6

Nudging Well

Roughly, a nudge is a way of influencing people to choose that works by changing aspects of the "choice architecture" (Thaler and Sunstein 2008, 6)—the context in which agents choose—to encourage better choices (better usually insofar as these choices promote the welfare of the choosing agent herself; occasionally nudges aim at the promotion of social welfare instead of individual welfare). It's a familiar fact that people often make choices they themselves recognize are in not in their own interests. Nudging can bring them—us—to choose better.

For example, people often have unhealthy diets. Making healthier foods more cognitively accessible (for example, by putting them at eye level) increases their consumption relative to less healthy foods (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. 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). Options can be *framed* in ways that change behavior: agents are risk averse when options are framed positively, but risk seeking when options are framed negatively, even though the options may be identical in arithmetical terms (Tversky & Kahneman 1981).

While nudges may be in the interests of individuals, however, they're extremely controversial. Critics often argue that nudging *manipulates* us (Bovens 2008; Wilkinson 2013; Saghai 2013). Nudging threatens our autonomy because autonomous choice is rational choice, and nudging bypasses our capacities to reason.

In this chapter, I'll argue that nudging doesn't manipulate us. Nor do nudges bypass reasoning. Instead, nudges work by providing genuine evidence to agents, and when they change behavior, the change occurs

in response to this evidence. Insofar as I advocate improving belief formation by thoroughgoing engineering of the epistemic environment—not merely by clamping down on pollutants, but also by altering cues to belief—the success of this argument is essential for my project. At very least, it would be a large cost to its acceptability if I were forced to advocate manipulation on a large scale. But I have an even broader aim in mind. Coming to see how nudges work allows us to better understand the mechanisms that underlie the processes discussed in previous chapters (social referencing and reliance on environmental cues, for example) as themselves reasons-providing. They're ways of changing minds through the provision of higher-order evidence.

Nudging and Autonomy

Autonomy—the capacity of an agent to govern herself—is highly prized by us. Autonomy is nowhere more significant than in the life of the mind. Kant's injunction *Sapere aude!* (have the courage to use your own understanding) is motivated in important part by the connection between using one's own understanding and autonomy. Of course, a major theme of this book is that individual cognition is much less powerful than we think, and that our epistemic capacities are very importantly owed to the distribution of cognitive labor instead. In light of this fact, and the link between the use of one's own understanding and autonomy, it might be thought that I would respond to the worry that nudges threaten our autonomy by arguing that autonomy isn't so valuable after all; that we overvalue it because we overvalue individual cognition.

This isn't in fact a thesis I want to defend. While it may be true that we overvalue autonomy, due to our overvaluation of individual cognition, autonomy seems to me to be genuinely worth defending. There are at least two reasons why autonomy is genuinely valuable. First, agents are often in a better position than others to make decisions concerning the shape of their own lives. It is one thing to think (as I have argued we should) that we ought to defer to epistemic authorities about difficult and complex questions and quite another to think that we can't come to

reliable beliefs about ourselves. Perhaps even in the personal sphere the social sciences have much to teach us, but we retain *some* degree of epistemic privilege with regard to our own values and preferences.

Second, even if we're apt to decide some questions badly, we seem to have a *right* to settle the shape of our own lives and to decide on a conception of the good for ourselves. When the harms we foreseeably risk will fall primarily on ourselves, others have little or no right to override our decisions. We want our lives to be our own, reflecting our own values and priorities, and even our own mistakes. The value of autonomy may often be overstated but I'm not ready to abandon it. Any impingements on our autonomy require justification.

But nudging *does* seem to threaten our autonomy. In fact, both advocates and opponents of nudges accept that it does. Nudging seems *paternalistic* (Thaler and Sunstein, the original advocates of nudging, describe their program as 'libertarian paternalism'): it manipulates us into making decisions in our own best interests, rather than leaving us to make decisions for ourselves on the basis of our own reasons. There's nothing wrong with parents making decisions on behalf of their children, when (and because) the children lack the rational capacity to make these decisions for themselves. But fully rational agents rightly value making decisions for themselves, and nudging bypasses our capacities for rational agency.

