

## NUDGE, NUDGE, WINK, WINK: NUDGING IS GIVING REASONS

NEIL LEVY

*University of Oxford and Macquarie University*

Nudges are, roughly, ways of tweaking the context in which agents choose in order to bring them to make choices that are in their own interests. Nudges are controversial: opponents argue that because they bypass our reasoning processes, they threaten our autonomy. Proponents respond that nudging, and therefore this bypassing, is inevitable and pervasive: if we do not nudge ourselves in our own interests, the same bypassing processes will tend to work to our detriment. In this paper, I argue that we should reject the premise common to opponents and proponents: that nudging bypasses our reasoning processes. Rather, well designed nudges present reasons to mechanisms designed to respond to reasons of just that kind. In this light, it is refusing to nudge that threatens our autonomy, by refusing to give us good reasons for action.

**R**OUGHLY, a *nudge* is a way of influencing people to act that works by changing aspects of the “choice architecture” (Thaler & Sunstein 2008: 6)—the context in which agents choose—rather than by giving them explicit reasons, changing their incentives, or removing options. Nudges work by taking advantage of predictable dispositions of human beings to make decisions in ways that are influenced by (apparently) irrelevant features of the environment in which they find themselves. As a framework for public policy, nudges usually aim to solve widespread failures of individuals to act in their own best interests. There are many examples of such failures: for instance, very many people fail to save enough money for a comfortable (or even an adequate) retirement. They may do so despite judging that they ought to save more. There is evidence that they may be nudged to save much more. For example, changing the defaults on the superannuation policies they sign up to on taking up employment changes the savings rate, because people tend to accept the default (see Smith, Goldstein, &

---

**Contact:** Neil Levy <neil.levy@philosophy.ox.ac.uk>

Johnson 2013 for review). Increase the default rate of savings, and people tend to save more. For those who are already in a job, savings rates can be increased in a different way. People are reluctant to see a drop in their take home salaries, and therefore reluctant to increase the percentage of their current pay sequestered in a retirement savings plan. But they are willing to sign up to a program that will see a higher proportion of future salary sequestered, if it is triggered by a raise in their pay, so that they don't see a drop in their take home pay (Thaler & Benartzi 2004).

Nudges seem to have the potential to improve health and welfare by improving decision-making and behavior. But many people—philosophers, psychologists, economists, and ordinary people alike—have expressed a great deal of anxiety about nudges. Thaler and Sunstein call their policy framework *libertarian paternalism*. It is, they concede, *paternalistic*, because it manipulates people in their own best interests. Paternalism requires justification, when it is aimed at the behavior of competent adults. It is, however, *libertarian* because it doesn't close off any options. If you want to reject the default retirement option, you have only to say so (you are provided with an opportunity to do so when you sign up to the plan). It doesn't even burden any options (they claim): there are no penalties for choosing a different option. These facts have failed to reassure many critics. While they accept that nudges are not coercive, the critics allege that they are nevertheless manipulative. When something bypasses our reasoning, as nudges are supposed to, our autonomy as agents is threatened (see, e.g., Bovens 2009; Wilkinson 2013).

As John Doris (2015; 2018) has recently implied, the worry on which the critics fasten is much more general than they themselves seem to recognise. While there may be special concerns that arise from nudges, our behavior is pervasively and significantly influenced by seemingly irrelevant features of the environment, in ways that apparently bypass our reasoning processes and thereby threaten our autonomy. All that nudging adds to the picture is *intentional* bypassing. Indeed, that fact is central to Thaler and Sunstein's reply to the manipulation concern: since such influences are pervasive, and therefore our capacity to reason is bypassed all the time, why worry about one more such influence? Since we are going to be nudged whether anyone nudges us or not, why not put nudges to good use and increase our welfare?

In this paper, I will argue that the concerns about bypassing of reasoning, and therefore about threats to our autonomy, are misplaced. Nudging doesn't bypass our capacity to reason. When they are effective in changing behavior, manipulations of the context of choice typically (though perhaps not invariably) work by giving us reasons. These reasons may not be consciously recognized or responded to by agents, but they are reasons nevertheless, and it is in virtue of being reasons that they alter behavior. The mechanisms that respond to nudges

are reasoning mechanisms, and in most cases, at least, nudges no more bypass reasoning than do philosophical arguments.<sup>1</sup>

I don't have an account of the nature of reasoning in hand, but it may be useful to say a few words about what I think it takes for a mechanism to constitute a reasoning system. I take the link between 'reasons' and 'reasoning' seriously: a reasoning mechanism is one that has the proper function of responding appropriately to reasons. A reason, in turn, is a consideration in favour of (or against) a particular response or a particular doxastic state. To respond appropriately to a reason is to be better or worse disposed toward an action, or to raise or lower one's credence, in a way that reflects the actual force of the reason. Thus, a reasoning mechanism is one that has the proper function of responding to reasons by appropriately changing dispositions or credence (for accounts of "reasons" and "reasoning" in this spirit, see Sturgeon 1994; Hieronymi 2005; and Way 2017). I will argue that the mechanisms addressed by nudges have this function. Moreover, when they are nudged (appropriately), they actually play this role.

## 1. Nudges and Bypassing

The debate over nudges and the way in which they allegedly bypass our reasoning is familiar (see Saghai 2013 for an overview). Rather than rehearse it here, I will illustrate the threat they represent to our autonomy by reference to John Doris's recent work. In *Talking to Ourselves*, Doris (2015) describes a number of different manipulations or features of the (mainly social) environment in which agents find themselves that cause them to act in predictable ways, without them being conscious of the influence. *Talking to Ourselves* builds on the earlier, and very influential, *Lack of Character* (Doris 2002). The earlier book argued that variance in behavior owes more to features of situations than to agents' characters, a fact that appears to threaten the Aristotelian program in ethics. The newer book is concerned not with character, but with what Doris calls 'agency'. Nevertheless, the newer book develops themes from the older: The same kinds of features of context that better predicted behavior than did agents' characters serve as what Doris calls *defeaters* of their agency. Agency, as understood by Doris, is manifested in actions that express people's values.

---

1. In what follows, I will not be very careful in my use of words like "mechanism" (and "process", which I will use more or less synonymously). Introducing greater precision in the terminology would require taking a stance on some controversial issues regarding mechanism individuation. It would also introduce some potentially distracting complications: I would bet that the "mechanisms" to which my loose use of the word refers are sometimes identical to a relatively discrete mental process, and sometimes supervene on a set of such processes. Some "mechanisms" may be cobbled together more or less on the fly. Because I don't think anything important for the discussion here turns on this issue, I prefer to avoid it.

