Routledge
Taylor & Francis Group

Philosophical Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cphp20

What does the CRT measure? Poor performance may arise from rational processes

Neil Levy

To cite this article: Neil Levy (2023) What does the CRT measure? Poor performance may arise from rational processes, Philosophical Psychology, 36:1, 58-84, DOI: 10.1080/09515089.2022.2038123

To link to this article: https://doi.org/10.1080/09515089.2022.2038123

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 12 Feb 2022.

Submit your article to this journal

Article views: 1113

View related articles

View Crossmark data

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

PHILOSOPHICAL PSYCHOLOGY
2023, VOL. 36, NO. 1, 58–84
https://doi.org/10.1080/09515089.2022.2038123

Routledge
Taylor & Francis Group

ORIGINAL PAPER

OPEN ACCESS Check for updates

What does the CRT measure? Poor performance may arise from rational processes

Neil Levy [a,b]

[a]Macquarie University and University of Oxford, Sydney; [b]Department of Philosophy Uehiro Centre for Practical Ethics, Macquarie University, University of Oxford, Australia United Kingdom

ABSTRACT

The Cognitive Reflection Test is a widely used measure of the degree to which individuals override an intuitive response and engage in reflection. For both theoretical and practical reasons, it is widely taken to assess an important component of rational thought. In this paper, I will argue that while doing well on the CRT requires valuable cognitive capacities and dispositions, doing badly does not always indicate a lack of such capacities and dispositions. The CRT, I argue, offers respondents implicit (but misleading) testimony: some of those who do badly do so due to a stronger disposition to defer to testimony. Since deference is not irrational, those who do badly need not be irrational, not even in part.

ARTICLE HISTORY
Received 5 November 2020
Accepted 31 January 2022

KEYWORDS
Cognitive reflection test; rationality; testimony; deference

An ancient tradition defines humanity as the rational animal. But a very large body of work in psychology seems to indicate that human rationality is very limited. Study after study seems to reveal that we are bad reasoners. We are subject to a range of heuristics and biases that distort our cognition (Kahneman, 2011; Kahneman et al., 1982); we are motivated reasoners, deploying our capacities to defend our gut feelings, our ideological allies and speculations flattering to us rather than searching for the truth (Kahan, 2013), we go with our gut because we are too lazy to analyze (Pennycook & Rand, 2019c), and so on.

There are two main ways to challenge the claim that these studies show us to be irrational. The first is to challenge the ecological validity of the findings. Perhaps the irrationality apparently manifested is the product of the fact that while human cognition works well when faced with the kind of challenges it is designed (by evolution) for, it misfires in the face of carefully designed laboratory tasks. If that's so, then the findings might indicate little about how rational we are in our day to day lives.[1] A second kind of response would consist in arguing that the findings have been

CONTACT Neil Levy neil.levy@philosophy.ox.ac.uk Macquarie University and University of Oxford; Department of Philosophy Uehiro Centre for Practical Ethics, Macquarie University, University of Oxford, NSW, 2109 Oxford, OX1 1PT, Australia United Kingdom

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

misinterpreted, and that the responses elicited in the lab are really rational. In this paper, I will pursue the latter kind strategy. Focusing on the Cognitive Reflection Test (CRT), I will argue while agents may respond badly (in some sense) on the items that comprise the test, responding *badly* does not entail responding *irrationally*.

Since this claim is central to the paper, it is important to say a few words about how I am using the terms 'rational' and 'rationality'. The apparent finding that human cognition regularly and predictably diverges from (Bayesian) normativity has sparked a now well-developed debate. One familiar response insists that these responses are *ecologically* rational, even if they are not Bayesian rational. They may depart from the former in that they ignore evidence, or respond perversely to it, or fail to respect transitivity; nevertheless, they are adaptive responses to the kinds of challenges we tend to face (or, in another version, tended to face in ancestral environments). Indeed, they may even be *epistemically* adaptive responses: they tend to generate the right answer by processing information in ways that depart from Bayesian rationality. In a phrase, we may be *rationally irrational*: by departing from Bayesian normativity, we better secure our epistemic goals. In this paper, I take no stand on whether bad responses on the CRT are ecologically rational. I am concerned with something more like Bayesian rationality. Ecological rationality is indirect: to the extent to which ecologically rational responses are (only ecologically) rational, they secure rationality by departing from direct rationality. I aim to defend the *direct* rationality of bad responses on the CRT.

The outputs of cognition are indirectly rational when they allow us to better secure our epistemic goals despite failing to weigh evidence in the normative manner. They are directly rational when the mechanisms adhere to the norms of logic and decision theory (Sturgeon, 1994; Way, 2017). Ecological rationality is assessed by focusing on outputs: a process is ecologically rational when it tends to generate the correct (or a better) output in appropriate circumstances. How they are generated is irrelevant to their being ecologically rational. Direct rationality is assessed by reference to the normative appropriateness of the process, and not the outcome: rational response is compatible with reaching an incorrect answer. This usage of 'rational' and the way in which it allows for a dissociation between rationality and correctness of response is familiar. If a detective's evidence is misleading, she might rationally identify the wrong culprit, for instance. I describe those who give an incorrect response to CRT items as responding *badly*. It is a further question whether in doing so they respond *irrationality*. Since two agents may give divergent but equally rational responses to the same information (due, for instance, to different Bayesian priors), we can't immediately infer from incorrectness of response to irrationality of processing.

In this paper, then, I will put forward evidence that I claim suggests that some significant proportion of the bad responses to the CRT are not irrational: they are generated by appropriate deployment of a mechanism designed for responding to reasons, in an environment that is epistemically hostile.[2] While success in identifying the right solution may indicate valuable cognitive competences, these failures do not arise from defects in reasoning competences.

The cognitive reflection test

The CRT is a very widely used test,[3] typically understood as measuring the extent to which individual cognition relies on reflective versus intuitive processes. The CRT (like other paradigms understood as probing rationality) is usually conceptualized within the framework of dual process (or dual system) accounts of cognition. On these kinds of views, cognitive processes tend to exhibit one or the other set of mutually incompatible features (Evans, 2008; Evans et al., 2013). Type 1 cognition is (always or typically, depending on the version) fast, mandatory in its operation, effortless (both in its phenomenology and insofar as it is not dependent on the availability of cognitive resources), and typically unconscious. Type 2 cognition has the opposite profile: it is slow, must be engaged and sustained effortfully, degrades under cognitive load and is conscious, in the sense that agents know when they are engaged in it. Type 1 cognition is typically assumed to be evolutionarily ancient, whereas Type 2 cognition is a more recent evolutionary adaptation.

The CRT consists of three questions (Frederick, 2005)

(1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents
(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes
(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days.

These questions *seem* to be very easy, requiring only simple arithmetic for their solution (we shall return to the question of their difficulty later). But the questions are worded to make a particular, wrong, response 'intuitive', where a response is intuitive in case it comes to mind easily, and is accompanied by a feeling of rightness (Thompson et al., 2011). For instance, in *Bat and Ball*, it is tempting to answer 10c, because $1.10 minus $1.00 is 10c. Most people do indeed give that answer (many who give the right

answer give this wrong answer first; (Frederick, 2005)). Reflection is required to override the intuitive response and calculate the correct one. We might check our answer, and notice that if the patch doubles every day, and it covers half the lake on day 24, it would cover the entire lake on day 25. Having noticed this, we can easily see the right answer: the patch must cover half the lake the day before day 48. While the reasoning involved seems easy, once we have worked through it, a surprisingly large proportion of participants answer none of the questions correctly. In his original paper, Frederick reported CRT scores – assessed by giving one point for a correct answer and zero for an incorrect – ranging from 2.18 for MIT students to 0.57 for University of Toledo students. Nowhere did a majority of participants get all three questions right.