To see how nudges are supposed to bypass rational agency, let's look at a possible nudge in action. An unscrupulous (or perhaps a well-meaning) election official might nudge voters in an upcoming election to vote for a particular candidate by listing that candidate first on the ballot. The official would be taking advantage of the *ballot order effect*: the small but sometimes significant advantage that accrues to names listed higher on the ballot paper (Darcy & McAllister 1990; King & Leigh 2009). The ballot order effect serves as John Doris' prime example of how these (supposedly) non-rational influences may threaten autonomous agency (Doris 2015, 2018). As Doris emphasizes, that a candidate is positioned higher on the ballot is not a genuine *reason* to favor that candidate. Ballot order doesn't correlate with candidate quality (in most jurisdictions, name order is determined by drawing lots). It follows that in being influenced by ballot order, a person has their choices

shaped by facts that are *not* good reasons. To the extent that choice is influenced by ballot order, genuine reasoning is bypassed. Doris therefore insists that these kinds of influences don't involve genuine reasoning or genuine reasons. These influences are, he says, "deeply unintelligent."[1]

I treat you as a rational agent and respect your intellectual autonomy when I give you *reasons* why you should vote for one candidate rather than another. If I nudge you into voting for someone by listing them first on the ballot (or, say, making a vote for them the default option) I do neither. I bypass your capacity to deliberate and instead appeal to "deeply unintelligent" mechanisms. It's because nudges appear to influence choice without offering reasons that both opponents and proponents of nudges believe that nudging is paternalistic. Opponents point to this (alleged) fact as a major plank in their case against nudging, while proponents cite the inevitability of nudging in its defense. Thaler and Sunstein accept that nudges take advantage of the fact that we are—in their own words—"somewhat mindless, passive decision makers" (2008: 37), but argue that there's simply no alternative to nudging. The deeply unintelligent mechanisms that respond to nudges are ubiquitous and nudging is inevitable. If we don't nudge deliberately, people will be nudged nevertheless, either by bad actors who seek to manipulate them or by chance. Whatever we do, or fail to do, we'll all be nudged nevertheless; we might as well put nudging to good use.

Whether nudging is really inevitable and whether there's a normative difference between being intentionally and unintentionally nudged (as Alfano (2013) and Kumar (2016) each argue) are interesting questions, but they're not questions we need answer here. Nudges don't simply manipulate us by bypassing our capacities to reason. Instead, they provide us with evidence, which we typically weigh appropriately. Nudges don't tend to provide arguments or evidence that fit our paradigms, but that's because our paradigms are of *first-order* evidence. We neglect *higher-order* evidence, but higher-order evidence is genuine evidence.

---

[1] Doris is here quoting Stanovich (2004). But Stanovich is more careful than Doris (unsurprisingly, given he has long resisted the facile identification of type 1 processes with irrationality (Stanovich 2018)). Stanovich writes that these processes are "*in some sense* deeply unintelligent" (39; emphasis added); Doris drops the qualification.

Nudging Higher-Order Evidence

First-order evidence, our paradigm of evidence, is evidence that bears directly on the truth or falsity of a particular proposition. The pattern of blood spatter in the room is evidence that the killer used a knife; the fingerprints on the light switch are evidence that the killer was the butler.[2] Higher-order evidence is evidence about our evidence. In epistemology, the main focus of debates about higher-order evidence has been the reliability of the agents who assess the evidence; in particular, on how disagreement can provide evidence about such reliability. Consider this (by now hackneyed) case, *Restaurant Check*:

> Suppose that five of us go out to dinner. It's time to pay the check, so the question we're interested in is how much we each owe. We can all see the bill total clearly, we all agree to give a 20 percent tip, and we further agree to split the whole cost evenly, not worrying over who asked for imported water, or skipped desert, or drank more of the wine. I do the math in my head and become highly confident that our shares are $43 each. Meanwhile, my friend does the math in her head and becomes highly confident that our shares are $45 each
>
> (Christensen, 2007: 193).

