

# Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration

Luciano Floridi<sup>1</sup>

**Abstract** Toleration is one of the fundamental principles that inform the design of a democratic and liberal society. Unfortunately, its adoption seems inconsistent with the adoption of paternalistically benevolent policies, which represent a valuable mechanism to improve individuals' well-being. In this paper, I refer to this tension as the *dilemma of toleration*. The dilemma is not new. It arises when an agent A would like to be tolerant and respectful towards another agent B's choices but, at the same time, A is altruistically concerned that a particular course of action would harm, or at least not improve, B's well-being, so A would also like to be helpful and seeks to ensure that B does not pursue such course of action, for B's sake and even against B's consent. In the article, I clarify the specific nature of the dilemma and show that several forms of paternalism, including those based on *ethics by design* and structural *nudging*, may not be suitable to resolve it. I then argue that one form of paternalism, based on *pro-ethical design*, can be compatible with toleration and hence with the respect for B's choices, by operating only at the informational and not at the structural level of a choice architecture. This provides a successful resolution of the dilemma, showing that *tolerant paternalism* is not an oxymoron but a viable approach to the design of a democratic and liberal society.

**Keywords** Ethics by design · Liberalism · Nudging · Paternalism · Toleration · Value-sensitive design

---

✉ Luciano Floridi  
luciano.floridi@oii.ox.ac.uk

<sup>1</sup> Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK

## Introduction

Toleration has been a philosophical problem at least since the publication of Locke's famous *Letter* (Locke 1689, 1991). I am not referring to the important but *external* struggle with zealots of all kinds, but to the corrosive nature of two *internal* difficulties. One of them was already well known to Locke himself. Contrary to justice and peace—two other fundamental principles that orient the design of a democratic and liberal society—there is such a thing as too much toleration, which becomes socially risky and potentially immoral whenever it is exercised towards the intolerant. This so-called *paradox of toleration* has fuelled the philosophical debate ever since its unveiling. I have dealt with this paradox elsewhere<sup>1</sup> and it is not the topic of this article, although it is strictly related to it.<sup>2</sup> The other internal difficulty is perhaps less famous yet no less challenging or common. Lacking a specific label I shall call it the *dilemma*<sup>3</sup> of toleration.<sup>4</sup> Informally, this arises when Alice wishes to be tolerant and altruistic towards, but also “knows better” (has more or better information) than Bob, someone she cares about, so that she is caught in the following impasse. For example, Alice may decide to have a speed limiter fitted in Bob's car, to determine how fast he can drive (more realistically, Ford's new S-Max has a new feature that automatically slows one down so that one is traveling within the speed limit). On the one hand, Alice would like to be tolerant and respectful of Bob's choices but, on the other hand, she is altruistically concerned that a particular course of action would harm, or at least not improve, Bob's well-being, so she would like to be helpful in ensuring that he does not pursue it, for his sake and even against his consent.

I shall be more precise in the following pages, but the reader may have already guessed that this well-known conflict, between restraint and intervention, arises from the troublesome relation between toleration and paternalism, and the different ways in which they interact with a fundamental respect for individuals' choices. Such a conflict is the topic of the article. Broadly speaking, this may be framed within the debate on the optimal balancing of exercise of beneficence and respect for autonomy. Such a debate dates at least to Mill. The contemporary revival seems to have its roots in the eighties.<sup>5</sup> As (Brock 1988) wrote: “there has been substantial

---

<sup>1</sup> I have dealt with it in (Floridi 2015). The two articles form a diptych, but they do not presuppose knowledge of each other.

<sup>2</sup> In terms of justice, both papers endorse a Kantian–Rawlsian position, in terms of *respect* for others' freedom, autonomy, and dignity and hence fairness and equal opportunity for all members of society. In the other paper on the paradox of toleration I have argued that toleration can ground justice.

<sup>3</sup> I used the word “dilemma” here to refer to a situation in which a difficult choice has to be made between two or more un/desirable alternatives (cf. the Samaritan's dilemma). I am not using it in the more specific sense in which the alternatives are logically related and their mutual interaction may lead to unacceptable or less acceptable results (cf. the Prisoner's dilemma).

<sup>4</sup> For a similar analysis in the context of public health see Cohen (2013): “Libertarian paternalism's notion of ‘nudging’ refers to steering individual decision making so as to make choosers better off without breaching their free choice. If successful, this may offer an ideal synthesis between the duty to respect patient autonomy and that of beneficence, which at times favors paternalistic influence”, p. 3.

<sup>5</sup> See for example Husak (1981), and Brock (1988), who reviews vol. 3 of Feinberg (1984) and VanDeVeer (1986).