The CRT is understood as tapping into a central and important component of rational cognition, on both empirical and theoretical grounds. The CRT is understood as assessing the *disposition* and *capacity* to engage in Type 2 processing. The CRT is understood to tap into Type 2 processing because success at the task seems to require (successfully) overriding cognitive miserliness: the disposition to avoid effortful processing in favor of heuristic processing. Effortful processing, as we have seen, is the domain of Type 2 cognition. Type 2 cognition, in turn, is often regarded as central to rational cognition, for a range of reasons. Type 2 cognition is unique to (or very much more significantly developed in) human beings (Evans et al., 2013), who are widely held to owe their adaptive success to their intelligence (Pinker, 2010). Our paradigms of rational thought, like mathematics, philosophy and science, are very heavily dependent on Type 2 cognition (Kahneman, 2011). Moreover, Type 2 cognition is domain general, allowing for a flexibility of thought that is denied to relatively encapsulated Type 1 cognition (Evans et al., 2013; Shea & Frith, 2016). Whereas Type 1 cognition is domain specific, Type 2 cognition enables us to combine propositions that pertain to any domain whatsoever with propositions that pertain to any other, which is a necessary condition for certain kinds of innovative thought (Levy, 2014) .

On the empirical side, there is evidence that CRT scores correlate with and predict beliefs and dispositions that themselves seem indicative or even constitutive of rational cognition. Frederick reported that CRT scores correlated with scores on the SAT (the standardized test widely used for college admissions in the United States), and also predict reduced temporal discounting. Since the capacity to forgo immediate reward for later, larger, rewards is essential to long term planning, this is evidence that the CRT measures an important capacity, with real-world consequences. More recently, Pennycook et al. (2016) reviewed evidence that CRT scores predict better performance on multiple decision making and reasoning tasks. As they note, CRT scores also appear to predict religious disbelief

(Pennycook et al., 2012; Shenhav et al., 2012) and belief in evolution (Gervais, 2015). As we saw above, moreover, those at more selective institutions perform better on the CRT than those at less selective institutions.

No task and no behavioral measure is process pure, of course, and few researchers would be willing to claim that the CRT should be understood simply as tapping into rational thinking. Evans et al. (2013), in particular, warn against identifying type 2 cognition with rational processing. Nevertheless, the standard view sees the CRT as tapping into two valuable and dissociable components of thought: a central component of thinking style, on the one hand, and analytic processing, on the other (Pennycook et al., 2014a; Toplak et al., 2011). In order to succeed on CRT items, one must, first, inhibit the intuitive response and, second, effortfully generate the correct response. Incorrect responses may therefore reflect either a disposition to be satisfied with intuitive responses without checking, or a lack of capacity to identify the correct response. It is the contribution of these two independent components of cognition that justifies the claim that it taps into the disposition *and* the capacity to engage in Type 2 reasoning. Agents may possess one without the other: they may possess the disposition to inhibit intuitive responses, but lack the capacity to identify the correct response, or they may possess the capacity to identify the correct response but lack the disposition to try (they may not be motivated to check their gut feelings, for instance).

Since both disposition and capacity seem to be independently valuable and jointly necessary for doing well on a range of reasoning tasks, there are therefore plausible reasons for thinking that the processes into which the CRT taps are required for good cognitive performance (especially, but not only, in epistemically hostile environments).It is this kind of thought that justifies (to some extent, at least), the incautious identification of the CRT as a test of "rational thinking" (Toplak et al., 2011) or "thinking well" (Royzman et al., 2014).

Given the correlation between CRT scores and a host of other measures of cognitive performance (Toplak et al., 2011), it is difficult to deny that it taps into abilities and dispositions that are important for cognitive success. Nevertheless, I will suggest that while *good* performance on the CRT requires valuable abilities and useful dispositions, *bad* performance need not indicate bad thinking or lack of important capacities (where "capacity" ranges over both cognitive ability and appropriate thinking styles).

A widely held account of failures of good performance on the CRT holds that incorrect answers stem from intuition (whereas correct answers stem from reflection). But this answer conflicts with the available evidence. First, reverse scoring the CRT – so that respondents score one point for every

intuitive answer and zero for any other response – yields results that do not correlate well with the Faith in Intuition (FI) scale (Pennycook et al., 2016). The FI scale is a widely used self-report scale, designed to measure the degree to which people trust their hunches and instincts (Epstein et al., 1996), and which correlates well with beliefs and personality traits associated with intuitiveness (Pacini & Epstein, 1999). The low correlation between FI and the reversed score CRT suggests that those who give the intuitive (but wrong) answer do not do so, in the main, *because* that answer is intuitive. In fact, those participants who gave the intuitive responses did not score higher on the scale than those who gave unintuitive but wrong responses. Second, for some participants the correct answer itself seems to be intuitive (Bago & De Neys, 2019). Perhaps for these participants, the arithmetic or logical reasoning required for generating the correct answer is overlearned, such that they can easily and effortlessly produce the right response (De Neys & Pennycook, 2019).

Third, there is evidence that the processes underlying the CRT can be dissociated, and the reflective component manipulated while the (putative) intuitive component cannot; this suggests that the CRT is not a measure of intuition (McPhetres, 2018). The process dissociation procedure aims mathematically to isolate and measure the independent contribution of automatic and controlled process when they might otherwise be confounded. If the CRT measures both reflection and intuition, as commonly held, then isolating them should yield parameters that can each be independently manipulated. This is not what McPhetres found: while cognitive load and speeded response decreased scores on the reflective component, they failed to increase scores on the intuitive component. Moreover, differences across individuals seemed to be very largely due to differences in the former and not the latter. This suggests that the CRT is not measuring individual differences in intuitiveness.

Fourth, and contrary to the repeated claim in the literature that CRT items are easy, a substantial number of individuals fail to identify the correct response when they perform the CRT in the form of a multiple choice quiz, and the intuitive response is not available (Patel et al., 2019). Even when prompted to deliberate and assured that the correct response is one of the four options presented, participants failed to identify the correct response on an average of 27% of trials (guessing would result in 25% of responses being correct, so while this is well above chance, it is far from the near 100% we'd expect if the items were genuinely easy when the intuitive option is removed). Response times increased significantly, but longer RTs were not associated with more accurate answers. Again, this suggests that poor performance on the CRT cannot simply be explained as arising from reliance on intuition.

I will argue that while good performance on the CRT may indeed be a function of cognitive ability and valuable thinking styles, poor performance is often due to a disposition to *defer*. By 'deference', I mean the disposition to accept testimony. One agent defers to another when she accepts the first agent's word for some proposition. As a number of philosophers have emphasized, much – perhaps most – of what we know, we know in virtue of testimony (Coady, 1992; Lackey & Sosa, 2006), and our pervasive reliance on testimony is now increasingly recognized within cognitive science (e.g., Rabb et al., 2019). Deference is often rational: other people are often better placed to assess some proposition than we are, in virtue of differences in expertise, in experience or diligence. We therefore routinely defer and thereby acquire knowledge. I will argue that deference often occurs in the absence of explicit testimony. One agent may defer to a second by acting or believing in the way that they (consciously or unconsciously) take to be recommended by the second agent's behavior, or even by subtle features of the context. The CRT, I will suggest, may rationally be taken to provide implicit recommendations, and poor performance on it may be a function of differences in dispositions to accept such recommendations.

