Nudges and other moral technologies in the context of power : Assigning and accepting responsibility

Alfano, Mark and Robichaud, Philip

Nudges and other moral technologies in the context of power: Assigning and accepting responsibility

Mark Alfano, Delft University of Technology & Australian Catholic University
Philip Robichaud, Vrije Universiteit Amsterdam

Abstract

Strawson argues that we should understand moral responsibility in terms of our practices of holding responsible and taking responsibility. The former covers what is commonly referred to as *backward-looking responsibility*, while the latter covers what is commonly referred to as *forward-looking responsibility*. We consider new technologies and interventions that facilitate assignment of responsibility. Assigning responsibility is best understood as the second- or third-personal analogue of taking responsibility. It establishes forward-looking responsibility. But unlike taking responsibility, it establishes forward-looking responsibility in someone else. When such assignments are accepted, they function in such a way that those to whom responsibility has been assigned face the same obligations and are susceptible to the same reactive attitudes as someone who takes responsibility. One family of interventions interests us in particular: nudges. We contend that many instances of nudging tacitly assign responsibility to nudgees for actions, values, and relationships that they might not otherwise have taken responsibility for. To the extent that nudgees tacitly accept such assignments, they become responsible for upholding norms that would otherwise have fallen under the purview of other actors. While this may be empowering in some cases, it can also function in such a way that it burdens people with more responsibility that they can (reasonably be expected to) manage.

Word count: 5294 (excluding references)

Keywords: responsibility, reactive attitudes, nudge, power relations

1. Introduction

In his seminal essay, "Freedom and Resentment," Peter Strawson (1962) argues that we should understand moral responsibility in terms of our practices of holding responsible. These practices are canalized by reactive attitudes such as anger and resentment, guilt, and vicarious indignation. Strawson was primarily concerned with what is nowadays referred to as *backward-looking responsibility*, in which bad or problematic outcomes are traced back to an agent who is then held responsible and prima facie blameworthy (Watson 1987; Fischer & Ravizza 1998; van de Poel 2011). More recently, philosophers have added to the mix the notion of *forward-looking responsibility*, in which an agent undertakes to ensure that certain outcomes (do not) obtain, certain values are upheld, or certain relationships are maintained (Goodin 1998; Darby and Branscombe 2014).

In this chapter, we consider new technologies and interventions that facilitate not just holding responsible and taking responsibility, but assignment of responsibility (and assignment of assignment of responsibility, and so on). Assigning responsibility is most easily understood as the second- or third-personal analogue of taking responsibility. It establishes forward-looking responsibility. But unlike taking responsibility, it establishes forward-looking responsibility in someone else. And when such assignments are either tacitly or explicitly accepted, they function in such a way that those to whom responsibility has been assigned face the same obligations and are susceptible to the same reactive attitudes as someone who, of their own free will, takes responsibility in the more familiar sense.

One family of interventions interests us in particular: nudges and related moral technologies. As defined by Thaler & Sunstein, a nudge is, "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentive." (2008, p. 6). Nudges are a subset of moral technologies, which attempt "to bridge the gap between moral psychology and normative theory by recommending ways in which we, as moral psychology describes us, can become more as we should be, as normative theory prescribes" (Alfano 2013, p. 6). We contend that many instances of nudging tacitly assign responsibility to nudgees for actions, values, and relationships for which they might not otherwise have taken responsibility. In so doing, such nudges do not bypass the agent's reasoning or values; instead, they engage the agent's reasoning and values by prompting them (if only unconsciously) to accept responsibility. To the extent that nudgees tacitly or explicitly accept such assignments, they thereby become responsible for upholding norms that would otherwise have fallen under the purview of other actors, such as the state or those with more political, economic, or epistemic power. While this may be empowering in some

cases, it can also end up burdening people with more responsibility that they can (reasonably be expected to) manage.