Doris argues that agency in this sense is defeated when the facts that cause cognition or behavior would not be recognized as reasons by the agent, were she to become aware of them (2015: 64–65). Such defeaters are common, and therefore agency is pervasively at risk.

Doris provides many examples of such defeaters in *Talking to Ourselves*. Consider, for illustration, ballot order effects (which serve as Doris's major example in a *Brain and Behavioral Sciences* target article; Doris 2018). Candidates placed at the top of the ballot in elections enjoy a small but significant advantage over candidates lower down the order. In close elections, this advantage may be decisive. Why does this defeat agency? The order in which candidates are listed is typically alphabetical or (in recognition of the existence of the effect) drawn by lot. In neither case does the order in which candidates are listed tend to correlate with their quality. Since being at the top of the ballot is not a reason in favour of voting for the candidate, and would not be thought to be a reason were the agent to recognize its influence, being influenced by ballot order is being influenced in a way that bypasses agency. In Australia, political pundits take a particular ballot order effect into account in predicting the outcome of an election: the Donkey Vote. Australia has compulsory voting, and preferences must be expressed by numbering every candidate. Someone is a donkey voter when they simply number the candidates from 1 to  $n$ , from top to bottom. The donkey vote has been estimated to amount to 0.7% of total votes, which is sufficient to decide some electorates (Richardson 2010).

As the name suggests, we don't think of donkey voting as especially rational. No doubt, donkey voting satisfies Doris's test for being a defeater: donkey voters would not accept candidate order as a reason for voting as they do if they were asked. Some of them would probably report, accurately, that ballot order guided their voting *without* serving as a reason: they voted that way because it was easy, and they simply didn't care about the outcome (ballot order effects are much stronger for low information voters; Pasek et al. 2014), or as a protest against compulsory voting. Some would probably confabulate a reason for voting as they did (no doubt, in Australian elections some mixture of these different factors play a role, with some people choosing candidates at the top of the ballot because order effects lead to them being evaluated more positively, and those lower down simply being numbered in the order they are listed because the voter doesn't have a preference between them).

The ballot order effect illustrates how the processes that are engaged by nudges bypass genuine reasoning. When we prefer a candidate at the top of the ballot *because she is listed first* we are guided by a fact that is not a genuine reason. Since order does not correlate with quality, being guided by that fact is not being guided by reasons. A mechanism that is responsive to candidate order therefore doesn't seem to be a reasoning mechanism: it is not a mechanism that

has the proper function of responding appropriately to genuine considerations in favour of an action or adjusting a credence. Our reasoning is out of the picture.

On the view shared by Doris and both the opponents and proponents of nudging, our agency (in Doris's sense of that word) is constantly being bypassed in this kind of way. Our agency is bypassed because our reflective processes, impressive as they are, have severely limited capacities, and we are largely at the mercy of non-rational—worse “deeply unintelligent” (2005: 50)—processes.<sup>2</sup> Thaler and Sunstein (2008: 37) make much the same claim, saying that nudges take advantage of the fact that we are “somewhat mindless, passive decision makers”. Nudges are addressed to parts of the mind that are non-rational, bypassing genuine reasoning. That's why they're threatening: autonomous agents govern themselves in the light of reasons (that's why it's not problematic to act paternalistically toward children: they don't have the capacity to govern themselves appropriately, so we may permissibly substitute competent guidance by an adult). Because autonomy depends on the capacity to respond to reasons, bypassing our reasoning mechanisms entails subverting our autonomy.

## 2. The Dual Process View

A particular account of the nature of mental processes underlies both the worries about autonomy expressed by opponents of nudges and the optimism of their advocates. On this view (at least as it has traditionally been understood), intelligent information processing is the domain of what are often called type two processes.<sup>3</sup> Such processes are slow, effortful (both in phenomenology and in requiring depletable cognitive resources), at least typically conscious, and flexible. So-called type one processes, on the other hand, are fast, effortless, typically nonconscious, and inflexible. The properties characteristic of each kind of process are thought to entail that the former—which I will from now on refer to as reflective processes—alone are genuinely intelligent.<sup>4</sup>

---

2. Here Doris is quoting Stanovich (2004). Stanovich qualifies the phrase in a way Doris does not: he says they are “*in some sense* deeply unintelligent” (2004: 39; emphasis added).

3. Kumar (2016) notes that recent work on implicit learning greatly complicates this picture. Some unreflective processes satisfy plausible criteria for being intelligent. The lessons he thinks follow for the permissibility of nudges are important, and I will return to his view in later sections.

4. The standard terminology divides human psychology into System One and System Two (Kahneman 2011), or Type One and Type Two. I avoid this terminology for two reasons. First, the ‘System’ language suggests that the distinction is exclusive and exhaustive, picking out natural kinds, whereas in fact there are a number of processes with some of the features of each type (see Carruthers 2012 for discussion). Second, and more importantly, it ignores the fact that reflective cognition is not dependent on a set of mechanisms distinct from unreflective. Rather, reflective cognition is realized by unreflective mechanisms interacting in various ways, with this interaction sustained by attention.

What makes it the case that unreflective processes and the mechanisms that underlie them are unintelligent? We may point to two different properties. First, they are *inflexible*. Take, for example, our (alleged) hyperactive agency detection devices (Barrett 2004). Because agency is so important for us—because agents are crucial threats and opportunities for animals like us—we have mechanisms that are hypersensitive to cues indicative of its presence. We are therefore disposed to see agents, and faces, in the world at the (literal) drop of a hat. Even when we know the cue is not really indicative of agency, the mechanism continues to dispose us to respond as if it were. More generally, these unreflective mechanisms are *encapsulated* (Fodor 1983): that is, they are insensitive to information outside the narrow domain to which they are attuned. Reflective mechanisms, in contrast, are flexible with regard to input type: they may respond to any reason at all. The lack of flexibility of unreflective mechanisms is manifested in their firing in response to considerations that are not genuine reasons (like ballot order, or minimal cues for agency). A genuine reasoning mechanism is one that responds appropriately to the force of genuine reasons; for that, we require the domain-generality that is possessed by reflective processing alone.

The second property of unreflective mechanisms that ensures that they are not intelligent is their lack of fit to contemporary environments. Take the first example of irrational behavior mentioned at the beginning of this paper: under-saving for retirement. Our tendency to value the near term and (hyperbolically) discount future rewards may be explained by a mismatch between the environment to which we are evolutionarily adapted and our current environment. In the (so-called) environment of evolutionary adaptiveness—the Pleistocene environment in which our species emerged—it was typically adaptive to consume resources more or less as soon as they became available. Food couldn't be preserved long, and access was unreliable; hence resources forgone might be permanently lost. Today, however, many of us live in environments in which food is cheap and abundant, and in which resources may be stored indefinitely. In this environment, our evolved disposition toward immediate consumption ill serves us. Again, contrast this mismatch with the properties of reflective mechanisms. Such mechanisms do not remain fixated on considerations that are no longer reasoning giving. Instead, they are capable of updating, as the genuine force of considerations changes. They alone seem to be genuine reasoning mechanisms, responding appropriately to the reason giving force of considerations in favour of actions or credences.