Implicit testimony: Nudges

Testimony, as the term is used in philosophy, refers to information conveyed secondhand. The great majority of the philosophical debate over the epistemic status of testimony focuses on assertion: one agent provides testimony to another when the first intentionally and explicitly conveys information to the second (e.g., verbally); the second agent defers to the testimony (in the sense I mean here) when they accept the information conveyed. But testimony can be implicit as well as explicit, and deference to implicit testimony is a familiar phenomenon (though it is rarely thought of in these kinds of terms). Instead, deference to implicit testimony is usually seen through the same kind of prism that leads to perceiving low scores on the CRT as evidence of a lack of cognitive competencies. In this section, I will show how this same kind of misinterpretation occurs with a regard to a different set of studies: those concerned with so-called "nudges" (Thaler & Sunstein, 2008). Working through these examples will provide a sketch of the kind of hypothesis to be pursued here, by showing how a recognition of the role of implicit testimony in nudging aids a rethinking of apparently irrational responses in other contexts.

Nudges are canonically defined as ways of structuring the context of choice to encourage better choices (normally a choice is considered better if it promotes the interest of the choosing agent better than alternatives

would, though sometimes nudges aim at social welfare rather than the welfare of the individual chooser; see, Hagman et al. (2015) for discussion). Agents often make choices that are not in their own interests, but they may be nudged to choose better. For instance, there is evidence that quite trivial manipulations of food accessibility (for example, ensuring that healthier foods are on more easily reached shelves or at eye level) may lead to a higher consumption of such foods, relative to less healthy (Rozin et al., 2011; see, Bucher et al., 2016 for review). Similarly, people often fail to save an adequate amount of money to fund a decent retirement. But they can be nudged into saving more: if the default option presented to new employees sequesters a higher percentage of salary to a retirement savings account, people tend to save more, because they tend to accept the default (see, Smith et al., 2013 for review).

While such nudges may be in the interests of consumers, however, they are extremely controversial. Critics argue that (at least in the absence of further measures to mitigate these effects) nudges are manipulative or even coercive (Bovens, 2008; White, 2013; Wilkinson, 2013). These ethical issues are not our concern: rather, what matters in this context is that nudges are often held to threaten our autonomy because they rely on Type 1 cognition. On both sides, proponents and opponents of nudges typically hold that nudges bypass Type 2 cognition – and thereby the cognitive competencies realized by it – in favor of engaging with heuristics and biases. Proponents of nudges suggest that if we often make bad choices because our cognitive miserliness pushes us in unhealthy or unwise directions, the solution is to design the context of choice such that the same mechanisms promote better choice. Typical nudges supposedly take advantage of the fact that we are "somewhat mindless, passive decision makers" (Thaler & Sunstein, 2008, p. 37).

The argument that nudges threaten our autonomy because they bypass Type 2 cognition is most clearly developed by John Doris (Doris, 2015, 2018). He gives the example of the ballot order effect (Darcy & McAllister, 1990; King & Leigh, 2009). Candidates at the top of the ballot have an advantage over those lower down in the list, because people have a slight preference for higher ranked names. In close elections, this effect can be decisive. However, ballot order does not constitute a genuine reason in favor of voting for a candidate – ballot order is either alphabetical, or drawn by lot, and in neither case does ballot order correlate with candidate quality. It follows that in being influenced by ballot order, one is having one's choices affected by facts that are not reasons: one is not responding to the genuine reason-giving force of information (since, in fact, the information constituted by ballot order is not a consideration in favor of the candidates). Therefore, to the extent that choice is influenced by ballot order, genuine

reasoning is bypassed. Doris explicitly invokes dual process models of cognition in arguing that these kinds of influences do not involve genuine reasoning. They are, he says, "deeply unintelligent".[4]

If the CRT probes, at least in important part, our disposition and capacity to be guided by Type 2 cognition, then by this reasoning it assesses our capacity to override our tendency to fall prey to such "deeply unintelligent" mechanisms. On this view, success at the task is evidence that we reason well, and failure evidence of bad reasoning. There are good reasons, however, to think that some significant proportion of the variance in responses – on the both the CRT and to nudges – is explained by deference to implicit testimony, rather than by lack of competencies partially constitutive of rationality.

The three examples of nudges outlined above – defaults, ballot order effects and manipulations of the salience of options – may be understood as offering implicit testimony to agents. The standard view of default effects interprets them as the result of cognitive miserliness (the same disposition that explains why most of us are happy with the intuitively correct response on CRT items). But there's another interpretation available: defaults may be understood as communicative. Indeed, there is extensive evidence that agents tend to see the default option as authoritatively recommended to them: both experimental work (McKenzie et al., 2006) and modeling (Carlin et al., 2013) suggests that ordinary people see defaults as reflecting expert opinion, and they change their attitudes to the default accordingly (for instance, if organ donation is the default option, participants see it as much less of a sacrifice than if it must be actively chosen; Davidai et al., 2012). The presentation of defaults is likely understood as communicative because it *is* communicative: the selection of defaults is typically meant (implicitly) to convey a recommendation. There is evidence that the framing of options is communicative in this kind of way: for example, a research and development team is more likely to be described in terms of success rate, rather than its failure rate, if the person thinks highly of it (Sher & McKenzie, 2006). Framing of options is intended as communicative and understood as such. It is likely that selection of defaults functions the same way.

Ballot order effects and the effects of making items more or less accessible or salient may be understood in precisely the same way. While it is true that candidate quality does not, as a matter of fact, correlate with ballot order (except by chance), it may nevertheless be true that being guided by ballot order is a rational response to an implicit recommendation. That is, we may be disposed to respond to the order in which options are listed in a way that prefers items higher up the list to those lower down, because the order in which options are presented tends to correlate with their choiceworthiness,

and this correlation may hold because when we list options, we implicitly intend to communicate choiceworthiness by the order in which we list them.

Intuitively, something like this seems to be true in everyday life. An agent asked for a restaurant recommendation will tend to list those they prefer before those they think less of (and departures from this order, due to temporary forgetting, would likely be signaled in some way). Somewhat similarly, we expect more important items to be listed earlier in newspapers and their online equivalents, and to be reported earlier in TV broadcasts. In the light of the experimental evidence that we use framing to communicative recommendations, it seems a safe bet that we use order in the same way, and that our intention is understood. Responding to defaults, frames and manipulations of salience may therefore be understood as deference to implicit testimony.

Implicit testimony: The CRT

The CRT may also be understood as offering implicit testimony. It might be useful to frame the implicit recommendations encoded in CRT items in terms of *implicature*. To implicate something is to imply it contextually, without asserting it (Grice, 1991). Implicature can usefully be understood as implicit testimony: one agent provides testimony to another whether she explicitly asserts a claim ("the post office is open") or implicates it ("the post office is just around the corner"). In fact, almost all explicit testimony includes implicated information, since ordinary conversation relies on conversational implicature and it is only in special contexts in which we take care to attempt to convey all information literally (notoriously, we rarely succeed in such attempts). Without implicature, the content of testimony would be severely impoverished. We can also usefully understand many nudges and allied ways of influencing choice to work through implicature. The order in which we present information implicates its importance (Green, 1998); this fact explains the ballot order effect. Similarly, making an option salient, through the use of defaults or framing, may be understood as implicating that it is choiceworthy. The CRT offers implicit testimony by implicating certain responses.

As several commentators have noted, CRT items closely resemble trick questions (Haigh, 2016). Indeed, Thomson and Oppenheimer (2016) searched online for "trick questions" and "brain teasers" to develop a new version of the CRT (in light of the fact that the three item test is now well known). Like (or including) the CRT items, trick questions are in one sense very easy. It is easy to see why the right answer must be right, when we are told which one is right. But (again like, or including) the CRT, we tend to get them wrong. Consider these sample trick questions:

Q : Some months have 31 days, others have 30 days. How many have 28 days?

A : All of them.

Q : Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?

A : Emily.

Q : How many cubic feet of dirt are there in a hole that is 3' deep x 3' wide x 3' long?