Most philosophers agree that under certain conditions, a disagreement of this kind provides higher-order evidence, and that evidence puts rational pressure on the parties to reduce their confidence in their calculation. If the agents who disagree are *epistemic peers*, then they should each conciliate (i.e., lower their confidence in their judgment). Peer dissent is evidence for each person that at least one of them has made a mistake; given that neither has a reason to think that the other is more likely to be in error than themselves, they should treat the disagreement as evidence against their conclusion. It's not first-order evidence (not evidence that they failed to carry the 2 or that they left off one item). It's

---

[2] Most epistemologists prefer to talk about our mental states, rather than the objects of these states, as our evidence. As far as I can see, nothing of significance turns on which approach we take in this context.

higher-order evidence: evidence that they may not have processed their first-order evidence well.

Given that disagreement is ubiquitous, if we must conciliate whenever we encounter it, we seem to lose the right to confidence very broadly. Conciliationism thus gives rise to what has been called the problem of *spinelessness* (e.g., Elga, 2007; Fritz, 2018). Many epistemologists respond to this worry by defending an extremely demanding account of epistemic peerhood, according to which my peers have exactly the same evidence and the same capacities as I do.[3] Since each of us has few peers (so defined), we are each under less pressure to conciliate. As Jennifer Lackey (2010) points out, this maneuver risks cutting the debate over the epistemic significance of disagreement off from the real world cases of dissent that motivated interest in it in the first place. In fact, higher-order evidence, of various strengths and kinds is ubiquitous and we are pervasively and appropriately responsive to it.

In real-life analogues of cases like *Restaurant check*, we rightly treat disagreement as high-order evidence without needing to know the track record of a dissenter. The fact that a sober, apparently well-functioning adult disagrees with me about a sum that's difficult enough for a mistake to be unsurprising is *some* evidence that I have made a mistake. Even if I know that the agent is *not* my peer, in the exacting sense common in the literature, his dissent is still higher-order evidence against my conclusion: I can't entirely dismiss his dissent on the grounds that I'm (say) 5 percent more likely to be right than he is. Of course I can dismiss his dissent if the sum is trivially easy (Lackey discusses a case in which a dissenter disagrees on the sum of 2+2) or if he lacks the competence to perform a calculation like this one. But dissent quite routinely provides some degree of higher-order evidence.

---

[3] Setting the bar for peerhood extremely high is not the only response to the problem of spinelessness. One influential response turns on the attractiveness of a test for peerhood along the following lines: my peers are those agents who are as likely as I am to be right about the issue under dispute. Combined with the so-called *independence principle*, according to which an agent's reasons for discounting a dissenting peer's opinions must be independent of the dispute itself, the test allows may allow us to hold fast to our controversial opinions. *Setting aside the dispute and all the reasons implicated in it*, we have no basis for thinking that dissenters would be as likely as we are to come to the right response (Elga 2007; Fritz 2018; McGrath 2008).

For that matter, *agreement* also provides higher-order evidence. Given that a calculation is moderately difficult for me, if I come to the same answer as an independent agent I should raise my confidence in it. The greater the likelihood that I might have been mistaken, the stronger the evidence provided by independent agreement. The *number* of others who have independently tackled the problem should also make a difference to my confidence. If I'm the lone dissenter at a table of 8, my confidence in my answer should be low; conversely, if many others agree with me (and few disagree), my confidence should rise. Numbers make a difference for two reasons. First, the likelihood that I've made a mistake rises or falls as a function of the number of others who agree with me: the higher the proportion of agents who agree with me, the lower the likelihood that I've made a mistake. Second, sheer numbers make a difference to the plausibility of an appeal to what Lackey (2010) calls "personal information," such as my knowledge that I'm paying attention and I'm being sincere, to break the symmetry between me and dissenters. Again, the higher the proportion of dissenters, the more implausible an appeal to such information to dismiss them, at least when the dissenters are to some significant degree independent of one another.

Of course, all this is true only if other things are equal. Experts ought to give little or no weight to dissent when it comes from those who lack expertise. An expert on climate change shouldn't lower her confidence in her predictions and her models because Donald Trump declares that global warming is a hoax. Nor should she be impressed by the enormous number of dissenters, given that she's an expert and almost none of them have any of the specialist skills to understand her work.