It is in the light of this dual process psychology that nudging is so controversial. Because nudged agents are brought to act by processes that bypass their reasoning mechanisms, their autonomy is undermined. When we nudge others, we do not manifest the respect that is due to a rational agent. We manipulate them instead (albeit for their own good). Advocates of nudges dissent from these

worries only insofar as they claim that we can't realistically hope for better from ourselves. Because reflective processes are an expensive and limited resource, we have to reconcile ourselves to being nudged. Nudging is inevitable (Thaler & Sunstein 2008). If our agency is not bypassed by nudgers, who ensures that our intuitive processes are triggered in ways that work in our interests, it will be bypassed in any case, and often in ways that harm us. This may occur because we are manipulated by others who do not have our interests at heart (such as those who manipulate us into buying products we don't need or into eating foods that are unhealthy). If we are not intentionally manipulated, we will be nudged nevertheless, by chance features of our choice context. Unintentional nudges, too, will often produce behavior contrary to our interests. There may be more ways to go wrong than right in many domains, after all, so randomly distributed nudges will tend to be deleterious nudges. We will be nudged in any case; we might as well put nudging to our advantage.<sup>5</sup>

Proponents of nudges agree with their critics that nudges take advantage of "deeply unintelligent" mechanisms. For them, intelligence enters the picture at the only point it can: in the design of the nudges. They argue that given the pervasiveness of the kinds of processes that nudges target, and the limits of reflective processing, the best we can hope for is a kind of indirect autonomy, through intelligent design of the choice architecture.

I aim to show that both sides are wrong, and we can realistically hope for much more. We can intelligently and directly self-govern *despite* the limitations of conscious deliberation. In fact, that is what we already do, though very imperfectly. The proponents of nudges are right to think that behavior is inevitably governed by intuitive processes, with reflection playing only a limited role. But they, and their opponents, are wrong in thinking that this entails that it is governed by unintelligent processes. Intelligence should not be identified with reflective processes alone. They do not have a monopoly on it. Nor, for that matter, do intuitive processes have a monopoly on *un*intelligence. Both intelligence and unintelligence are distributed across information processing mechanisms.

---

5. Kumar (2016) denies that nudging is inevitable, for two reasons. First, he claims that it is not always true that if we are not nudged intentionally, we will be nudged anyway, because nudging may introduce additional structural influences on choice. Second, he notes that the value of autonomy is not consequentialist, and there is therefore a normative difference between intentional and unintended manipulation. I think both claims are mistaken. Even if it is true that a nudge introduces novel structural constraints, it is plausibly true that in the absence of the nudge, the agent would be (just as significantly) guided by features of the choice architecture. These features might be of a different kind, but the inevitability of nudge claim does not depend on sameness of kind of influence. From the fact (if it is a fact) that the value of autonomy is non-consequentialist, it does not follow that threats to autonomy arise only or even more seriously from intentional manipulation. If autonomous agents govern themselves in the light of reasons, then agents whose reasoning is bypassed have their autonomy undermined. Being governed by others is one way to lack autonomy; another is failing to govern oneself.

### 3. Unreflective Intelligence

Reflective processing is required for the acquisition, and in many cases for the implementation, of the kinds of cognition we regard as paradigmatically rational. Formal and rule-based thought is the domain of such processes: mathematics, formal logic, statistical reasoning all depend on effortfully acquired and (often, though not always) effortfully implemented conscious use of explicitly representable symbols. These are ‘unnatural’ ways of thinking, in McCauley’s (2011) sense of natural, where to be natural is to be such as to be the expected outcome of cognitive maturation. Such rational thought is indeed the domain of reflective cognition.

Perhaps because we tend to identify reasoning with such ‘unnatural’ activities as science and mathematics, or because we tend to identify reasoning with conscious processes, we are likely to overlook the extent to which automatic, uncontrolled or unconscious processes are rational processes. Ironically, despite his opposition to what he calls reflectivism (which identifies agency with the products of conscious deliberation), and his explicit recognition that “the analytic/automatic distinction crosscuts the intelligent/unintelligent distinction” (2018: 2), Doris provides a neat illustration of how powerful is the intuition that conscious deliberation is the province of true reasoning. As mentioned above, Doris’s test for whether something constitutes a defeater of rational agency explicitly makes our intuitions about reasoning decisive: “when the causes of her cognition or behavior would not be recognized by the actor as reasons for that cognition or behavior, were she aware of these causes at the time of performance, these causes are defeaters” (2015: 64–65). In doing so, he makes the conscious judgments of naïve agents decisive. If the threat to agency consisted in a threat to our self-conception alone, then such a test might be appropriate. We might be incorrigible with regard to whether a particular process is consistent with our self-image. But the threat Doris aims to identify is supposed to have dimensions that escape such subjective judgments: the subjective test is supposed to index whether or not certain attitude-independent facts about human agency prevail.

For Doris, the ballot order effect and related dispositions are supposed to show not only that our conception of ourselves as governed by consciously accessible reasons is false, but also that our behavior is caused by facts that do not constitute reasons. If your vote was determined by the ballot order effect, it had causes “that are not plausibly taken as reasons for that behavior” (2018: 3). Hence (whatever your attitude to it), “your conduct was not appropriately responsive to reasons” (2018: 3). Tapping into the extent to which conduct is guided by agents’ genuine reasons (and not merely into whether they *take* it to be guided by reasons) is required, in turn, because for Doris agency is expressive of values: “behaviors are exercises of agency when they are expressions of the actor’s values” (2018: 7). At



least in principle, an action might express an agent's values despite her failure to recognize that it does (and vice-versa). Hence, it is only if the subjective test really indexes the facts about agency that it can serve Doris's purposes. However, there is no particular reason to think that his is test is reliable. I would bet that the test would fail by its own standards: the considerations to which these processes are responsive, in producing the judgment, often wouldn't be recognized by the agent as reasons, were she aware of them. Rather than rely on a subjective test, I will attempt to assess whether a mechanism manifests intelligence by careful probing of how it operates and the considerations to which it responds, not by consulting our intuitions. Such a direct assessment should be acceptable to Doris, since his subjective test is supposed to tap into the facts about agency.<sup>6</sup>

The argument for the claim that nudges are addressed to reasoning mechanisms will come in two parts. First, I will argue that attention to how nudges work shows that they have the function of making considerations salient to us, and disposing us to respond to them in a way that reflects their actual reasoning force. That is, nudges work by targeting mechanisms that satisfy the conditions for being reasoning mechanisms. I will then turn to the (alleged) features of unreflective mechanisms that are supposed to entail that they are unintelligent: their inflexibility, and the mismatch between the environment for which they are designed and the contemporary world. These two facts together are taken to entail that intuitive processes are atavisms, which systematically generate dispositions to believe and behave which are at odds with the reasons that genuinely prevail. Inflexibility, I will argue, is not distinctive of, or limited to, unreflective processes, and mismatch, though genuine, is no reason to conclude that such processes are not reasoning mechanisms.