Here is the plan for this paper. We begin by arguing that the Strawsonian framework is not sufficient for understanding the full range of responsibility-practices that need to be explained. Next, we enrich the Strawsonian framework with the concepts of assigning, accepting, and repudiating responsibility. Each of these moral psychological practices can take place either tacitly or explicitly. We then we explain how to incorporate nudging into the enriched framework. We argue that many instances of nudging are best understood as moral technologies aimed at assigning responsibility to agents who will be inclined to tacitly accept the assignment. We conclude by reflecting on the conditions under which assigning responsibility by nudging is morally acceptable.

2. Strawson's inadequacy

The Strawsonian framework presupposes a largely horizontal power structure, in which each agent both is able to hold others to account and feels accountable to others (McKiernan 2016; Todd forthcoming). While such horizontal power structures undeniably characterize much of human activity both now and in our evolutionary niche (Dunbar 1992, 1993), they are clearly not the only ways in which people relate to one another morally and politically.

Agents whom the naive Strawsonian might feel entitled to hold responsible via blame, sanctions, punishments, and reactive attitudes such as anger and resentment are often unaccountable not just legally but also -- to their own minds and from the point of view of their social circles -- morally. Someone who is excluded from a community via contempt or disgust may be taken to have no standing to hold those within the community responsible (Bell 2013, p. 6; Mason 2018; Darwall 2018). Kate Manne (2017, p. 186) discusses an example from *The Talented Mr. Ripley* in which a woman's concerns are dismissed as irrelevant because, in her misogynistic culture, she lacks standing to voice them. Further examples from real life are not hard to come by, as the #metoo movement demonstrated in recent months. Conversely, agents whom the naive Strawsonian might view as beyond reproach are often blamed, punished, resented, and scapegoated for behavior that seems morally unobjectionable or even praiseworthy (e.g., whistleblowers like Chelsea Manning and Edward Snowden).

Similar counterexamples crop up in the context of taking responsibility. Agents who might feel entitled to take responsibility for a particular outcome or value may sometimes be denied the authority to do so. For example, when Dr. Tamika Cross, who is black, raised her hand to volunteer to treat an ailing passenger on a flight from Detroit to Minneapolis on 9 October 2016, the flight attendant who had requested medical assistance demanded to see her credentials and

said, “Oh no, sweetie put your hand down; we are looking for actual physicians or nurses or some type of medical personnel. We don’t have time to talk to you” (Hauser 2016). Conversely, agents may be allowed to take responsibility for outcomes or values that they manifestly lack the competence to handle. For instance, in 2017 Jared Kushner sought and was accorded responsibility for managing the American opioid epidemic, resolving the Israel-Palestine conflict, and overhauling the entire US federal bureaucracy (Bartlett 2017).

The Strawsonian might object to these alleged counterexamples by insisting that they represent abuses of our practices of taking responsibility and holding responsible. Presumably there is some slack built into these practices, allowing both for cases in which agents aren’t held responsible when they could or should be, and for cases in which agents are held responsible when they shouldn’t be. This is easier said than done, however. Strawson argues that our practices of holding responsible and taking responsibility, along with “their reception, the reactions to them, really are expressions of our moral attitudes [....] Our practices do not merely exploit our natures, they express them” (1963 [2003], p. 93). If this is right, then whatever practices humans systematically employ in the context of responsibility, punishment, and blame must be incorporated into the framework.

Strawson and his fellow travelers attempt to systematize exceptions by articulating theories of excuses, exempting conditions, and the like (Austin 1979). But the sorts of counterexamples mentioned above do not fit into this framework and are, we contend, systematic enough to force us to rethink the Strawsonian framework. Scapegoating, victim-blaming, and retaliating against whistleblowers are as much a part of human practices as the horizontal reactive attitudes and activities that Strawson recounts. Manuel Vargas argues that they might be elements of the optimific (from his particular consequentialist standpoint) moral responsibility system. (2013, pp. 177-180). We are not just talking about corrigible problem-cases that present anomalous counterexamples to the Strawsonian framework. We are talking about cases where the framework systematically and constitutively delivers problematic results. If people are enculturated to respond to praise, blame, admiration, resentment, and the whole suite of reactive attitudes as Strawson suggests, then these outcomes are -- if not inevitable, then highly likely.