A : None.[5]

In each case, there is an intuitive answer to the question, but that intuitive answer is incorrect (or at any rate, not the answer we're looking for). We tend to get these trick questions wrong not because they are complicated but because they misdirect us. The wording makes salient to us, for example, the month which has only 28 days (as opposed to those that have at least 28 days), and renders relatively pallid those facts that would lead us to the correct answer. Trick questions work by violating Grice's Cooperative Principle: "Make your conversational contribution what is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice, 1991). In particular, trick questions tend to violate the 'supermaxim' "be perspicuous", which enjoins us, among other things, to avoid "obscurity of expression" and ambiguity. Trick questions mislead us into giving the wrong answer because we interpret them in line with the cooperative principle, but they deliberately violate this principle. They implicate responses which are incorrect (or, again, not the response sought).

CRT items work in precisely the same way. The intuitive response, in virtue of being intuitive, is implicated (that is, implicitly recommended to our attention).[6] Just as, for example, the third question above misleads us into thinking we should do some basic arithmetic to solve the question (cube 3), so *Lily Pads* misleads us by making the concept 'half' salient, disposing us to halve '48' to get the answer. *Bat and Ball* and *Machines*, too, looks like arithmetic problem of a very simple sort: subtract 1.00 from 1.10 in the first case; multiply the number of machines by 1 in the second. These questions are worded to implicate, and thus implicitly recommend, certain responses. It is not because people struggle with basic arithmetic or with concepts like 'doubling' that they find these problems difficult: it is because they direct us to look in certain directions for the answer, and even if we notice or are told that the answer is wrong, it is difficult to reframe the problem.[7]

Patel et al. (2019) provide evidence that supports the view that the intuitive response is implicated and that this implication plays an important role in what response is given. They used a multiple-choice format, where the intuitive response was not available but (contrary to the common claim

that the questions are easy) many participants still struggled. That indicates that the questions are, for many people, difficult. They are not mathematically difficult, of course: they are difficult because they are worded in unhelpful ways, or because they need to be reframed to answer them (indeed, Patel and colleagues found that the CRT loaded onto numeracy and insight problem solving ability). It is not just because people have trouble inhibiting the wrong response that they give the incorrect answer on these problems. It is also because they find them difficult (Patel and colleagues note that time spent reflecting did not predict more accurate responses). Importantly, when we find ourselves unable to answer a question, it is *rational* to defer to implicated testimony (whereas when we can see the answer for ourselves, we should reject conflicting testimony). Indeed, this is what we do: it is when we find questions difficult that we defer to testimony (Mercier, 2017), and such deference is, in circumstances like these, appropriate. Giving the intuitive response on the CRT isn't an index of a failure of rationality. When an agent is faced with a difficult and confusing problem, she should defer to implicated testimony. That's the rational thing to do.

If one does *not* defer to the implicit testimony offered by the question, then identifying the correct answer will require valuable cognitive capacities, like numeracy and the capacity to reframe problems (whether by some kind of stroke of insight or by more effortful variation of parameters). But deferring to testimony itself is not irrational, and therefore giving the wrong answer does not index a *lack* of rationality. Instead, I suggest, it is (at least partly) an index of a disposition to defer. Agents who give the intuitive answer do so because it is implicitly recommended; thus those who are more strongly disposed to defer ought to give the intuitive answer more often than those who are less strongly disposed to defer.

This is a testable hypothesis. If it is true that generating the wrong answer is an index of a disposition to defer, rather than of cognitive laziness or lack of numeracy, we should expect a correlation between epistemic deference and low CRT scores, and we should expect that manipulating the disposition to defer should cause changes in CRT. As far as I know, there have been no attempts to test this hypothesis directly. However, there is indirect evidence in its favor.

Roughly speaking, political liberals have higher CRT scores than (social, but not economic) conservatives (Deppe et al., 2015; Pennycook & Rand, 2019a).[8] Conservativism, in turn, is characterized by a higher degree of ingroup loyalty and respect for authority (Graham et al., 2009; Talhelm et al., 2015). Ingroup loyalty and respect for authority, especially the latter, entail a greater willingness to take things on trust: to defer. Of course, epistemic authorities are a subset of authorities (not only do conservatives defer more than liberals, they are also more prone to see moral and political

authorities as possessing epistemic authority). This correlation between conservatism, epistemic deference and CRT is likely causal. As Talhelm et al. (Talhelm et al., 2015) note, liberals may be more analytical because they are more individualistic. As a result, liberals are more likely to accept norms that run counter to the social consensus. For instance, they are less condemning of actions that are usually judged to be immoral; (Pennycook et al., 2014b; Royzman et al., 2014).[9] These findings are what we would expect, given that conservatism is predicated on a more communal cognitive style and therefore a greater willingness to defer epistemically.

Consistent with this hypothesis Barr et al. (2015) found that participants who scored lower on measures of analytic thinking, including the CRT, report greater smartphone use and in particular greater use of search engines than those who score higher on such measures. The authors interpret this through the lens of dual process theories: that is, they suggest that those who score lower are cognitive misers, and avoid engaging in effortful thinking by having Google do their thinking for them. The evidence Barr et al. cite for the claim that the use of search engines reflects cognitive miserliness, and not research into competing viewpoints, is correlational: those who scored higher on cognitive miserliness were also more likely to report greater use of search engines. To whatever extent the use of search engines reflects an unwillingness to engage in critical reflection, this evidence can equally be understood as supporting reliance on a kind of epistemic deference: the substitution of the advice of a (presumably more accurate) algorithm for our own judgment.[10] The evidence indicates that those people who are more apt to defer also tend to do worse on the CRT. I suggest that this is no coincidence: they do worse, in part, *because* they tend to defer.

The rationality of deference

There is ongoing debate about whether political ideology is correlated with, or caused by, cognitive dispositions that are criticizable (Baron & Jost, 2018; Ditto et al., 2018). The evidence that social (but not economic) conservatives score worse on the CRT might be thought to provide indirect evidence that conservatism is caused by or causes bad thinking. If the worse performance of conservatives arises from a stronger disposition to defer, however, that would be a mistake: epistemic deference is often rational.

Knowledge is very significantly a social product (Rabb et al., 2019; Levy, 2022) and both liberals and conservatives must and should defer to epistemic authorities[. Given that much of what we know we know by testimony, rather than direct observation, and much of our knowledge can only be generated by cooperating (and sometimes conflicting) individuals and

groups, not by individuals on their own, deference is rational. In science, for example, an individual cannot make discoveries on her own: she relies on instruments and techniques designed by others and which she may not fully understand; on teams of other researchers to gather data and interpret it; and on networks of (pre- and post-publication) peer review to detect weaknesses and ambiguities in her work. The question facing us, as rational individuals, is not *whether* to defer, but *when* and *to whom* to defer.

In this light, conservative deference and liberal suspicion represent different strategies for information processing, the success of which depends on the environment in which they are deployed.[11] In an environment in which implicit communications are reliable – in which defaults are selected so that genuinely good options are made salient, in which questions are worded so that incorrect responses are not prompted, in which order tends to correlate with quality, and so forth – the conservative strategy is adaptive. It saves on cognitive resources and time, and allows for the identification of correct responses to problems that may be intrinsically difficult. As we saw (Patel et al., 2019), coming to the right answer on CRT items is surprisingly difficult for many people, but the difficulty of that task pales into insignificance compared to the difficulty of identifying the optimal savings rate, investment plan or diet: we do far better to defer to experts on these topics, assuming the experts are reliable and their advice is reflected in explicit and implicit recommendations.

But it is not simply adaptive to defer to implicit recommendations: it is also rational. Framing effects are often regarded as paradigmatically irrational, on the grounds that how options are framed has a significant effect on whether they are preferred (e.g., Shafir & LeBoeuf, 2002). Since it is the frame, and not the content of the option, that has altered, this is thought to be irrational. But if framing is communicative, responding to it is rational: it is an appropriate response to the information communicated. It is obviously not irrational to prefer option A to option B because someone has recommended A over option B. The fact that had the testifier advocated B instead the person would have preferred B is no reason to think that accepting testimony is irrational.