### *3.1. Nudges Are Addressed to Reasoning Mechanisms*

Let's begin with Doris's central example in his recent précis (2018) of *Talking to Our Selves*: ballot order effects. Doris highlights this example because it seems obvious that such effects are irrational: nudging us to prefer one candidate over another in this way is bringing us to be influenced by irrelevant facts. The appellation of the best-known ballot order effect as the donkey vote indicates what we usually think of it. Candidate order is allocated alphabetically or (as in Australia)

---

6. I have expanded this discussion of Doris's subjective test for agency at the urging of a reviewer for *Ergo*. As the reviewer points out, nudges will routinely fail Doris's test, whether or not they are genuinely addressed to reasoning mechanisms. But since Doris's test is not suitable for the task he sets himself, I don't take that fact as a mark in their disfavour. Perhaps under pressure, Doris would prefer to identify agency with conscious expression of consciously held values, and retain the subjective test. He would need an account of why such expression is valuable, given that it does not underwrite control by reasons.

by lot, so ballot order does not track quality. The property to which we respond is not, in fact, a good reason. But that fact doesn't entail the conclusion that Doris wants: that the mechanism does not have the function of disposing us to respond, appropriately, to reasons. While being at the top of the order may not *be* a good reason, the mechanism may nevertheless be a reasoning mechanism, which happens to misfire in this context.

The ballot order effect probably arises because the order in which candidates are listed is taken to be an implicit recommendation (Gigerenzer 2015). There is evidence that people do tend to see options that have been made salient to them as having been recommended to them (see McKenzie, Liersch, & Finkelstein 2006 for supportive data). But being guided by recommendations is rational. We can generalize this recommendation paradigm: canonical nudges and supposedly irrational heuristics and biases very often work by functioning as implicit recommendations. Take framing effects. They are often cited as paradigmatically irrational, on the grounds that they cause us to have conflicting preferences over options depending on how they are presented (e.g., Shafir & LeBoeuf 2002). But if the choice of frame is understood as an implicit recommendation (Gigerenzer 2015), there is nothing irrational about reliance on it. No one ever argued that it is irrational to choose a recommended option on the basis that if exactly the same option hadn't been recommended, it wouldn't have been chosen. If it is rational to be guided in our judgments by testimony (and it surely is, under many conditions), it is no less rational to be guided by implicit recommendations (of course, we would worry if these implicit recommendations disposed us to ignore countervailing considerations, but there is no evidence that they do: recall that the ballot order effect is strongest for low information voters). Setting defaults or framing options is communicative: people frame options in ways that highlight particular choices because they take them to be good ones, and their communicative intent is recognized by those who respond to the framing.

So even the central examples invoked in the debate over nudging and bypassing—the ballot order effect, the selection of defaults, framing, anchoring effects—actually have the properties we expect of reasoning mechanisms: making particular considerations salient to us and disposing us to respond appropriately to them. However, this may not be sufficient to show that nudges are addressed to reasoning mechanisms. A number of theorists have suggested that nudges work by *combining* reason-giving with non-rational influence. There is extensive evidence that under cognitive load (when processing resources are scarce, for instance because the person is required to multi-task, or is fatigued or stressed), agents are more powerfully influenced by heuristics and biases (e.g., Gilbert & Osborne 1989; Krull & Erikson 1995). The fact that our susceptibility to heuristics and biases increases under load seems to indicate that they work (in part at least) by taking advantage of cognitive laziness or the fact that it is tempo-

rarily too difficult for us to make a decision. Since nudges harness the heuristics and biases, that suggests that at minimum they *partially* bypass our reasoning. Ansher et al. (2014) make the point explicit: changing defaults at once provides recommendations to agents *and* takes advantage of non-rational dispositions.<sup>7</sup>

However, the fact that susceptibility to defaults increases under load isn't a compelling reason to think they work by bypassing our reasoning, even in part. Once again, it is helpful to think of setting defaults (and framing and anchoring effects) as functioning as implicit testimony. There is nothing irrational about putting more weight on testimony when we lack the resources to assess a claim for ourselves. We all accept that I ought to place more weight on your testimony when you are more expert in the relevant domain than I am. We might think of load analogously: while I am under load, I should give greater weight to testimony because I am temporarily less expert. Nor is the other piece of evidence Ashner et al. cite to show that defaults function in part as non-rational status quo effects persuasive. In the study they cite, pulmonologists were asked whether they would prescribe a CT for a patient. In the control condition, 54% of pulmonologists ordered the scan. That establishes a baseline: given the symptoms described, roughly half will think a scan is warranted. In the other condition, participants were told that a scan had already been ordered but not yet performed. In this condition, only 29% of physicians cancelled the scan. The original authors of this study (Aberegg, Haponik, & Terry 2005), and following them Ansher et al. suggest that a mere (non-rational) bias drives the difference between conditions, for the following reason: "clinical information should dictate whether or not a CT scan should be performed [ . . . ] whether or not it has been ordered or discontinued by the emergency department physician should be irrelevant" (Aberegg, Haponik, & Terry 2005: 1499). But that's false. Clinical information is of course an important piece of evidence, but the decisions of other qualified people are also evidence. The attitudes formed by other physicians is higher-order evidence concerning the correct attitude to take to the case at hand. There is nothing irrational about being guided by such evidence.<sup>8</sup>

---

7. I am grateful to a reviewer for this journal for raising this issue, as well as bringing the paper by Ansher et al. to my attention.