With these facts in mind, we can begin to glimpse the ubiquity of higher-order evidence. We can also begin to see how often I've appealed to it throughout this book. In the previous chapters, for instance, I discussed the role that markers of expertise (possession of relevant qualifications, of a track record of publication, of prizes and citations, and so on) should and do play in guiding our response to testimony. In appealing to such markers, I appealed to higher-order evidence. In giving more weight to some opinions on the basis that they come from someone who possesses these markers, we are taking higher-order evidence in favor of

their views into account. In appealing to an expert consensus, we're also appealing to higher-order evidence.

Markers of expertise are just one of the more obvious kinds of higher-order evidence I've appealed to in this book. Once we've seen how nudges, too, provide higher-order evidence, we'll be in a better position to see just pervasive such appeals have been—that is, we'll be in a position to begin to glimpse just how much we lean on higher-order evidence in ordinary and expert cognition, and how important such evidence is in our epistemic lives.

Nudges as Evidence

How do nudges work? Exactly how they cause behavior, on the standard understanding of their influence, remains elusive: theoretical models often invoke vague notions like "salience," which seem more like place-holders for mechanisms than explanations. There is, however, more or less universal agreement that however they work, they bypass rational cognition. While I don't claim to be able to do very much better at pro-viding a proper account of how they function, I suggest that nudges *do not* bypass rational cognition. Instead, at least typically their influence is due to the manner in which they provide implicit recommendations, and therefore higher-order evidence in favor of the option nudged.

Let's begin with the ballot order effect, which was Doris' prime exam-ple of how these kinds of influences may threaten autonomy. As we saw, the ballot order effect is supposed to be irrational, because candidate order doesn't correlate with the quality of candidates. Of course that's true: since candidate order is settled by lot or in some other way that doesn't track quality (e.g., by the order in which they registered for the election), order doesn't provide *reliable* evidence in favor of any candi-dates. But it doesn't follow that it doesn't provide evidence at all. The order in which items are listed implicates their importance (Green 1998). Think of how news is presented online or on TV. The most important information is presented first ("our headlines at the top of the hour…"); being listed first is therefore implicit testimony that an item is impor-tant. Similarly, though we've taken steps to randomize ballot order, we

may nevertheless be communicating implicit testimony to individuals that some candidates are better than others simply by the order in which we list them.

While I know of no direct evidence that this is how candidate order is understood by voters, there is indirect evidence. This evidence is best approached by considering other nudges. We've already mentioned the nudges that tend to feature as examples in the literature: the way in which changing the visual accessibility of food in the cafeteria line changes consumption patterns (Bucher et al. 2016) and changing the default option to a higher rate on an employment contract increases savings (Smith et al. 2013). Both of these nudges can be understood as providing implicit recommendations to agents.

The use of defaults to change behaviors has widely been seen as taking advantage of our cognitive laziness. But there's evidence that defaults are understood as communications (Fisher 2020a, 2020b; Levy 2019b). Agents tend to see default options 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. 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. That's how the framing of options works: 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's likely that selection of defaults functions the same way.

The selection of a default is the provision of implicit testimony: *this* option is best, or at least sufficiently good to be choiceworthy (isn't this exactly what we'd expect pretheoretically: wouldn't you be extremely surprised if a default option on a form or an employment contract wasn't a reasonable option? Imagine if your employment contract had as the default option saving 98 percent of your $60,000 salary into a retirement fund. Wouldn't you think whoever drew up the contract was incompetent?) Ballot order effects and the effects of making items more or less accessible or salient (by, for example, placing them at eye level) may be

understood in precisely the same way. While it's true that candidate quality doesn't, 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.

This account of how (canonical) nudges function to guide behavior doesn't merely provide an explanation of how they work (albeit an explanation falling well short of a full mechanistic account); it also *rationalizes* them. That is, it shows that and how it's rational to be guided by a nudge. It's rational to be guided by a nudge because it's rational to give due weight to a recommendation, implicit or not. A recommendation is higher-order evidence that an option is choiceworthy, and higher-order evidence is genuine evidence. Let's see how this works.