Stanovich (2018) has recently emphasized that we increasingly live in epistemically hostile environments.[12] In such environments, in which the correlation between implicit recommendations and good enough actions is not reliable, the conservative strategy will not do as well. Even in such environments, though, the liberal strategy will outperform the conservative strategy only on a narrow range of questions: those in which it is possible for the person to do better by thinking for herself or by actively seeking alternative sources of information. In these environments, we don't do well by deferring uncritically. But we will often not do better by refusing to defer: detecting signal from noise in this environment may often be

beyond us (Levy, 2022). This important social and political question aside, it should be noted that the relative success of the conservative and the liberal strategy at guiding action is somewhat independent of the question whether deference to testimony is rational. A strategy may be maladaptive without being irrational; in many cases, it is rational to accept testimony from others, even though this leaves agents vulnerable to exploitation by liars.

If it is true that the conservative strategy tends to do better in many environments, and does little worse in others, we have few grounds for criticism of those who deploy it. Indeed, given our reliance on deference and the social production of knowledge, there is a strong case for thinking we ought to respond to the fact that the conservative strategy is often unreliable not by criticizing those who deploy it, but by changing the environment in which it is deployed so that it is more reliable: regulating the epistemic environment so that implicit recommendations are more reliable than they currently tend to be.[13]

Are lures required for poor performance?

This proposal faces an important objection.[14] I have argued that bad responses on the CRT are explained (in important part) by a disposition to defer to testimony. Because CRT items make salient a particular (wrong) response, they implicitly recommend that response and those people more apt to defer to testimony respond accordingly. But while the standard CRT makes wrong responses salient, nonstandard versions need not. Moreover, there have recently been claims that nonstandard versions, without 'lures', correlate with other measures of thinking styles and capacity to the same degree as do the standard versions (Baron et al., 2015; Attali & Bar-Hillel, 2020). If this evidence is robust, then performance on the CRT is not explained by a disposition to defer. If performance on CRT items with and without lures is equally predictive of thinking styles, then it cannot be a disposition that is cued only by the first set that underlies performance on either.

It is worth emphasizing that this evidence is inconsistent with the standard interpretation of the CRT. On the standard view, good performance on the CRT depends, in part, on inhibiting the intuitive response the tasks make salient (Campitelli & Gerrans, 2014; Travers et al., 2016). Equally, there are some CRT-like items that do not principally involve numeracy (Thomson & Oppenheimer, 2016). The challenge to the standard view that Baron et al. and Attali and Bar-Hillel present is a major one, and must be taken seriously. But siding with the consensus remains, for the moment, reasonable. Much more evidence is needed before we reject the view that lures explain a large proportion of variance in responses. Because the

heterodox views may turn out to be correct, however, I devote the rest of this section to discussing what remain of my proposals on the assumption it is vindicated.

Even if these heterodox views are correct, it remains plausible that performance on CRT items (and mathematical items, and so on) with lures is at least partly explained by reference to implicit testimony. But if items without lures are equally predictive of thinking styles and capacity, the way in which lures render a particular wrong salient is secondary to the explanation of performance on the CRT. It would seem that poor performers respond to implicit testimony when it is available, but in its absence, they continue to be poor performers, with the same dispositions underlying performance on both items with and those without lures. The evidence of performance on items without lures seems to rule out an explanation of poor performance according to which it is *explained by* the salience of implicit testimony (e.g., by the salience of lures preempting any further processing).

Even so, it remains possible that performance on CRT items with and without lures is directly rational. On standard views, poor performance is explained by some bias or irrational disposition: say a disposition to rely on intuition over reasoning, or (on Baron's own view), on impulsivity. But poor performance might instead be at least partly due to a lack of mindware (Stanovich, 2018); a lack, that is, of acquired tools for thinking. A lack of mindware is not an indication of irrationality: an ideally rational thinker cannot compensate through sheer thinking power for tools that must be developed and effortfully acquired (conversely, being able to use mindware effectively is a good, though not perfect, index of thinking ability because it is often somewhat difficult to acquire mindware, and its intelligent deployment often requires understanding the conceptual structure of problems).

Evidence that poor CRT performance is partly due to a lack of mindware is plentiful. First, there is evidence that repeated exposure to the CRT (Meyer et al., 2018) and even ongoing accuracy feedback on multiple versions of the same type of CRT (Janssen et al., 2020) fails to improve performance for most participants. Feedback does increase reaction times, but without a corresponding increase in accuracy. Conversely, brief explanation of the structure of the problem leads to a significant boost in performance (Boissin et al., 2021; Hoover & Healy, 2017). Knowing *that* one is mistaken does not lead to an improvement in response, unless one also knows *how* one is going wrong. This strongly suggests that good performance is often partly due to the availability of the requisite mindware, whereas poor performance is explained by its unavailability, either because the arithmetical principles are rarely used and therefore inaccessible or because they were never acquired in the first place. It is noteworthy that many poor performers give covert indications that they are

aware that they are failing to solve the problem, as indicated by an increase in response latencies (Bago & De Neys, 2019) and a decrease in confidence (De Neys et al., 2013). Even in the absence of feedback they may be aware that their responses are unsatisfactory, but they lack the mindware to improve them.

When implicit testimony is available, poor performers defer to it. So doing is rational because those who use it have every reason to be guided by the evidence available to them. They may be aware they cannot solve the problems on their own, so they defer to the testimony offered to them. Implicit testimony is evidence for everyone, but only for those with little else to go on is it decisive (similarly, only for low information voters is the ballot order effect – which also functions by offering implicit testimony – decisive; Pasek et al., 2014). When it is *not* available, they may resort to the best strategy that remains available to them, which may be nothing better than guessing. Either way, they're responding rationally: using the thinking tools they have as well as they can, given the information available to them.

Baron et al. might therefore be correct that the salience of lures does not explain poor performance, while it remains true that the CRT does not measure a lack of rationality. A lack of mindware is not a lack of reasoning ability. This suggestion would also explain why CRT items with and without lures are equally predictive of thinking styles and other dispositions. Both sets of problems are deceptively difficult, especially for those who lack the mindware to solve them (of course some participants lack the ability to acquire or to deploy the mindware too). In both cases, agents use what evidence they have available to make up for their shortfall in mindware. They may have developed a disposition to rely on implicit testimony to compensate for their lack. Conversely, they may not have acquired the mindware due to a commitment to deference. In either case, it is their use of evidence or of the next best strategy available to them that explains performance, not a lack of ability.

This proposal could be further developed in one of two ways. On the first, the disposition to defer would play a secondary role. The agent might engage in two separate processing stages: first, they assess the problem as beyond their capacities; only subsequently do they seek an alternative strategy, such as deferring to implicit recommendations or simply guessing when no such recommendation is available. If that suggestion were accurate, it would not be true that a disposition to defer underlies performance: rather, this disposition would be a factor that is relevant to the actual response given, but not a factor that explains poor performance. On an alternative proposal, however, the disposition to defer would remain an important part of the explanation of poor performance.

On this proposal, the disposition to defer would be active from the start: the agent is sensitive to any testimonial evidence embedded in the items from the outset. When it is available, they respond to it at the first stage of processing, attempting to integrate this evidence with the first-order evidence they are able to glean from the problem structure. This proposal is more plausible than the two-stage proposal, because the two-stage proposal predicts longer reaction times for poor performers; this does not seem to be the case. It is also more plausible because it is unlikely that anyone, poor performer or not, ignores available evidence. Higher-order evidence like testimony is evidence, and a well-designed cognizer is sensitive to all the evidence they are able to process.