These observations suggest that, while the naive Strawsonian framework may be adequate to a society inhabited only by saints and guilt-prone sinners, it must be supplemented for a world like ours in which the shameless ride on the buoyant unaccountability made possible by toadies, sycophants, scapegoaters, and the rest of their ugly menagerie. This chapter is an attempt to move beyond Strawson’s cheerful perspective by incorporating relations of power into the theory of responsibility.

3. Enriching the Strawsonian framework

Define a responsibility-community as a set A = {*a*1, *a*2, *a*3, …, *a*n} of agents (potentially including group agents), a set V = {*v*1, *v*2, v3, …, *v*n} of values (broadly construed to include desirable outcomes, relationships, and so on), and a relation R mapping from agents to values. Let *a*R*v* mean that agent *a* takes responsibility for value *v*. Such a community is dynamic insofar as additional agents can be added to A, existing agents can be eliminated from A, additional values can be added to V, existing values can be eliminated from V, agents can assume new responsibilities, and agents can be stripped of or otherwise lose responsibilities. The familiar Strawsonian framework incorporates three main functions related to such dynamics. First, an agent can *take responsibility* for preserving, promoting, pursuing, or protecting a value. While Strawson himself does not mention this, it's clear that in many cases the community possesses veto power over such taking responsibility. If your community rejects your bid to take responsibility for a value, whether because you lack standing to assume it in the first place or because you are deemed untrustworthy, you cannot take responsibility for it in a way that licenses others to direct reactive attitudes and the sanctions they express towards you. Second, an agent can be held responsible (via reactive attitudes, sanctions, and so on) for failing to preserve, promote, pursue, or protect some value for which they have or should have taken responsibility. Third, an agent can be (temporarily) excluded from the community by taking an "objective stance" rather than a "participant stance" towards them.

These functions, when enacted explicitly, are part of what we might call the language game of responsibility (Wittgenstein 1953; Sellars 1954), in which the speech-act of declaring oneself responsible plays a pivotal role (Austin 1975; Searle 1995). However, they can also occur without any words being uttered. Gestures, eye-contact, shared assumptions, default rules, and a wide variety of other non-linguistic activities contribute to our responsibility practices. In this section, we supplement the Strawsonian picture by theorizing further functions available to members of responsibility-communities. These include assigning responsibility, accepting responsibility, and repudiating responsibility.

3.1. Assigning responsibility

Subject to the veto of the responsibility-community, when *a* bids to take responsibility for *v*, she succeeds in doing so. In second-person and third-person assignments of responsibility, *a*2 assigns *a*1 responsibility for *v*. As before, such assignments are typically subject to veto by the community. The assignment might be rejected because *a*2 lacks the authority to make the assignment (either *tout court* or specifically to *a*1), because some other agent is already responsible for *v* (and there is a limit on the number of agents who can be responsible for it),

because the community deems *a*1 untrustworthy, because *a*1 rejects the assignment, or for some other reason.

These considerations point to the role of power in responsibility-communities. In extremely egalitarian communities, assignments are always only invitations, which the assignee can reject. At the opposite extreme, dictatorial communities vest the power to assign responsibility in one or a few individuals, whose declarations must be accepted -- both by the assignee and by the community at large. In between are a wide variety of communities in which some assignments are mandatory but others can be rejected (either by the assignee or by other members of the community). As before, assigning responsibility can be done explicitly through a declarative speech-act, but it can also be enacted through non-linguistic actions such as gestures, eye-contact, shared assumptions, default rules, and so on. There are also various means by which responsibility assignments can be vetoed by the community. The veto can be an explicit speech-act, but it may also be enacted through non-linguistic actions such as gestures (shaking one's head, rolling one's eyes), facial expressions (glowering), incorporating prohibitions into bureaucratic forms or online interfaces, and so on.