Framing effects are often regarded as paradigmatically irrational, on the grounds that how identical options are framed has a significant effect on whether they're preferred (e.g., Shafir and LeBoeuf 2002). Since it's the frame, and not the content of the option, that has altered, a change in preference is thought to be irrational: if your preference for A over B flips in response to a change that is irrelevant to which is better, than your preference flip is irrational. But how options are framed *isn't* irrelevant to which is better—far from it. While the first-order evidence is fixed across frames, the higher-order evidence is not: framing just is changing the higher-order evidence. *Of course* it's rational to be guided by a recommendation!

After all, no one ever objected to guidance by testimony on grounds like these. No one ever said "it's irrational to go to restaurant A rather than B just because your friend recommended it. After all, had she recommended B instead, that's where you'd have gone. Her recommendation doesn't change the options themselves." That's just how recommendations are *supposed* to work. They're supposed to provide higher-order evidence. They're guides to what the first-order facts are and therefore it's no objection to them that they leave these facts unaltered.

To be sure, it would be worrying if nudges were *compelling* causes of behavior: if they overrode other and better sources of information, for example. Recommendations don't overwhelm our better judgment and compel us to act. We respond judiciously to them: we integrate them

with other information available to us, including information about the person providing the recommendation. We probably won't go to Tofu Hut on our friend's recommendation if we know it's just failed a health inspection or we know she's very unreliable on questions like these. We take recommendations into account alongside other information and they are decisive for us only when we lack better information. The evidence strongly suggests that nudges work in the same way: they provide information we take into account, not mechanical shoves. The ballot order effect, for instance, may provide implicit testimony to everyone, but it makes a difference to the choices of only two groups of agents: those who have no real preference between candidates and those who have little information about them (Pasek et al. 2014). Ballot order effects influence behavior in just the way, and to the same degree, as testimony from a source regarded as somewhat reliable does, making a difference for those who have little else to go on. That's because ballot order effects *are* implicit testimony.

Nudges are sometimes defended on the grounds that it's *ecologically rational* to be guided by them (Gigerenzer 2015).[4] Use of a heuristic or bias is ecologically rational just in case it enables us reliably to get the right answer. *How* we get to that answer is irrelevant to its ecological rationality. In principle, ecological rationality can vindicate anything, if it turns out to work. If a benevolent God ensured that throwing dice was a reliable way to make investment decisions, it would be ecologically rational to rely on the dice. My defense of nudging doesn't depend on the notion of ecological rationality. Framing options, the selection of defaults and other ways of making options salient are ways of providing implicit testimony, and it's *directly* rational to be guided by testimony. Implicit testimony isn't merely correlated with the right answer, in the way in which (say) being banded might correlate with being venomous in snakes. Rather, implicit testimony is *evidence* for the right answer, and in being guided by it appropriately—in giving it the weight in our

---

[4] Schmidt (2019) defends the rationality of nudges on grounds like these. On his view, nudges are content rational—they enable us to achieve epistemic ends—but they sacrifice a great deal of process rationality by ignoring or failing to track the features that make an option choiceworthy. In contrast, I claim that nudges are process rational in an entirely orthodox manner. Nudges provide evidence for all agents. Their influence is directly rational, not (merely) ecologically rational.

cognition that reflects its actual evidential value as testimony—we're being guided by the evidence. Perhaps there are nudges that are merely ecologically rational, but most are (also) directly rational and do not bypass rational cognition at all.[5]

Those few people sympathetic to the thought that nudging might be rational see nudges as working through rational and non-rational channels simultaneously. One piece of evidence that's been cited for the conclusion that nudge-style influences work non-rationally is that their power increases when we're under cognitive load (when processing resources are scarce, for instance because the person is required to multi-task, or is fatigued or stressed; see, e.g Gilbert and Osborne, 1989; Krull and Erickson, 1995). This fact has been taken to indicate that they work (in part at least) by taking advantage of cognitive laziness or the fact that it is temporarily too difficult for us to make a decision. Ansher et al. (2014) make the point explicit: changing defaults at once provides recommendations to agents *and* takes advantage of non-rational dispositions.