It is worth saying a little more about why we ought to expect the correlations Baron et al. (2015) report, on the proposal just sketched. They report a negative correlation with utilitarian thinking and with 'active open-minded thinking' (AOT). Neither is surprising. I have suggested that the disposition to defer is stronger in conservatives than liberals. As Baron et al. themselves recognize, a preference for AOT is also associated with liberals. Indeed, AOT items seem well-designed as probes of current political allegiances, rather than of thinking dispositions. People are, in the main, well aware that being open to persuasion and taking into account a wide range of evidence are (supposedly) characteristics on which liberals claim to distinguish themselves. Responses to AOT items may be expressive, rather than sincere reports of dispositions. Of course, we have already seen evidence that liberals (and economic but not social conservatives) are also are less condemning of actions that are usually judged to be immoral; (Pennycook et al., 2014b; Royzman et al., 2014).

Conclusion

The CRT is widely held to be an index of good thinking. Bad performance on it is often held to arise from difficulty in inhibiting intuitive thoughts, cognitive laziness or low numeracy. It is true that hitting upon the right solution to CRT items is cognitively demanding, and that those who succeed thereby exhibit capacities that are valuable. But many of those who perform badly may not thereby manifest criticizable cognitive dispositions or a lack of capacity. I suggest that they defer to the implicit testimony encoded in CRT items.

The hypothesis advanced in this paper swims against a strong tide in cognitive science, which tends to suggest that we are much less rational than we had hoped. To that extent, it joins forces not only with Gigerenzer and other proponents of ecological rationality, but also with those thinkers urging a rethink of specific tasks alleged to demonstrate widespread irrationality (for instance, Hedden (2019) on hindsight bias, Oaksford and

Chater (2001) on the Wason selection task and Crupi et al. (2007) on the conjunction fallacy). Most of these approaches differ from that urged here inasmuch as they aim to vindicate the rationality of human beings by suggesting that they use strategies that are appropriate to the kinds of situations we regularly face, *rather than* the kinds of rules captured by standard logic. Stenning and van Lambalgen (2008), for instance, argue that subjects performing the Wason selection task interpret the problem as involving conditionals that allow for exceptions, rather than as the conditionals of first-order logic, while Oaksford and Chater (1994) suggest that participants understand it as an inductive, rather than a deductive task. Similarly, Crupi et al. (2007) hold that the conjunction fallacy arises when participants substitute assessments of confirmation for probability judgments. In one way or another, all of these are attribute substitution theories: they maintain that participants who fail at the task do so because they substitute a different task for the one the experimenters intend to demand of them.

The hypothesis put forward here is different. It turns not on the substitution of a different task for the one that is actually intended, but a different weighting of information the questions contain in an attempt to solve it. Given that participants find it hard to solve the original task on the basis of the logical form, they turn to the implicated testimony to solve it. They use a rational *strategy* to solve the original task. Ecological rationality, recall, is assessed on the basis of outputs: a process is ecologically rational if it is a process that tends to secure epistemic goals in appropriate circumstances. I have argued that bad CRT responses are directly rational (whether or not they are also ecologically rational). They are the output of mechanisms that weigh available evidence in the normatively appropriate manner. The evidence to which they respond consists in the higher-order evidence of the apparent testimony CRT items offer to respondents.

While I have cited experimental and correlational evidence in favor of this hypothesis, the evidence is indirect. High confidence in the hypothesis must await direct evidence. How can the hypothesis be tested? The extent to which testimony is perceived as reliable is open to experimental manipulation. The simplest manipulation would involve the intuitive answer being provided by a source that was perceived as trustworthy versus untrustworthy in a between-subjects experiment. Insofar as people find the problems difficult to solve, however, they may be expected to rely on testimony from the untrustworthy informant as much as the trustworthy: some evidence might be better than none. Competing sources of testimony could be used to control for this worry. We ought also to find that individual differences in the disposition to rely on testimony are independent predictors of generating the intuitive (but not unintuitive) wrong answers on the CRT.

If the hypothesis is borne out, we should see CRT items as implicating, or implicitly recommending for our attention, certain responses. Given that knowledge production is social and much of our knowledge is gained by testimony, deferring to the implicit testimony encoded in these items is adaptive. Moreover, it is arguably rational: an appropriate response to the recommendation implicitly conveyed by the way CRT items make information salient. The dispositions manifested in deferring to this implicit testimony are dispositions that are central to human ecological and cognitive success. Perhaps we ought to focus on ensuring that we live in an epistemic environment in which such dispositions are successful, rather than criticize those who deploy them.[15]

Notes

1. While this kind of response is not uncommon – Gerd Gigerenzer's adaptive toolbox response to the heuristics and biases tradition (Gigerenzer & Selten, 2002; Gigerenzer et al., 2000) turns on it – the claim that standard paradigms are mismatched to the kinds of challenges ordinary agents typically face may be false because hostile agents may deliberately take advantage of our cognitive vulnerabilities. Pressing this objection, Stanovich (2018) argues that what critics see as a bug in these paradigms should actually be regarded as a feature, "because the modern world is, in many ways, becoming hostile for individuals relying solely on miserly processing" (427).
2. For the notion of epistemic hostility, see, (Evans et al., 2013; Stanovich, 2018); and Sterelny (2003).
3. According to Google Scholar, Frederick (2005), which introduced the test, has been cited 4323 times.
4. Doris is here quoting Stanovich (2004). But Stanovich is more careful than Doris (unsurprisingly, in light of the care he takes to correct the identification of type 1 processes with irrationality (De Neys & Pennycook, 2019). Stanovich writes that these processes are "*in some sense* deeply unintelligent" (39; emphasis added); Doris drops the qualification.
5. The second and third of these questions are among the four that constitute Thomson and Oppenheimer's (2016) CRT-2.
6. It's worth flagging a potential worry that might arise at this point. One might expect that being alerted (by features of the context or the content) to the possibility that the question is a trick might implicate that the intuitive answer is unlikely to be correct, and therefore dispose some participants to reject implicit recommendations. Anecdotally, trick questions 'work' for a high proportion of respondents who are aware that the question is a 'trick'; so do CRT items (I've seen items from newer versions of it presented to audiences at professional conferences *as* new CRT items; many audience members still gave the 'intuitive' responses). Perhaps the implicature constituted by the wording functions implicitly whereas the implicature constituted by explicit wording or other features of context or content that alert the person to the fact that this *is* a trick question functions at the personal level. If the person is unable to identify the correct response, the former implicature may be sufficiently strong to determine the response. The process dissociation procedure might shed light on this question by disentangling the contribution of these two independent processes.