3.2. Accepting and repudiating responsibility

Because extant discussions of forward-looking responsibility are built around first-personal acts of taking responsibility, the distinction between assigning responsibility and accepting the assignment of responsibility may at first seem moot. Why would anyone take responsibility for a value in one breath just to reject that same responsibility in the next? While it is true that a community could veto the agent's bid to take responsibility, the prospect does not loom large unless the assigner and the assignee are distinct. When we enlarge the universe of agents and allow for second-person and third-person assignments of responsibility, it becomes obvious that the assignee or other members of the community might resist a particular assignment. These observations make it clear that, at least in non-dictatorial communities, when *a*1 makes a bid to assign *a*2 responsibility for *v*, both *a*2 and other members of the community typically have the opportunity to say no. How easy and effective such naysaying is depends on whether the assignment was an invitation, a request, a plea, an offer of bribe, a threat of blackmail, an order, or some other speech-act.

For our purposes, it's important to bear in mind that the assignment may be non-verbal and tacitly accepted. In such cases, the assignee and the community may end up inadvertently ratifying the assignment of responsibility without even realizing it. This points to the need for agents and the communities they inhabit to audit the distribution of responsibilities and sometimes to repudiate some assignments. Such repudiation may be called for when someone who initially took responsibility for a value is no longer trusted to protect, promote, pursue, or

preserve it. It may be called for when the assigner no longer needs to trust another agent to uphold the value. It may also be called for when someone realizes that they have tacitly accepted the assignment of responsibility for a value that they should not own. If this is right, then at the moments of taking and assigning, and for an indefinite period thereafter, members of responsibility communities need to monitor and occasionally revise the distribution of responsibilities amongst themselves. They manage this via tacit or explicit acts of accepting and repudiating responsibility -- essential elements of our responsibility-practices that Strawson and his fellow travelers neglect.

3.3 Higher-order responsibilities: powers and immunities

In this section, we use Hohfeld's (1913) structural analysis of first-order and higher-order legal relations as an analogy for first-order and higher-order responsibility relations. Hohfeld's scheme is built on an ontology of agents and actions. There are four first-order legal relations that characterize agents and their actions: *right*, *no-right*, *duty*, and *privilege*. In this system, agent $a1$ has a right to perform action φ just in case all other agents $a2$, $a3$, … , an have a correlative duty to allow (i.e., not to prevent) $a1$ doing φ. By contrast, $a1$ has no-right to φ just in case at least one other agent ai does not have a duty to allow $a1$ doing φ; in other words, $a1$ has no-right to φ just in case at least one other agent ai has the privilege to prevent $a1$ from φing.

Higher-order legal relations concern not base-level actions but actions that do or would alter existing legal relations. As with first-order legal relations, there are four higher-order legal relations: *power, disability, liability, and immunity*. In Hohfeld's scheme, agent $a1$ has the power to change R (where R is a first-order or higher-order relation) just in case there is some agent ai who has a liability to changes in R. For instance, while you may currently have property rights over some object (i.e., all others have a duty not to use the object without your consent), a judge may have the power to strip you of that right (e.g., when applying a legal penalty). And while you may currently lack property rights over some object, a judge may grant you that right (e.g., when granting compensatory damages). Powers can also be over higher-order legal relations. For instance, in the United States a judge may be stripped of the second-order power to alter property rights through Congress's exercise of the third-order power of conviction for impeachable high crimes and misdemeanors.

Just as powers are interdefinable with correlative liabilities, disabilities are interdefinable with correlative immunities. Agent $a1$ has immunity to changes in relation R just in case all other agents have the disability to change R (equivalently, no agent other than $a1$ has the power to change R). An inalienable right is a right that is also protected by immunity (and immunity of that immunity, and immunity of that immunity of that immunity, and so on). For example, not only does everyone have a right not to be enslaved (everyone has a duty not to enslave them), but

that right is immune to all powers that might alter it (everyone has a disability to change that right). These relations are summarized in Tables 1 and 2.