The thought seems to be that if defaults brought about their effects through rational processes alone, we wouldn't see their influence increase under load. If they provided only rational inputs, then the person should respond to them when her processing capacity is undiminished just as strongly as when it is depleted. But that's a mistake. 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're more expert in the relevant domain than I am. We might think of load analogously: while I'm under load, I should give greater weight to testimony because I'm temporarily less expert.

Ashner et al. cite another piece of evidence in arguing that default effects are partially irrational. The evidence comes from a study in which pulmonologists were asked whether they would prescribe a CT scan for

---

[5] It's difficult to come up with an example of a nudge that might be ecologically rational but is not directly rational. Perhaps priming behavior might be an example; whether primes should be understood as presenting us with evidence is difficult to assess. It's important to note that the priming literature has been a principal victim of the replication crisis in psychology: priming is (to my mind) real, but is a fragile and very weak influence on behavior.

a patient. In the control condition, 54 percent 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 percent of physicians cancelled the scan. Following the original authors of this study, Ansher et al. suggest that a mere (non-rational) bias drives the difference between conditions: "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 et al., 2005: 1499). Of course, that's false. Clinical information provides first-order evidence, and that evidence is obviously of critical importance to clinicians. But the attitudes of our epistemic peers—here represented by the decisions of other physicians—provides us with higher-order evidence, and when a question is difficult to settle (as the fact that pulmonologists split on whether to order the scan indicates), such evidence *should* be given significant weight. Again, the best explanation of how this information guides behavior is via the provision of higher-order evidence, and higher-order is genuine evidence.

Disagreement provides evidence about how well we have responded to our first-order evidence; it provides evidence we may have made a mistake in responding to it. Nudges and the like may not provide evidence like that (though they may). A recommendation, implicit or not, provides evidence about some other agent's attitude to an option. It may entirely replace first-order evidence. A recommendation may lead me to choose an option about which I know nothing; that is, about which I lack *any* first-order evidence. It provides evidence not about the facts that make an option choiceworthy, but about its choiceworthiness itself. It's not evidence about *my* evidence, but it's evidence about what the evidence, properly understood, supports.

In Praise of Nudges

Most (if not all) nudges provide agents with higher-order evidence. They are understood, implicitly, as encoding testimony. They're understood

in this kind of way because they *do* provide testimony: to nudge some-one in the direction of an option is to recommend it to her. Just as agents frame communications to convey their opinions, so we make options salient or highlight them in other ways in order to recommend them. Giving greater weight to an option than we otherwise would just because it has been made salient to us in this kind of way is rational, because it's rational to give weight to recommendations (unless we have counter-vailing evidence, or reasons to distrust the person providing them). This remains true even holding first-order evidence fixed. That's just how testimony works and is supposed to work.

Of course, nudges (or their naturally or stochastically occurring ana-logues) may not present us with *reliable* testimony. Candidate order doesn't in fact correlate with candidate quality. The options may have been framed by agents who know little about the costs and benefits, or who seek to manipulate us to their own ends. But these facts are no objection to the claim that nudges work by offering us implicit testi-mony, nor to the claim that giving them due weight in our choices is rational. Exactly the same points, after all, apply to *explicit* testimony. Agents may offer testimony inadvertently (perhaps unaware that we're listening). They can offer testimony despite knowing little about the options, and I don't need to point out that agents may offer testimony in order to manipulate us to their own ends: that's obviously a central tactic of salespeople and advertisers.

Once we see that (most) nudges work by offering implicit testimony to agents, we're in a good position to see that many of their opponents have got things completely backwards. They demand, in effect, that we leave things as they are so that people are offered misleading testimony,[6] rather than change the context of choice so that people are offered testi-mony that genuinely tracks option quality. Though they don't recognize it, they're advocating the deception of others, rather than taking steps to ensure that they're told the truth. That's not respectful of agency: quite the opposite. Nudging well is offering honest testimony, and refusing to nudge is refusing to ensure that bad testimony is no longer offered.