8. Philosophers acquainted with the debate on the epistemic significance of disagreement will recognize this kind of experiment as parallel to the thought experiments that populate this literature. While some philosophers defend positions according to which the higher-order evidence represented by dissenting peers should not influence our credences (e.g., Kelly 2005), many other defend a conciliatory view, according to which we should give significant weight to such higher-order evidence (see Matheson 2015 for discussion and defence of the equal weight view). *At best*, then, the claim that it is irrational for the physicians to give weight to the opinions of other experts is controversial. The case presented is not a case of actual disagreement—physicians do not first form a view and then learn of a dissenting position—and it is not apparent whether those philosophers who defend a steadfast view would embrace the suggestion that higher-order evi-

Far from showing that defaults guide behavior in part by taking advantage of irrational biases, working through these examples strengthens the case for seeing them as functioning as evidence. Nor are these isolated examples. Kumar (2016) identifies nudges (proper) with manipulations of choice architecture that take advantage of Kahneman (2011) style heuristics and biases. But in most, if not all cases, these heuristics and biases have the proper function of making us appropriately sensitive to genuine reasons (and, indeed, routinely play this role, even in our current environment). There are, for example, compelling arguments that the hindsight bias is not a bias (Hedden 2019), and that the confirmation bias is an adaptation for reasoning (Levy 2019). There is a good reason for this: these mechanisms are designed to make us respond adaptively to considerations, and—at least in general—it is adaptive to respond to considerations in ways that follow their actual reason-giving force. That is not to say, of course, that these dispositions never mislead us; that a mechanism has the proper function of making us sensitive to reasons doesn't entail that it always plays that role. We will return to this question with regard to the heuristics and biases in the next section.

In light of this evidence, we ought, therefore, to be much less impressed than we commonly are by the fact that our paradigms of reasoning involve effortful processing over explicit representations. A range of data indicates that unreflective processes may embody intelligence. Kumar (2016) cites extensive evidence that some unreflective processes embody sophisticated learning biases that are sensitive to statistical regularities. For instance, unreflective processes implement Bayesian reasoning optimally, whereas we are notoriously bad at conscious probabilistic reasoning tasks; even nonhuman animals, including many species that are usually supposed to lack reflective processing capacities altogether, appear to update in conformity with Bayesian principles; Valone 2006, and see Brownstein 2018 for further evidence that implicit processes are responsive to personal-level goals). These unreflective processes surely satisfy plausible tests for manifesting intelligence.

In fact, there is extensive evidence that reflective processing often *lowers* decision quality and accuracy, relative to these kinds of unreflective mechanisms. For instance, Halberstadt and Levine (1999) provide evidence that thinking about reasons lowers the accuracy of predictions about the outcomes of basketball games, compared to the production of intuitive judgments. Wilson and Schooler (1991) provide evidence that thinking about reasons made participants' judgments about jams less likely to match those of experts. The capacity for intuitive judgments in some domains seems to be learnable, and in those domains expert judgment is better when made immediately compared to after reflection. Johnson and Raab (2003), for example, present evidence that expert handball players

---

dence should not guide initial decision-making, so it may be that even they would not accept the irrationality claim.

choose worse options when given more time to reflect than when asked to respond immediately. Reviewing a number of studies, Plessner and Czenna conclude that “the assumption of a general superiority of analytic processes proves to be wrong even in the domain of judgments about matters of fact” (2008: 253).

Some psychologists have gone further, arguing that not only does reflective processing often lower decision quality, relative to immediate judgment, but that unreflective processing may raise the quality of the decision, relative to both reflective processing and immediate judgement, even in domains for which we are unlikely to have either innate or acquired learning algorithms. The principal advocate of this view is Ap Dijksterhuis (Dijksterhuis, Bos, Nordgren, & Baaren 2006). On his view, when there are multiple competing considerations to juggle, the higher capacity of nonconscious processing produces better decisions than both conscious deliberation and immediate judgment. Under these conditions, people make objectively better decisions (Dijksterhuis 2004; Dijksterhuis et al. 2006), are more satisfied with their decisions (Dijksterhuis & van Olden 2006), and are more consistent in their judgments (Nordgren & Dijksterhuis 2009). Interestingly, he also produces evidence that conscious deliberation *increases* the influence of the availability heuristic (Dijksterhuis et al. 2008: 94).

Dijksterhuis’s work is controversial. A number of studies comparing immediate judgment, nonconscious deliberation (operationalized by a distraction condition) and conscious reflection have failed to find an advantage of nonconscious thought over immediate judgment (Payne, Samper, Bettman, & Luce 2008; Waroquier, Marchiori, Klein, & Cleeremans 2010). However, we need not accept Dijksterhuis’s central claim for his evidence to help break the grip of reflective processing on our imaginations: even if he is wrong that unreflective processing is *better* than reflective in domains in which we lack learning algorithms, his work does seem to demonstrate that reflective processing is sometimes *worse* than immediate judgment in these domains. In some domains, unreflective mechanisms are intelligent because they are able to integrate a range of information. In others, we may lack such intelligent learning mechanisms, but reflection nevertheless lowers decision quality relative to immediate judgment.

In light of this evidence, there seems little reason to accept the common view that associates reflective processing with intelligence and unreflective with its lack. The intelligent/unintelligent distinction seems to cross-cut the reflective/unreflective distinction (Kumar 2016). How can we square the apparent intelligence of unreflective mechanisms with evidence of their inflexibility?

### 3.2. *Intuitive, Inflexible?*

As Kumar (2016) emphasises and we noted in the previous section, some unreflective processes manifest considerable evidence of flexibility. This fact is cen-

tral to his distinction between “nudges”, which take advantage of heuristics and biases and bypass our reasoning, and “bumps”, which instead work by addressing our sophisticated implicit reasoning capacities. The distinction between implicit processes Kumar draws is genuine. Only some manipulations of the choice architecture target heuristics and biases, and only some work via mechanisms that are to some significant degree encapsulated. It doesn’t follow, however, that there is a normatively significant distinction to be drawn between nudges and bumps, or that nudges addressed to heuristics and biases are addressed to inflexible mechanisms.

The flexibility of response characteristic of reflective processing is not the product of unencapsulated central cognition. There is no such thing. Rather, it is realised by otherwise inflexible mechanisms working in concert: reflective deliberation works through cycles of repeated querying of intuitive processes (Carruthers 2006; Levy 2014). It is true that reflective processing is often more flexible than unreflective. When reflective processing is engaged, attention is (typically) sustained and the outputs of intuitive processes are made available to yet further processes. Without such processing, fewer such mechanisms tend to be engaged for a shorter period of time, and as a result reasoning may reflect a narrower range of the agent’s beliefs, desires, and other attitudes. But nudges are typically processed consciously, and therefore attentional mechanisms are engaged (attention need not be directed *to* the nudge for the mechanisms to play their role; Baars 2002). Nudges *never* engage only a single mechanism, and we should therefore expect a considerable degree of flexibility in the processing of their contents—at least typically sufficient flexibility for them to count as reasoning mechanisms.