Table 1: first-order legal relations. This table illustrates the four Hohfeldian first-order legal relations.

a has a right to φ	← contraries →	*a* has no-right to φ
↑ correlates ↓		↑ correlates ↓
$\forall b \in A$ *b* has a duty to allow *a* to φ	← contraries →	$\exists b \in A$ s.t. *b* has a privilege to prevent *a* from φing

Table 2: higher-order legal relations. This table illustrates the four Hohfeldian higher-order legal relations.

a has a power over R	← contraries →	$\forall b \in A$ *b* has a disability to change *a*'s possession of R
↑ correlates ↓		↑ correlates ↓
$\exists b \in A$ s.t. *b*'s possession of R is liable to changes by *a*	← contraries →	*a* has immunity with respect to R

We can understand the possession of forward-looking and backward-looking responsibilities as analogous to first-order legal relations. Agent *a* has a forward-looking responsibility for *v* just in case other members of the community have a moral right to hold *a* responsible (subject to exempting and excusing conditions) for failing to protect, promote, pursue, or preserve *v*. Likewise, we can understand taking responsibility, assigning responsibility, accepting responsibility, and repudiating responsibility as analogous to higher-order legal relations. Taking, assigning, or repudiating responsibility is the enactment or expression of a *power* over forward-looking responsibilities. Thus, in order for *a*1 to take or repudiate responsibility for *v*, *a*1 must have a *liability* with respect to forward-looking responsibility for *v*. Likewise, in order for *a*2 to assign *a*1 responsibility for *v* or veto *a*1's taking responsibility for *v*, *a*1 must have a liability with respect to forward-looking responsibility for *v*. In cases where *a*2 lacks standing to assign forward-looking responsibility for *v* to *a*1, we can say that *a*1 has *immunity* with respect to that responsibility while *a*2 has a disability.

In the next section, we show that responsibility-assigning and -repudiating nudges presuppose higher-order moral powers, liabilities, and immunities. While such powers and immunities are sometimes embodied, they are not always. In such cases, nudging is morally problematic.

4. Nudge

This supplementation of the Strawsonian framework offers a helpful way of understanding how nudges, though a recent development spurred on by advances in behavioral economics, jibe with established responsibility practices. So-called *choice architects* design and structure choice situations in ways that conduce to the performance of certain target actions (or omissions). These nudges work by exploiting widespread decisional heuristics and biases of which most of us are generally unaware (Thaler & Sunstein 2008). Although most of the discussion about nudges focuses on their use by allegedly well-meaning policy makers who are motivated to steer citizen behavior in ways that promote well-being, the use of such interventions is by no means restricted to benevolent government agents. By setting the default for participation in some scheme or plan as opt-out rather than opt-in, policy makers have increased organ-donation status and companies have increased participation in corporate pension schemes (Shepherd et al 2014 ; Beshears et al. 2017). By diminishing the effort an agent has to make in order to secure some benefit or status, universities increased the likelihood that students of lower socioeconomic status would matriculate just by sending them application forms that were filled in with information from the family's previous year's tax returns (Castleman & Page 2016). A third kind of nudge involves the explicit or implicit communication of socially accepted norms. The Guatemalan government decreased rates of tax evasion by sending out letters that indicate that evasion rates are quite low (Kettle et al. 2016). These cases exemplify three nudge types, and it will suffice for our purposes to focus on them.

The key question here is how the practice of nudging can be understood in light of the framework developed above. The relevant agents in this framework are policy makers and business managers, who are potential responsibility assignors, and citizens and employees, who are candidate responsibility assignees. The nudges that are implemented by the former groups target those in the latter groups and, if successful, result in the nudgee having taken responsibility for various distinct values. Regarding these values, there are three candidates that are the most relevant here. The first and perhaps most obvious is *the value of whatever boost in well-being obtains as a result of the nudgee performing the target action*. This may be an increase in her own well being, as in the case of nudges that impact decisions that influence health or financial stability. If a1 successfully nudges a2 to save more for retirement, a2 will effectively be accepting responsibility for the value of being in a position to live comfortably in retirement. A second value is that of the nudged agent *coming to meet her obligations* (or forward-looking responsibilities) as a result of performing the nudged action. Sticking with the

savings nudge, a1's nudge, if successful, induces a2 to perform an action that amounts to the meeting of a standing obligation to her future self and her family, and which is necessary for the performance of downstream actions that are also plausible moral obligations (such as making effective charitable donations or supporting friends in financial need). In this case the nudged agent accepts responsibility for the value of meeting obligations that she has. The third value is more normatively momentous. It involves a1 making it the case that the nudgee is in a position to realize a new value or meet some new obligation (or responsibility). When a nudge is successful at getting prospective students to matriculate to university, these nudged agents will incur all the new forward-looking responsibilities associated with being university students (such as not plagiarizing). Not only will these nudged agents be in a position to realize the particular values associated with being students, but they will also have a new set of moral obligations. In a similar fashion, someone defaulted into a pension plan accepts new responsibilities -- those that come with managing and preserving one's nest egg.