---

[6] In the event the testimony isn't misleading, of course, there's no dispute: no one advocates intervening to change it.

There are nudges that don't appear to work through this kind of mechanism, but most nudges that have been proposed for the improvement of choice provide reliable higher-order evidence and thereby testimony. These nudges are defensible both on the grounds that they provide good evidence to agents, *and* on the grounds that they enhance individual or collective welfare. Some nudges may work through the provision of *unreliable* (first-order) evidence. For instance, fake potholes or fake speed bumps (in both cases painted on the road surface) have been used to slow traffic (Hamill 2008). I take no stand on whether such nudges are defensible; I note only that if they're indefensible it won't be on the ground that they bypass rational cognition. Deliberately misleading evidence is still evidence.[7]

Canonical nudges, however—and, as we'll soon see, other ways of engineering the epistemic environment—don't suffer from this problem. They're properly respectful of agency, because they work by providing the agent with reliable (usually higher-order) evidence and *thereby* improve our welfare. We may utilize such nudges in good conscience.

Stepping Back

In previous chapters, I suggested we needed to go beyond removing pollutants from the epistemic environment, and actually structure the environment to nudge agents toward better beliefs. I've suggested, for example that we should ensure that markers of expertise correlate with genuine expertise; that a view is represented in the media in rough correlation with the proportion of experts who hold it, and so on. Ensuring that higher-order evidence is reliable suggests policies that may be contentious: for instance, it suggests that news organizations shouldn't seek to balance competing experts when one view is very much better supported than another, because balancing speakers falsely conveys the impression that there's no consensus.

---

[7] Nudges like this are paradigm cases of paternalistic interventions: they're noble lies, intended to bring people to behave better by first getting them to believe falsehoods. They're defensible if (or when) such lies are defensible.

Elsewhere, I've suggested that a policy of ensuring that higher-order evidence is conveyed appropriately may support no-platforming certain speakers, on the grounds that provision of a platform itself provides higher-order evidence in favor of a view (Levy 2019c). While there are surely grounds for worry about some of these proposals (we may legitimately worry who can be appropriately trusted to implement nudges, about the potential for their misuse, and about restrictions on free speech), recognizing that these proposals aim at the provision of higher-order evidence disarms some objections. They shouldn't be seen as manipulative or disrespectful of agency: on the contrary, they're maximally respectful, in just the same way and for just the same reasons as giving people arguments and (first-order) evidence is maximally respectful of agency.

Throughout this book, I've argued that we're deeply social agents, agents who owe our epistemic success to the division of epistemic labor and the ways in which we scaffold cognition. Much of the scaffolding we rely on involves the flexible use of higher-order evidence, I now emphasize. Our use of social referencing—our use of cues as to what others believe to form our own beliefs—the conformity bias and the prestige bias, our outsourcing of belief to the environment and our reliance on distributed networks of agents and artifacts; all of these should be seen as reliance on higher-order evidence. Evidence about what the majority believes is higher-order evidence, as we saw in our discussion of how the numbers count when it comes to peer agreement and disagreement. The prestige bias consists in the use of indirect evidence—higher-order evidence—that certain ways of behaving bring success. That someone is prestigious is higher-order evidence that they behave, or think, well. We make certain facts salient to one another—sometimes through the design of the physical environment—to recommend them (and sometimes to provide implicit warnings, which is another form higher-order evidence can take). Peer review is, in part, the institutionalization of higher-order evidence: that a paper has received its imprimatur is (some) evidence in favor of its quality and its reliability.

If we engage in the kinds of strategies I recommend—nudging better belief—we won't be doing anything new. We've been nudging better belief forever: our epistemic success has *always* been dependent on

ensuring that higher-order evidence is reliable. Higher-order evidence is the real secret of our success. Correlatively, epistemic engineering is not dependent on our biases (understood as ways in which we fall short of rationality), or our (putative) cognitive laziness or even our bounded rationality. Rather, it takes advantages of our rational faculties.