7. As noted above, a significant proportion of participants find the right answer on the bat and ball problem intuitive (Bago & De Neys, 2019). This is not, however, evidence that the format of the CRT does not provide implicit testimony guiding us toward the response usually seen as intuitive. Bago and De Neys report evidence (in the form of response latencies and lowered confidence) that these participants find *both* the correct and the incorrect response intuitive. Speculatively, the correct response is intuitive for them because arithmetical reasoning is sufficiently overlearned to be automated, while the incorrect response is intuitive due to the format of the question. Conversely there is evidence that many participants who generate the wrong response on CRT items and other classic reasoning problems (base rate neglect, for example), also find the objectively correct response intuitive (De Neys & Pennycook, 2019). Which response will be offered depends on the strength of often competing intuitions, among other variables (see, Stanovich (2018) for a detailed proposal).
8. Note that the evidence for the association between lower CRT scores and conservatism comes almost entirely from the United States (indeed, Yilmaz and Alper (2019) provide evidence that it is much weaker elsewhere). This should be borne in mind in interpreting the hypothesis.
9. Note that Pennycook et al. (2014b) found that analytic thinking predicted a willingness to accept heterodox opinions independently of political leanings. Thus, while liberalism predicts lower deference to authority, which in turn predicts higher CRT scores there is an independent route that runs directly via analyticity.
10. In contrast, Pennycook and Rand (2019c) found that conservatives were less trusting of mainstream news than liberals, and that those lower in CRT were worse at discerning trustworthy sources. There was an interaction between CRT scores and conservatism: higher CRT scores enabled better discernment for both liberals and conservatives, but the effects were much weaker for conservatives. This suggests that with regard to some sources, those who score lower on the CRT are less, not more, trusting. This evidence is not inconsistent with the claim that conservatives are (on average and slightly) more disposed to rely on deference than liberals. Agents may have a greater disposition to defer without being disposed to defer at all to particular sources. Epistemic vigilance mechanisms may be upstream of those mechanisms involved in deference. Deference is always selective (Sperber et al., 2010); *who* one defers to is a distinct question from how deferential one is.
11. It is important to notice that on the most recent evidence available on this correlation (Pennycook & Rand, 2019a) the effect sizes in question are actually quite small – in the order of $r = .11$ or even less – so talk of a conservative strategy versus a liberal strategy is somewhat misleading. Moreover, the disposition to defer may be continuous: it might be better to talk about the adaptive success of more versus less deferential strategies.
12. It is worth saying a word about the relationship between explanations that appeal to the notion of an epistemically hostile environment and the kind of explanation of poor performance on heuristics and biases tasks offered by Gigerenzer and other proponents of ecological rationality. On both, a lack of fit between our cognitive processing mechanisms and the task explains why we go wrong. But accounts of ecological rationality postulate mechanisms that are fast and frugal and ignore a great deal of information. There are grounds for thinking of these mechanisms as adaptive, rather than rational: they generate the right answers, in appropriate environments, using cues

that may not be genuine reasons. My account holds that the relevant mechanisms are genuine reasoning mechanisms, responding appropriately to the information they are presented with; they go wrong because the evidence is misleading.

13. Of course, the best strategy would be to defer when and only when we are in contexts in which testimony can be expected to be reliable. It should be emphasized that we are all somewhat discerning consumers of testimony: we filter testimony by reference to its plausibility, and by reference to cues for competence and reliability (Harris, 2012; Mascaro & Sperber, 2009; Sperber et al., 2010). But we operate under ineliminable constraints. First, greater epistemic vigilance risks too many false negatives; given the extent of our epistemic reliance on others, this is a cost that might be devastating. Second, strategies that focus on honing individual capacities, rather than policing the epistemic environment, are vulnerable, because bad epistemic actors can usually mimic whatever cues to reliability individuals use.
14. I thank a reviewer for this journal (who identified himself as Jonathan Baron) for highlighting this objection and making me see how important it is.
15. I am grateful to two reviewers for *Philosophical Psychology*, who identified themselves as Jonathan Baron and Wim De Neys, for helpful and challenging comments on this paper. .

Disclosure statement

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

Funding

This work was supported by the Australian Research Council [DP180102384].

Notes on contributor

Neil Levy is professor of philosophy at Macquarie University, Sydney, and a Senior Research Fellow at the Uehiro Centre for Practical Ethics, University of Oxford.

ORCID

Neil Levy http://orcid.org/0000-0002-5679-1986

References

Attali, Y., & Bar-Hillel, M. (2020). The false allure of fast lures. *Judgment and Decision Making*, *15*(1), 93–111. http://journal.sjdm.org/19/191217/jdm191217.pdf

Bago, B., & De Neys, W. (2019). The smart system 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, *25*(3), 257–299. https://doi.org/10.1080/13546783.2018.1507949

Baron, J., & Jost, J. T. (2018). False equivalence: Are liberals and conservatives in the US equally "biased? *Perspectives on Psychological Science*. 14(2), 292–303. doi:10.1177/1745691618788876 .

Baron, J., Scott, S., Fincher, K., & Emlen Metz, S. (2015). Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, *4*(3), 265–284. https://doi.org/10.1016/j.jarmac.2014.09.003

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that Smartphones are used to supplant thinking. *Computers in Human Behavior*, *48*, 473–480. https://doi.org/10.1016/j.chb.2015.02.029

Boissin, E., Caparos, S., Raoelison, M., & De Neys, W. (2021). From bias to sound intuiting: Boosting correct intuitive reasoning. *Cognition*. June , 211. https://doi.org/10.1016/j.cognition.2021.104645

Bovens, L. (2008). The ethics of nudge. In M. J. Hansson & T. Grüne-Yanoff (Eds.), *Preference change: Approaches from philosophy, economics and psychology* (pp. 207–220). Springer, Theory and Decision Library A.

Bucher, T., Collins, C., Rollo, M. E., McCaffrey, T. A., Vlieger, N. D., der Bend, D. V., Truby, H., & Perez-Cueto, F. J. A. (2016). Nudging consumers towards healthier choices: A systematic review of positional influences on food choice. *British Journal of Nutrition*, *115*(12), 2252–2263. https://doi.org/10.1017/S0007114516001653

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, *42*(3), 434–447. https://doi.org/10.3758/s13421-013-0367-9

Carlin, B. I., Gervais, S., & Manso, G. (2013). Libertarian paternalism, information production, and financial decision making. *The Review of Financial Studies*, *26*(9), 2204–2228. https://doi.org/10.1093/rfs/hht025

Coady, C. A. J. (1992). *Testimony: A Philosophical Study*. Clarendon Press.

Crupi, V., Fitelson, B., & Tentori, K. (2007). Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning*, *14*(2), 182–199. https://doi.org/10.1080/13546780701643406

Darcy, R., & McAllister, I. (1990). Ballot position effects. *Electoral Studies*, *9*(1), 5–17. https://doi.org/10.1016/0261-3794(90)90038-A

Davidai, S., Gilovich, T., & Ross, L. D. (2012). The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences*, *109*(38), 15201–15205. https://doi.org/10.1073/pnas.1211695109

De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, *28*(5), 503–509. https://doi.org/10.1177/0963721419855658

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, *20*(2), 269–273. https://doi.org/10.3758/s13423-013-0384-5

Deppe, K. D., Gonzalez, F. J., Neiman, J. L., Jacobs, C., Pahlke, J., Smith, K. B., & Hibbing, J. R. (2015). Reflective liberals and intuitive conservatives: A look at the cognitive reflection test and ideology. *Judgment and Decision Making*, *10*(4), 18. http://journal.sjdm.org/15/15311/jdm15311.html

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., Celniker, J. B., & Zinger, J. F. (2018). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*, 14(2), 273–291. doi:10.1177/1745691617746796 .

Doris, J. M. (2015). *Talking to our selves: Reflection, ignorance, and agency*. Oxford University Press.

Doris, J. M. (2018). Précis of talking to our selves: Reflection, ignorance, and agency. *Behavioral and Brain Sciences*, *41*, 1–12. https://doi.org/10.1017/S0140525X16002016

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390–405. https://doi.org/10.1037//0022-3514.71.2.390

Evans, J. S.,T.B., . (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. S., T, B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gervais, W. M. (2015). Override the controversy: Analytic thinking predicts endorsement of evolution. *Cognition*, *142*(September) , 312–321. https://doi.org/10.1016/j.cognition.2015.05.011

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT Press.

Gigerenzer, G., Todd, P. M., & Group, A. R. (2000). *Simple heuristics that make us smart* (1 ed.). Oxford University Press.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Green, M. S. Direct reference and implicature. (1998). *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *91*(1), 61–90. JSTOR. https://doi.org/10.1023/A:1004212614842

Grice, H. P. (1991). *Studies in the way of words*. Harvard University Press.

Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public views on policies involving nudges. *Review of Philosophy and Psychology*, *6*(3), 439–453. https://doi.org/10.1007/s13164-015-0263-2

Haigh, M. (2016). Has the standard cognitive reflection test become a victim of its own success? *Advances in Cognitive Psychology*, *12*(3), 145–149. https://doi.org/10.5709/acp-0193-5

Harris, P. (2012). *Trusting what you're told*. Harvard University Press.

Hedden, B. (2019). Hindsight bias is not a bias. *Analysis*, *79*(1), 43–52. https://doi.org/10.1093/analys/any023

Hoover, J. D., & Healy, A. F. (2017). Algebraic reasoning and bat-and-ball problem variants: Solving isomorphic algebra first facilitates problem solving later. *Psychonomic Bulletin & Review*, *24*(6), 1922–1928. https://doi.org/10.3758/s13423-017-1241-8

Janssen, E. M., Raoelison, M., & de Neys, W. (2020). "You're wrong!": The impact of accuracy feedback on the bat-and-ball problem. *Acta Psychologica*, *206*(May) , 103042. https://doi.org/10.1016/j.actpsy.2020.103042

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 18. http://journal.sjdm.org/13/13313/jdm13313.pdf

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin UK.

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

King, A., & Leigh, A. (2009). Are ballot order effects heterogeneous? *Social Science Quarterly*, *90*(1), 71–87. https://doi.org/10.1111/j.1540-6237.2009.00603.x

Lackey, J., & Sosa, E. (2006). *The epistemology of testimony*. Oxford University Press.

Levy, N. (2014). *Consciousness and Moral Responsibility*. Oxford University Press.

Levy, N. (2022). *Bad Beliefs*. Oxford University Press.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*(3), 367–380. https://doi.org/10.1016/j.cognition.2009.05.012

McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, *17*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2006.01721.x

McPhetres, J. (2018). *What does the cognitive reflection test really measure: A process dissociation investigation*.

Mercier, H. (2017). How gullible are we? A review of the evidence from psychology and social science. *Review of General Psychology*, *21*(2), 103–122. https://doi.org/10.1037/gpr0000111

Meyer, A., Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the cognitive reflection test. *Judgment and Decision Making*, *13*(3), 246–259. http://journal.sjdm.org/18/18228a/jdm18228a.pdf

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608–631. https://doi.org/10.1037/0033-295X.101.4.608

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*(8), 349–357. https://doi.org/10.1016/S1364-6613(00)01699-5

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*(6), 972–987. https://doi.org/10.1037//0022-3514.76.6.972

Pasek, J., Schneider, D., Krosnick, J. A., Tahk, A., Ophir, E., & Milligan, C. (2014). Prevalence and moderators of the candidate name-order effectevidence from statewide general elections in California. *Public Opinion Quarterly*, *78*(2), 416–439. https://doi.org/10.1093/poq/nfu013

Patel, N., Baker, S. G., & Scherer, L. D. (2019). Evaluating the cognitive reflection test as a measure of intuition/reflection, numeracy, and insight problem solving, and the implications for understanding real-world judgments and beliefs. *Journal of Experimental Psychology. General*, *148*(12), 2129–2153. Pagination Specified-No Pagination Specified. https://doi.org/10.1037/xge0000592

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014a). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition*, *42*(1), 1–10. https://doi.org/10.3758/s13421-013-0340-7

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014b). The role of analytic thinking in moral judgements and values. *Thinking & Reasoning*, *20*(2), 188–214. https://doi.org/10.1080/13546783.2013.865000

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, *48*(1), 341–348. https://doi.org/10.3758/s13428-015-0576-1

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*(3), 335–346. https://doi.org/10.1016/j.cognition.2012.03.003

Pennycook, G., & Rand, D. G. (2019a). Cognitive reflection and the 2016 U.S. Presidential election. *Personality & Social Psychology Bulletin*, *45*(2), 224–239. https://doi.org/10.1177/0146167218783192

Pennycook, G., & Rand, D. G. (2019c). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*(July), 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, *107*(Suppl. 2), 8993–8999. https://doi.org/10.1073/pnas.0914630107

Rabb, N., Fernbach, P. M., & Sloman, S. A. (2019). Individual representation in a community of knowledge. *Trends in Cognitive Sciences*, *23*(10), 891–902. https://doi.org/10.1016/j.tics.2019.07.011

Royzman, E. B., Landy, J. F., & Goodwin, G. P. (2014). Are good reasoners more incest-friendly? Trait cognitive reflection predicts selective moralization in a sample of American adults. *Judgment and Decision Making*, *9*(3), 16. http://journal.sjdm.org/14/14226/jdm14226.pdf

Rozin, P., Scott, S., Dingley, M., Urbanek, J. K., Jiang, H., & Kaltenbach, M. (2011). Nudge to obesity I: Minor changes in accessibility decrease food intake. *Judgment and Decision Making*, *6*(4), 10. http://journal.sjdm.org/11/11213/jdm11213.html.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, *53*(1), 491–517. https://doi.org/10.1146/annurev.psych.53.100901.135213

Shea, N., & Frith, C. D. (2016). Dual-process theories and consciousness: The case for 'Type Zero' cognition. *Neuroscience of Consciousness*, *2016*(1), niw005. https://doi.org/10.1093/nc/niw005

Shenhav, A., Greene, J. D., R, D. G., R, D. G., Greene, J. D., Shenhav, A., Greene, J. D., & R, D. G. (2012). Divine intuition: Cognitive style influences belief in god. *Journal of Experimental Psychology. General*, *141*(3), 423–428. https://doi.org/10.1037/a0025391

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*(3), 467–494. https://doi.org/10.1016/j.cognition.2005.11.001

Smith, N. C., Goldstein, D. G., & Johnson, E. J. (2013). Choice without awareness: Ethical and policy implications of defaults. *Journal of Public Policy & Marketing*, *32*(2), 159–172. https://doi.org/10.1509/jppm.10.114

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Stanovich, K. E. (2004). *The robot's rebellion*. University of Chicago Press. https://www.press.uchicago.edu/ucp/books/book/chicago/R/bo3627517.html

Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science*. MIT Press.

Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition* (1 ed.). Wiley-Blackwell.

Sturgeon, S. (1994). Good reasoning and cognitive architecture. *Mind & Language*, *9*(1), 88–101. https://doi.org/10.1111/j.1468-0017.1994.tb00217.x

Talhelm, T., Haidt, J., Oishi, S., Zhang, X., Miao, F., & Chen, S. (2015). Liberals think more analytically (More 'Weird') than conservatives. *Personality & Social Psychology Bulletin*, *41*(2), 250–267. https://doi.org/10.1177/0146167214563672

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. Yale University Press.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140. https://doi.org/10.1016/j.cogpsych.2011.06.001

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113. http://journal.sjdm.org/15/151029/jdm151029.pdf

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275. https://doi.org/10.3758/s13421-011-0104-1

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the cognitive reflection test. *Cognition*, *150*(May) , 109–118. https://doi.org/10.1016/j.cognition.2016.01.015

Way, J. (2017). Reasons as premises of good reasoning. *Pacific Philosophical Quarterly*, *98* (2), 251–270. https://doi.org/10.1111/papq.12135

White, M. (2013). *The manipulation of choice: Ethics and libertarian paternalism*. Palgrave Macmillan US. https://doi.org/10.1057/9781137313577

Wilkinson, T. M. (2013). Thinking harder about nudges. *Journal of Medical Ethics*, *39*(8), 486. https://doi.org/10.1136/medethics-2013-101687

Yilmaz, O., & Alper, S. (2019). The link between intuitive thinking and social conservatism is stronger in WEIRD societies. *Judgment and Decision Making*, *14*(2), 156–169. http://journal.sjdm.org/18/181212/jdm181212.html