There are thus three candidate values that nudged agents, as a result of being nudged, can accept responsibility for (some bearer of value, meeting one's existing obligations, and meeting new obligations). Against this background we can see how the higher-order responsibility relations discussed in the previous section apply. When choice architects succeed, they are engaged in the practice of assigning responsibility, and so we can question whether they (should) have the power to do so and, correlatively, whether nudgees are relevantly liable to this assignment. Another relevant issue concerns accepting and repudiating responsibility. If the nudged agent accepts responsibility for the relevant values, then we can also say that she's liable to the responsiblity assignment. Similarly, her repudiation of the responsibility assignment, if the nudger had the power to make the assignment in the first place, can only occur after the assignment has been made. Just to take one example, you might have been nudged via an enrollment default to have employer-subsidized health insurance, but later decide to cancel it as a money-saving measure. You thereby repudiate the forward-looking responsibility to maintain access to affordable health services. Finally, consider the issue of immunity and correlative disability. If there's reason to think that agents are immune from being nudged into accepting responsibility for some value, then it follows that all possible nudgers have the disability to make the relevant responsibility assignment. This might occur if an accepted social norm entails that certain choice domains ought to be free of influence by government agents or employers. Presumably, our decisions about whom to vote for ought to be immune from influence by policy makers or political parties in power. Few people would accept a policy of having one's ballot filled in with preferences for candidates that promise to support policies that promote one's self interest. Plausible and widely accepted democratic principles disable governments from nudging votes in this way, even if the effect of voting defaults allow citizens to realize some political outcome that is valuable to them (or to satisfy their civic duty of supporting non-fascistic

candidates, for example). Another area in which we might think such immunity obtains is being nudged into marital relationships with particular partners.

At this point one might wonder whether the mechanism by which nudges assign these responsibilities -- by capitalizing on (or exploiting) our cognitive and behavioral quirks, many of which we are unaware of -- raises worries about manipulation. The same considerations that justify immunity against voting nudges, namely that they manipulate agents to making decisions that they might not otherwise make in the absence of the nudges, might generalize. Why not also conclude that governments and companies that nudge us to making responsible choices are also manipulating us? If manipulation suffices for immunity in the one nudge, perhaps it should do so for all of them. One might think that nudgers are taking what Strawson (1962) called the objective attitude with respect to nudgees. Rather than viewing nudgees as equal participants in the moral community whose autonomy should be respected, the nudger sees them as objects to be controlled and manipulated. The question whether nudges are manipulative in a way that entails universal immunity is too broad for our purposes, but what we do in the final section is appeal to the framework developed in this chapter in order to highlight other features of particular nudges that might make them morally problematic. In so doing we show that our framework offers a novel way of mapping this normative terrain.

5. When (not) to assign responsibility via nudging

It is constructive to think about the ethics of nudges by viewing them through the lens of power relationships and the kind of moral principles that govern them. Of central relevance here are considerations that are typically taken to confer and legitimate the power of governments or private agents to assign responsibility to others. On the other side are considerations that support claims of immunity of individuals against such assignments. The permissibility of some nudge will depend in part on whether the balance of reasons supports the nudger's power to assign responsibility or the claim that some agent has to immunity. A related factor is the nudged agent's (and her community's) acceptance or repudiation of the responsibilities assigned. In this final section we consider two distinct ways of grounding the claim that a nudge is problematic.