Just as importantly, it doesn’t follow from the fact that sometimes unreflective mechanisms produce outputs that are encapsulated against domain-general reasoning that they are especially irrational, compared to reflective processes. True, unreflective mechanisms are sometimes responsible for cognitive ‘seemings’ that the agent knows to be misleading. Such seemings are recalcitrant to domain-general knowledge: *that the snake is made of rubber, that the lines are the same length*, and so on. But this kind of resistance to update is not specific to unreflective mechanisms. Consider Huck Finn cases, for example. In Twain’s novel, famously, Huck knows that Jim the escaped slave is a human being, who in virtue of that fact deserves the same rights and the same dignity as anyone else. But he finds compelling the argument for the view that Jim is property and he therefore wrongs his ‘owner’ in helping Jim to escape. He knows that deliberation is misleading here, but this personal-level knowledge is powerless to render the output of deliberative processes any less compelling for him.

While Huck Finn cases are unusual, cases in which we find arguments powerful even though we know their conclusions are false are common. Philoso-

phy has long specialized in generating such arguments (consider brain-in-a-vat skepticism), but they are a routine feature of ordinary epistemic lives. Most of us are unable to counter the arguments of moderately sophisticated climate science ‘sceptics’, for example, since we lack the scientific expertise. We may be confident that the premises are true, and we may be unable to find fault with the argument itself, but we may be very confident—in fact we may *know* - that the conclusion is false (see Fantl 2018 for a book-length discussion of cases like this, in which we retain knowledge despite our inability to counter the argument). Again, recalcitrance of the output to personal-level knowledge is not distinctive of unreflective processes alone.

It might be responded that reflective processing is rational even when its outputs are recalcitrant to domain-general knowledge. Consider the influential default interventionist model of rational cognition (Evans & Stanovich 2013). On this model, the role of reflective processing is to intervene in reasoning, to correct the outputs of unreflective mechanisms. This model accepts that sometimes the outputs will continue to seem compelling to the agent, but the rational agent will override the seeming. As Steven Jay Gould (1988) reports, even when he knows that a particular seeming is the product of the conjunction fallacy, “a little homunculus in my head continues to jump up and down, shouting at me.” This response is plausible, but does nothing to show that nudging is addressed to irrational mechanisms. If the conjunction of recalcitrant output and intervening process is sufficient for reasoning, then nudging may be addressed to reasoning mechanisms. The outputs of nudges are not so compelling that agents cannot ignore them: indeed, when they have good reason to ignore them, they routinely do (recall that the ballot order effect is strong only for low information voters).

In light of this fact, the inflexibility that sometimes arises from unreflective processes should not lead us so quickly to conclude that nudges bypass our reasoning in virtue of being addressed to such processes. Since nudges are never addressed to a single mechanism at a time, we ought to expect some degree of flexibility in their processing. Even when they output a recalcitrant seeming, a broader set of mechanisms allows the agent to ignore or override the seeming. In this respect (too), there is no difference between nudging and the presentation of explicit arguments.

### *3.3. Mismatch to the Contemporary World*

While it may be true that the intuitive processes engaged by nudges have the proper function of prompting appropriate response to genuine reasons, in a manner that qualifies them as reasoning mechanisms, there remains the worry that they may be systematically mismatched to the contemporary world. Due to this mismatch, they might fail to produce judgments or dispositions that answer

to the genuine reason-giving force of the properties to which they respond. Consider our informational context. We are (allegedly) adapted for the small groups in which our ancestors lived on the African savannah: groups of roughly 150 individuals (Dunbar 1996). Today, however, most of us live in giant cities, alongside millions of other people. Moreover, we are constantly exposed to reports of events from across the world. If our cognitive processes are adapted for small group living, and to processing information about events that occur nearby, we ought to expect them to misfire routinely. For example, we might be designed to enter a state of heightened alert in response to reports of violence. In contemporary life, we will be exposed to such reports daily, whenever we watch the news. We may therefore find ourselves in a state of constant vigilance.

There is persuasive evidence that people process information about risks badly, underestimating relatively high probability risks like heart disease, while overestimating much less likely risks like terrorist attacks (Sunstein & Zeckhauser 2011). This implicit assessment is maladaptive, insofar as it leads to underinsuring for health risks, on the one hand, and unjustified support for restrictions on civil liberties, on the other. Our evolved biases plausibly play a role in our attitude to risk: for example, the salience bias—the bias toward information that is striking—and the closely related availability heuristic ensure that terrorist attacks play an obtrusive role in our thinking, out of proportion to their true risk. Plausibly, this dysfunction arises out of a mismatch between the environment for which the relevant mechanisms are designed and the informational environment we currently inhabit.

Overall, I think there is persuasive evidence that maladaptive belief and behavior is often produced by intuitive mechanisms, because they are ill-matched to the inputs they receive. However, it is a mistake to think that this evidence indicates that the intuitive mechanisms are not reasoning mechanisms. Once again, using susceptibility to mismatch as a criterion will fail to distinguish the mechanisms involved in processing nudges from reflective processing. Reflective processes, too, perform badly given inputs of a kind for which they are not designed.

Most obviously, reflective processes work badly when the inputs are false. The adage ‘garbage in, garbage out’ applies as much to these processes as intuitive. The problem of fake news, which—arguably—has had seriously negative consequences for the world over the past few years, is at least partially a problem that arises from the way that reflective processes may be targeted with bad inputs. Less obviously, reflective processes can be nudged. Consider the influence of how decisions are framed (Koralus & Alfano 2017). Agents overlook reasons an action might be *impermissible* when they are asked which actions are *obligatory*, and vice-versa. The questions are designed to encourage reflection, yet the



framing nudges the response in one direction or another.<sup>9</sup> Again, bad inputs will lead to bad outputs, for reflective processes as much as intuitive.

Intuitive processes may indeed often go wrong because the inputs they are asked to handle are a poor match for those to which they are designed to be sensitive. But that failing is not well understood as involving the bypassing of reasoning. Rather, we do better to understand the failing as involving the presentation of bad reasons to these mechanisms. Such mismatches occur all too frequently, because we have allowed corporations, bad intuitions about institutional design, and sheer drift to ensure that we too often make choices in environments in which bad reasons are presented to us—in which, for example, the best options are not implicitly recommended or the wrong properties are made salient to us.

The right response to failures of inputs to match those the mechanisms are designed to respond to is not to avoid nudging. That's impossible, as Thaler and Sunstein have argued. But it's also absurd; as absurd as refusing to give people explicit reasons why some options are better than others. The right response is to ensure that the context in which we choose is better matched to the mechanisms which process information, so that they are presented with adequate reasons for choice. Just as we shouldn't respond to the problem of fake news, say, by concluding that reflective processing is hopeless, but instead by ensuring that reflection is fed more accurate information, so we should respond to the predictable failures of intuitive processes by ensuring that they are fed the kinds of inputs that serve as reasons for them. That's what well designed nudges do: they present reasons to mechanisms designed to respond to reasons of that kind.

