/jocn_a_00750

Forstmann, M., & Burgmer, P. (2015). Adults are intuitive mind-body dualists. *Journal of Experimental Psychology: General*, *144*(1), 222–235. https://doi.org/10.1037/xge0000045

Forstmann, M., Burgmer, P., & Mussweiler, T. (2012). The mind is willing, but the flesh is weak": the effects of mind-body dualism on health behavior. *Psychological Science*, *23*(10), 1239–1245. https://doi.org/10.1177/0956797612442392

Gruber, D., & Dickerson, J. A. (2012). Persuasive images in popular science: Testing judgments of scientific reasoning and credibility. *Public Understanding of Science*, *21*(8), 938–948. https://doi.org/10.1177/0963662512454072

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. https://doi.org/10.1038/nature06288

Hood, B., Gjersoe, N. L., & Bloom, P. (2012). Do children think that duplicating the body also duplicates the mind? *Cognition*, *125*(3), 466–474. https://doi.org/10.1016/j.cognition.2012.07.005

Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind-brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience*, *25*(9), 1397–1405. https://doi.org/10.1162/jocn_a_00407

Hopkins, E. J., Weisberg, D. S., & Taylor, J. C. V. (2016). The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition*, *155*, 67–76. https://doi.org/10.1016/j.cognition.2016.06.011

Hopkins, E. J., Weisberg, D. S., & Taylor, J. C. V. (2019). Does expertise moderate the seductive allure of reductive explanations? *Acta Psychologica*, *198*, 102890. https://doi.org/10.1016/j.actpsy.2019.102890

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*(1), 227–254. https://doi.org/10.1146/annurev.psych.57.102904.190100

Kiley Hamlin, J., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226. https://doi.org/10.1111/desc.12017

Koestler, A., & Butterfield, H. (1989). The sleepwalkers: A history of man's changing vision of the universe. https://archive.org/details/sleepwalkershist00koes_0

Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5-month-old infants see humans as material objects? *Cognition*, *94*(1), 95–103. https://doi.org/10.1016/j.cognition.2004.02.007

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402–408. https://doi.org/10.1111/1467-9280.01454

Leslie, A. M. (1987). Pretense and representation:The origins of 'theory of mind'. *Psychological Review*, *94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in Cognitive Sciences*, *8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. *Developmental Science*, *16*(3), 327–335. https://doi.org/10.1111/desc.12018

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*(1), 343–352. https://doi.org/10.1016/j.cognition.2007.07.017

Mccloskey, M. (1983). Intuitive physics. *Scientific American*, *248*(4), 122–130. https://doi.org/10.1038/scientificamerican0483-122

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., & Garry, M. (2013). On the (non) persuasive power of a brain image. *Psychonomic Bulletin & Review*, *20*(4), 720–725. https://doi.org/10.3758/s13423-013-0391-6

Minahan, J., & Siedlecki, K. L. (2016). Individual differences in need for cognition influence the evaluation of circular scientific explanations. *Personality and Individual Differences*, *99*, 113–117. https://doi.org/10.1016/j.paid.2016.04.074

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *Journal of Speech, Language, and Hearing Research*, *52*(4), 1073–1081. https://doi.org/10.1044/1092-4388(2009/07-0276)

Rhodes, R. E., Rodriguez, F., & Shah, P. (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1432–1440. https://doi.org/10.1037/a0036844

Riekki, T., Lindeman, M., & Lipsanen, J. (2013). Conceptions about the mind-body problem and their relations to afterlife beliefs, paranormal beliefs, religiosity, and ontological confusions. *Advances in Cognitive Psychology*, *9*(3), 112–120. https://doi.org/10.5709/acp-0138-5

Roach, N. W., & Hogben, J. H. (2007). Impaired filtering of behaviourally irrelevant visual information in dyslexia. *Brain*, *130*(Pt 3), 771–785. https://doi.org/10.1093/brain/awl353

Robbins, J., & Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly*, *81*(2), 407–420. https://doi.org/10.1353/anq.0.0005

Russeler, J., Becker, P., Johannes, S., & Munte, T. F. (2007). Semantic, syntactic, and phonological processing of written words in adult developmental dyslexic readers: An event-related brain potential study. *BMC Neuroscience*, *8*(1), 52. https://doi.org/10.1186/1471-2202-8-52

Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: An event-related potential study. *Psychological Science*, *11*(1), 46–50. https://doi.org/10.1111/1467-9280.00213

Schweitzer, N. J., Baker, D. A., & Risko, E. F. (2013). Fooled by the brain: Re-examining the influence of neuroimages. *Cognition*, *129*(3), 501–511. https://doi.org/10.1016/j.cognition.2013.08.009

Shtulman, A. (2015). How Lay cognition constrains scientific cognition. *Philosophy Compass*, *10*(11), 785–798. https://doi.org/10.1111/phc3.12260

Spelke, E. S. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*(1-3), 431–445. https://doi.org/10.1016/0010-0277(94)90039-6

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605–632. https://doi.org/10.1037/0033-295X.99.4.605

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, *13*(2), 113–142. https://doi.org/10.1111/j.2044-835X.1995.tb00669.x

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, A. J. Premack, D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). Clarendon Press/Oxford University Press.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Stanovich, K. E. (1989). Implicit philosophies of mind: The Dualism scale and Its relation to religiosity and belief in extrasensory perception. *The Journal of Psychology*, *123*(1), 5–23. https://doi.org/10.1080/00223980.1989.10542958

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392. https://doi.org/10.1016/j.ijpsycho.2011.12.007

Valenza, E., & Bulf, H. (2011). Early development of object unity: Evidence for perceptual completion in newborns. *Developmental Science*, *14*(4), 799–808. https://doi.org/10.1111/j.1467-7687.2010.01026.x

Weisberg, D. S., Hopkins, E. J., & Taylor, J. C. V. (2018). People's explanatory preferences for scientific phenomena. *Cognitive Research: Principles and Implications*, *3*(1), https://doi.org/10.1186/s41235-018-0135-2

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, *20*(3), 470–477. https://doi.org/10.1162/jocn.2008.20040

Weisberg, D. S., Taylor, J. C., & Hopkins, E. J. (2015). Deconstructing the seductive allure of neuroscience explanations. *Judgement and Decision Making*, *10*(5), 429–441. http://journal.sjdm.org/15/15731a/jdm15731a.pdf

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, *18*(1), 53–57. https://doi.org/10.1111/j.1467-8721.2009.01605.x