5.1 No power to assign responsibility

There are certain domains in which governments, companies, and institutions clearly have the power to assign responsibilities and other domains in which they clearly don't. Policies that require parents to take responsibility for their children's welfare will strike many as legitimate. No individuals are better-placed than parents to discharge this obligation, and it is a weighty obligation indeed. These facts plausibly confer power on the state to assign parents caregiving responsibilities. But are the nudges currently deployed by governments relevantly similar? There are at least two dissimilarities. First, the kinds of values at stake in the assignments of

responsibility involved in nudges are multifarious and don't always involve getting an agent to meet her standing obligations. As discussed above, nudges that default into organ-donation schemes have a primarily axiological aim. Plausibly, the strength of whatever consideration underwrites the power to nudge will vary with the goal of the nudge. When the nudge would merely realize some good, the case for the state having the power to meddle in the lives of its citizens might be weaker than it would be if the nudge would realize the meeting of one's obligations. That is, when it comes to power-conferral, axiological purposes are weaker than deontic. Furthermore, there might be room within these two types of purposes for further relevant distinctions. Regarding value, a nudge might target the general societal good or something good for the nudged agent herself, and regarding obligation, a nudge might enable an agent to meet an obligation that she has to herself or an obligation that she has regarding others. Table 3 represents a plausible rendering of how much these aims would support a nudger's claim to the power to assign responsibility.

Table 3: A rough rank ordering of the degree to which different types of nudges confer power on the nudger to assign responsibility.

Nudge aim	Degree of power-conferring support	Example
Realize value to society	Low	Organ-donation defaults
Realize value to the nudgee herself	Low-moderate	Pension-plan defaults
Meet obligations to nudgee herself	Moderate	Easier-to-complete college applications
Meet obligations to others	High	Information about neighbor's energy consumption

To be sure, this is just one way of characterizing the strength or power-conferring support that these nudges have. The norms accepted by a particular community concerning the importance of these generally-described aims may differ from this one. What this motivates is a demand that particular nudge proponents proffer reasons to accept their claim to have the power to assign responsibilities via their nudges. When the degree of power-conferring support is plausibly low, there's a risk that the balance of reasons will not ground the power to nudge.

5.2 Immunity against responsibility-assignments

Even when there is a high degree of power-conferring support for some nudge, perhaps because it enables a nudgee to meet many obligations, immunity may nevertheless obtain. As mentioned above, there are certain domains in which many communities are likely to accept norms that rule out paternalism of any kind. In addition to voting and marital decisions, one might think that paternalism, even the soft paternalism of nudges, should be forbidden in the context of healthcare decisions (White 2016). A second way of being immune is repudiation by individuals or the community. This repudiation can be actual or hypothetical. For instance, In 2010 the Swiss municipality of St. Galen voted in a referendum to shift its energy production to greener sources; the local power utility then used default nudges to assign residents (in the first instance) to slightly more expensive but greener mixes of electricity sources. While citizens could choose to switch to cheaper, higher-polluting mixes (or an even greener mix), the default nudge had a large effect (Chassot et al. 2014). In voting to take responsibility for shifting their energy supply, the denizens of St. Galen conferred democratic legitimacy on this nudge, relinquishing any claim to immunity. We can easily imagine another municipality holding a similar referendum and choosing not to shift their energy mix. In such a case, it seems clear that nudging them to accept responsibility for paying for greener power would violate their immunity.

5.3 Conclusion

In this section we've shown that the framework of assigning responsibilities offers a novel orientation for thinking about the ethics of nudges and of other moral technologies. By shifting the focus to the power wielded by choice architects and the potential immunity of their targets, we can move beyond concerns about manipulation and autonomy. Nudges are just one way in which we engage in the practice of assigning, accepting, and repudiating responsibilities, and so the same normative frameworks that are appropriate to the latter should be brought to bear on the former.

References

Alfano, M. (2013). *Character as Moral Fiction*. Cambridge University Press.

Austin, J. L. (1975). *How to Do Things with Words*. Harvard University Press.

Austin, J. L. (1979). A plea for excuses. In J. O. Urmson & G. J. Warnock (eds.), *Philosophical Papers*. Oxford University Press.

Bartlett, B. (2017, April 3). Jared Kushner, the assistant with the big portfolio. *The New York Times*. Url = < https://www.nytimes.com/2017/04/03/opinion/jared-kushner-the-assistant-with-the-big-portfolio.html >. Accessed 5 January 2018.