#### 4. Nudge in Peace?

Most actual and proposed nudges function by presenting reasons to agents. They often present higher-order evidence, and higher-order evidence is evidence. It is, of course, rational to guide our decisions and our beliefs in the light of evidence. There is no reason to think, therefore, that most nudges bypasses reasoning. Should we conclude that nudging raises no special worries about autonomy or paternalism?

While well-designed nudges do not threaten autonomy, it is certainly possible to nudge in an autonomy threatening way. Perhaps, first, there are nudges that might promote welfare or pro-social behavior *without* presenting agents

---

9. It is plausible that this kind of framing works like other kinds: people treat the framing as an implicit recommendation. When we are nudged into thinking that obligatoriness is the normative dimension that is most significant here, we give it greater weight than when impermissibility is the dimension recommended for our attention.

with reasons. Perhaps priming behavior might be an example: whether priming works by presenting evidence or by simply biasing processing mechanisms is difficult to assess. Second, whether or not such non-rational nudges exist, nudges that are addressed to reasoning mechanisms can be used in an autonomy-threatening manner. We can nudge people by designing the choice architecture so that they are presented with bad reasons: bad defaults, or inappropriate frames, for example. Such nudges don't *bypass* reasoning, anymore than overt lies bypass reasoning. Rather, they subvert autonomy by giving *bad* reasons. They aren't nudges, as classically conceived, insofar as such nudges are thought to bypass reasoning. But they subvert autonomy nevertheless.

In fact, that's exactly how we should think of the features of choice architecture that proponents of nudges aim to change. Bad defaults or badly designed environments subvert autonomy by feeding bad reasons to us. The opponents of nudges are right to worry about threats to autonomy, but they find them in *precisely* the wrong places. It is when we fail to change these features that we subvert autonomy, by giving bad reasons to agents. The mechanisms addressed by nudges are reasoning mechanisms. They may be addressed well or badly: with good reasons or bad. When they are addressed with good reasons, they tend to promote autonomy, for exactly the same reasons why giving people good arguments promotes their autonomy. Conversely, when they are addressed with bad reasons, agents' autonomy tends to be impaired.

## 5. Conclusion

In this paper, I have argued that the worry that nudges bypass reasoning, and therefore threaten agents' autonomy or are unacceptably paternalistic, is misplaced. Rather than bypassing reasons, nudges are addressed to reasoning mechanisms. This is true whether the underlying mechanisms are sophisticated implicit learning mechanisms capable of integrating disparate information, or whether they are the kinds of mechanisms that underlie comparatively simple heuristics and biases. Good nudges aim to substitute genuine reasons for the badly designed cues which cause maladaptive behavior, considerations to which the mechanisms have the proper function of responding. The intuitive mechanisms to which nudges are addressed are reasoning mechanisms: they are disposed to respond as they do because they are attuned to features of the context of choice in just the way such mechanisms are supposed to be. Nudges may subvert autonomy, but that fact doesn't distinguish them from the presentation of explicit reasons. When they subvert autonomy, they do so for the same reasons as bad arguments do.

Rather than there being a special problem with nudging, a reason to worry

that nudging threatens autonomy, good nudges promote autonomy. Indeed, when our agency is undermined by badly designed features of the context of choice, we may have an obligation to nudge. We treat one another respectfully when we ensure that the context in which we choose is one that supports rational agency.

## Acknowledgments

I am grateful to Mark Alfano, audiences at Deakin University and the University of Sydney, and especially two reviewers for *Ergo*, for very helpful comments. I am also grateful to the Australian Research Council and the Wellcome Trust (grant WT104848/Z/14/Z) for their generous support of this research.

## References

- Aberegg Scott K., Edward F. Haponik, and Peter B. Terry (2005). Omission Bias and Decision Making in Pulmonary and Critical Care Medicine. *Chest*, 128(3), 1497–1505.
- Ansher, Cara A., Dan Ariely, Alisa Nagler, Mariah Rudd, Janet Schwartz, and Ankoor Shah (2014). Better Medicine by Default. *Medical Decision Making*, 34(2), 147–158.
- Baars, Bernard J. (2002). The Conscious Access Hypothesis: Origins and Recent Evidence. *Trends in Cognitive Science*, 6(1), 47–52.
- Barrett, Justin L. (2004). *Why Would Anyone Believe in God?* AltaMira.
- Bovens, Luc (2009). The Ethics of Nudge. In Till Grüne-Yanoff and S.O. Hansson (Eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Springer.
- Brownstein, Michael (2018). *The Implicit Mind: Cognitive Architecture, the Self, and Ethics*. Oxford University Press.
- Carruthers, Peter (2006). *The Architecture of Mind*. Oxford University Press.
- Carruthers, Peter (2012). The Fragmentation of Reasoning. In P. Quintanilla (Ed.), *La co-evolución de mente y lenguaje: Ontogénesis y filogénesis*. Fondo Editorial de la Pontificia Universidad Católica del Perú.
- Chater, Nick, Joshua B. Tenenbaum, and Alan Yuille (2006). Probabilistic Models of Cognition: Conceptual Foundations. *Trends in Cognitive Sciences*, 10(7), 335–344.
- Dijksterhuis, Ap (2004). Think Different: The Merits of Unconscious Thought in Preference Development and Decision Making. *Journal of Personality and Social Psychology*, 87(5), 586–598.
- Dijksterhuis, Ap, Maarten W. Bos, Loran F. Nordgren, and Rick B. van Baaren (2006). Complex Choices Better Made Unconsciously? *Science*, 313(5788), 760–761.
- Dijksterhuis, Ap and Zeger van Olden (2006). On the Benefits of Thinking Unconsciously: Unconscious Thought Increases Post-Choice Satisfaction. *Journal of Experimental Social Psychology*, 42(5), 627–631.
- Dijksterhuis, Ap, Rick B. van Baaren, Karin C. A Bongers, Maarten W. Bos, Matthijs L. van Leeuwen, and Andries van der Leij (2008). The Rational Unconscious: Conscious versus Unconscious Thought in Complex Consumer Choice. In Michaela Wänke (Ed.), *Social Psychology of Consumer Behavior* (89–108). Psychology Press.

- Dijksterhuis, Ap and Henk Aarts (2010). Goals, Attention, and (Un)consciousness. *Annual Review of Psychology*, 61(1), 467–490.
- Doris, John M. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge University Press.
- Doris, John M. (2015). *Talking to Our Selves: Reflection, Ignorance, and Agency*. Oxford University Press.
- Doris, John M. (2018). Précis of *Talking to Our Selves: Reflection, Ignorance, and Agency*. *Behavioral and Brain Sciences*, 41, 1–12.
- Dunbar, Robin (1996). *Grooming, Gossip, and the Evolution of Language*. Harvard University Press.
- Evans, Jonathan St. B. T and Keith E. Stanovich (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives in Psychological Science*, 8(3), 223–241.
- Fantl, Jeremy (2018). *The Limitations of the Open Mind*. Oxford University Press.
- Fodor, Jerry (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.
- Frankish, Keith (2010). Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass*, 5(10), 914–926.
- Gigerenzer, Gerd (2015). On the Supposed Evidence for Libertarian Paternalism. *Review of Philosophy and Psychology*, 6(3), 361–383.
- Gilbert Daniel T. and Randall E. Osborne (1989). Thinking Backward: Some Curable and Incurable Consequences of Cognitive Busyness. *Journal of Personality and Social Psychology*, 57(6), 940–949.
- Gould, Stephen J. (1988, August 18). The Streak of Streaks. *The New York Review of Books*. Retrieved from <https://www.nybooks.com/articles/1988/08/18/the-streak-of-streaks/>
- Halberstadt, Jamin B. and Gary M. Levine (1999). Effects of Reasons Analysis on the Accuracy of Predicting Basketball Games. *Journal of Applied Social Psychology*, 29(3), 517–530.
- Hedden, Brian (2019). Hindsight bias is not a bias. *Analysis*, 79(1), 43–52. <https://doi.org/10.1093/analys/anyo23>
- Hieronymi, Pamela (2005). The Wrong Kind of Reason. *Journal of Philosophy*, 102(9), 437–457.
- Johnson, Joseph and Markus Raab (2003). Take the First: Option Generation and Resulting Choices. *Organizational Behavior and Human Decision Processes*, 91(2), 215–229.
- Kahneman, Daniel (2011). *Thinking Fast and Slow*. Allan Lane.
- Kelly, Thomas (2005). The Epistemic Significance of Disagreement. In Tamar Gendler and John Hawthorne (Eds.), *Oxford Studies in Epistemology* (Vol. 1, 167–196). Oxford University Press.
- Koralus, Philipp and Mark Alfano (2017). Reasons-Based Moral Judgment and the Erotetic Theory. In Jean-Francois Bonnefon and Bastien Trémolière (Eds.), *Moral Inferences* (77–106). Routledge.
- Krull Douglas S. and Darin Erickson (1995). Judging Situations: On the Effortful Process of Taking Dispositional Information into Account. *Social Cognition*, 13(4), 417–438.
- Kumar, Victor (2016). Nudges and Bumps. *Georgetown Journal of Law and Public Policy*, 14(Special Issue 2016), 861–876.
- Levy, Neil (2014). *Consciousness and Moral Responsibility*. Oxford University Press.
- Levy, Neil (2019). Due Deference to Denialism: Explaining Ordinary People’s Rejection of Established Scientific Findings. *Synthese*, 196(1), 313–327.
- Matheson, Jonathan (2015). *The Epistemic Significance of Disagreement*. Palgrave Macmillan.

- McCauley, Robert N. (2011). *Why Religion Is Natural and Science Is Not*. Oxford University Press.
- McKenzie, Craig R. M., Michael J. Liersch, and Stacey R. Finkelstein (2006). Recommendations Implicit in Policy Defaults. *Psychological Science*, 17(5), 414–420.
- Mercier, Hugo and Dan Sperber (2017). *The Enigma of Reason*. Harvard University Press.
- Nordgren, Loran F. and Ap Dijksterhuis (2009). The Devil Is in the Deliberation: Thinking Too Much Reduces Preference Consistency. *Journal of Consumer Research*, 36(1), 39–46.
- Pasek, Josh, Daniel Schneider, Jon A. Krosnick, Alexander Tahk, Eyal Ophir, and Claire Milligan (2014). Prevalence and Moderators of the Candidate Name-Order Effect: Evidence from Statewide General Elections in California. *Public Opinion Quarterly*, 78(2), 416–439.
- Payne, John W., Adriana Samper, James R. Bettman, and Mary Frances Luce (2008). Boundary Conditions on Unconscious Thought in Complex Decision Making. *Psychological Science*, 19(1), 1118–1123.
- Plessner, Henning and Sabine Czenna (2008). The Benefits of Intuition. In Henning Plessner, Cornelia Betsch, and Tilmann Betsch (Eds.), *Intuition in Judgment and Decision Making* (251–265). Lawrence Erlbaum Associates.
- Richardson, Charles (2010, August 2). What's the Donkey Worth? Plenty, the Ballot Draw Reveals. *Crikey*. Retrieved from <https://www.crikey.com.au/2010/08/02/whats-the-donkey-worth-plenty-the-ballot-draw-reveals/>
- Saghai Yashar (2013). Salvaging the Concept of Nudge. *Journal of Medical Ethics*, 39(8), 487–493.
- Shafir, Eldar and Robyn LeBoeuf (2002). Rationality. *Annual Review of Psychology*, 53(1), 491–517.
- Smith, N. Craig, Daniel G. Goldstein, and Eric J. Johnson (2013). Choice without Awareness: Ethical and Policy Implications of Defaults. *Journal of Public Policy & Marketing*, 32(2), 159–172.
- Stanovich, Keith E. (2004). *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. University of Chicago Press.
- Sturgeon, Scott (1994). Good Reasoning and Cognitive Architecture. *Mind & Language*, 9(1), 88–101.
- Sunstein, Cass R. and Richard Zeckhauser (2011). Overreaction to Fearsome Risks. *Environmental & Resource Economics*, 48(3), 435–449
- Thaler, Richard H. and Shlomo Benartzi (2004). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, 112(S1), 164–187.
- Thaler, Richard H. and Cass Sunstein (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. Yale University Press.
- Valone, Thomas J. (2006). Are Animals Capable of Bayesian Updating? *Oikos*, 112(2), 252–259.
- Waroquier, Laurent, David Marchiori, Olivier Klein, and Axel Cleeremans (2010). Is It Better to Think Unconsciously or To Trust Your First Impressions? A Reassessment of Unconscious Thought Theory. *Social Psychological and Personality Science*, 1(2), 111–118.
- Way, Jonathan (2017). Reasons as Premises of Good Reasoning. *Pacific Philosophical Quarterly*, 98(2), 251–270.

- Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies*, 61(2), 341–355.
- Wilson, Timothy D. and Jonathan W. Schooler (1991). Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions. *Journal of Personality and Social Psychology*, 60(2), 181–192.