Bell, M. (2013). *Hard Feelings: The Moral Psychology of Contempt*. Oxford University Press.

Beshears, J. Benartzi, S., Mason, R., & Milkman, K. (2017). How do consumers respond when default options push the envelope? Available at SSRN: url = < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3050562 >.

Castleman, B. & Page, L. (2014). Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence. *Journal of Human Resources*, 51(2): 389-415.

Chassot, S., Wüstenhagen, R., Fahr, N., & Graf, P. (2014). Introducing green electricity as the default option. In C. Herbes & C. Friege (eds.), *Marketing Renewable Energy: Concepts, Business Models, and Cases*. Springer.

Darby, D. & Branscombe, N. (2014). Beyond the sins of the fathers: Responsibility for inequality. *Midwest Studies in Philosophy*, 38: 121-37.

Darwall, S. (2018). Contempt as an other-characterizing, "hierarchizing" attitude. In M. Mason (ed.), *The Moral Psychology of Contempt*. Rowman & Littlefield.

Dunbar, R. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6): 469-93.

Dunbar, R. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4): 681-735.

Fischer, J. M. & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.

Goodin, R. (1998). *Social Welfare and Individual Responsibility*. Cambridge University Press.

Hauser, C. (2016, October 14). Black doctor says Delta flight attendant rejected her; sought 'actual physician'. *The New York Times*. Url = < https://www.nytimes.com/2016/10/15/us/black-doctor-says-delta-flight-attendant-brushed-her-aside-in-search-of-an-actual-physician.html >. Accessed 5 January 2018.

Hohfeld, W. (1913), Some fundamental legal conceptions as applied in judicial reasoning. *Yale Law Journal*, 23(1): 16-59.

Kettle, S., Hernandez, M., Ruda, S., & Sanders, M. (2016). Behavioral interventions in tax compliance: Evidence from Guatemala. *World Bank Policy Research Working Papers*. URL = < https://doi.org/10.1596/1813-9450-7690 >.

Manne, K. (2017). *Down Girl: The Logic of Misogyny*. Oxford University Press.

Mason, M. (2018). Contempt: At the limits of reactivity. In M. Mason (ed.), *The Moral Psychology of Contempt*. Rowman & Littlefield.

McKiernan, A. (2016). Standing conditions and blame. *Southwest Philosophy Review*, 32(1): 145-51.

Searle, J. (1995). *The Construction of Social Reality*. Free Press.

Sellars, W. (1954). Some reflections on language games. *Philosophy of Science*, 21(3): 204-28.

Shepherd, L., O'Carroll, R., & Ferguson, E. (2014). An international comparison of deceased and living organ donation/transplant rates in opt-in and opt-out systems: A panel study. *BMC Medicine* 12(131). URL = < https://doi.org/10.1186/s12916-014-0131-4 >.

Strawson, P. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48: 1-25, reprinted in G. Watson (Ed.) (2003). *Free Will, 2nd edition.* Oxford: 72-93

Thaler, R. & Sunstein, R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

Todd, P. (forthcoming). A unified account of the moral standing to blame. *Nous*.

Van de Poel, I. (2011). The relations between forward-looking and backward-looking responsibility. In N. Vincent, I. van de Poel, & J. van den Hoeven (eds.), *Moral Responsibility: Beyond Free Will and Determinism*. Springer.

Vargas, M. (2013)*Building Better Beings: A Theory of Moral Responsibility*. Oxford University Press.

Watson, G. (1987). Responsibility and the limits of evil. In F. Schoeman (ed.), *Responsibility, Character, and the Emotions*. Cambridge University Press.

White, M. D. (2016). Bad Medicine: Does the unique nature of health care decisions justify nudges? In edited by I.G. Cohen, H.F. Lynch, and C. Robertson *Nudging Health: Health Law and Behavioral Economics*. Johns Hopkins University Press

Wittgenstein, L. (1953). *Philosophical Investigations*. G. E. M. Anscombe & R. Rhees (eds.), G. E. M. Anscombe (trans.). Blackwell.