interest in the conflict between autonomy and paternalism and in working out an account of the limits of justified paternalism that gives due weight to the importance of individual autonomy” (550–565). And indeed, such a debate, especially relevant in medical ethics,<sup>6</sup> provides a valuable backdrop for the contextualization of the dilemma. However, in the following pages I shall not focus on it because the fundamental question I intend to address is whether an agent A can be both paternalistic *and* tolerant towards another agent B, *independently* of whether B is autonomous or not. In more scholarly terms, I shall argue that, *pace* Locke, toleration and paternalism can be compatible.<sup>7</sup> For the sake of simplicity, I shall assume that B is indeed autonomous, but even if B were not, even if B’s choices were entirely heteronomous, the dilemma of toleration would still arise, and this, rather than the tension between paternalism and autonomy, is the topic of the article. For if toleration and paternalism can never be compatible, then this can be used as an argument against the adoption of toleration as a design principle that should inform our society, and it is this argument that I wish to counteract (more presently). One final disclaimer: the reader will not find in the following pages a scholarly debate of the dilemma.<sup>8</sup> The plan is to engage with its logical solvability not its interpretative history. I intend to show that the dilemma of toleration is real and pressing but not irresolvable. More specifically, in section two I shall clarify the nature of the dilemma and how it arises. I shall then argue, in section three, that there are several forms of paternalism and that almost all of them, including two currently popular ones, based on *ethics by design* and structural *nudging*, cannot resolve the dilemma of toleration. I specify “almost all of them” because, in section four, I shall argue that one form of choice architecture, which I shall label *pro-ethical design*, can influence choices, while still being fully respectful of freedom. This provides a successful approach (Floridi 2008, 2009) to solve the dilemma, as I shall show in section five, where I argue that *tolerant paternalism* is not an oxymoron. As I anticipated, if I am correct, the consequence is that we can remove a major obstacle that blocks the possibility of adopting tolerant paternalism as one of the fundamental principles informing the design of a democratic and liberal society, as I shall maintain in the conclusion. Let us begin.

## The Dilemma of Toleration

In order to analyse the nature of the dilemma and determine whether *paternalism* may be compatible with *toleration*, we need a more precise understanding of these two italicised concepts. So let me suggest two definitions. They are summarised in

---

<sup>6</sup> See in particular Maehle and Geyer-Kordesch (2002), Bayer and Beauchamp (2007), Nys et al. (2007); and Rothstein (2014).

<sup>7</sup> I am grateful to one of the anonymous reviewers for pointing out this scholarly perspective. As the reviewer argues, and I agree, Locke’s *Letter* and the *Treatises* are forms of anti-paternalism (anti-patriarchalism), see Peter Laslett’s Introduction to Filmer (1949) p. 20 and p. 34. I also agree that Locke’s theory of social justice and how the latter interacts with his ideal of toleration, are problematic issues see for example Euchner (1969), Dunn (1979), von Leyden (1982), and Leyden in Locke (2002).

<sup>8</sup> On this see for example Scanlon (2003), Scoccia (1990), Cohen (2004).

Table 1, at the end of the analysis. They are not meant to be new. And neither definition should be particularly contentious, in the sense that both are consistent with our current understanding of toleration and paternalism.<sup>9</sup> However, if the reader does not recognise them immediately as familiar this may be because I shall formulate them in such a way as to make their logical structures more salient and useful for our comparative analysis. Let us consider toleration first:

TOLERATION: An agent A  $\psi$ -s (acts, behaves, does or does not do something, etc.) tolerantly towards some  $\phi$ -ing (acting, behaving, doing or not doing something, etc.)<sup>10</sup> by another<sup>11</sup> agent B if and only if:

- (i) A is informed<sup>12</sup> about B's  $\phi$ -ing
- (ii) A could  $\psi$  to interfere with B's choice<sup>13</sup> to  $\phi$
- (iii) A disapproves of B's  $\phi$ -ing
- (iv) A refrains from  $\psi$ -ing.

For example, Alice is informed about Bob's smoking and she could forbid him from smoking in her house but, although she disapproves of his habit, she refrains from doing so.

The definition comprises four necessary and jointly sufficient conditions. Each of them comes in degrees, and may generate different issues, either separately or in various combinations. As an illustration, consider (i). Alice may wish not to be entirely or officially informed about Bob's  $\phi$ -ing. This was the "solution" behind the "don't ask, don't tell" official United States policy on service by homosexuals in the military between 1994 and 2011. It is also the most common way of dealing with the sexual life of one's own teenagers. Yet pretending not to know is not a way of being tolerant, it is rather a way of not having to be tolerant. Here, the condition that interests us is the fourth. Strictly speaking, A may refrain from interfering with B's choice to  $\phi$ —for example, the state may decide to make it legal to buy cigarettes for its citizen who are not under age—and yet still interfere with B's choices in order to obtain the same result, through direct persuasion, social pressure, economic disincentives, physical impediments, or other constraints. Of course, such deterrence strategies may be indistinguishable from cases of intolerance by stealth. This happens when an intolerant

---

<sup>9</sup> I have reviewed the literature on toleration in (Floridi 2015). For a valuable review of our understanding of paternalism see Dworkin (2010). Although the format of Dworkin's definition is different, its conceptual content overlaps with the one provided here and any difference is philosophically negligible in this context.

<sup>10</sup> I avoid making things more complex by omitting any propositional attitude modalities, such as B's intending to  $\phi$ -ing, for this would add nothing to the analysis but only muddle it.

<sup>11</sup> The assumption is that A and B are different. When A and B are the same agent there is no dilemma and other problems arise, e.g. self-indulgence, which are beyond the scope of this article.

<sup>12</sup> By "is informed" I mean, throughout the article, "holds the truth" that, or about,  $p$ , not "receives the message" that or about  $p$ .

<sup>13</sup> In this article I refer to B's choice instead of B's freedom, B's autonomy, and B's liberty to do something because such a more-finely grained analysis is unnecessary here. The reader interested in my position on this distinction may wish to see Floridi (2013b).

A would like, but is actually unable to, prevent B from  $\phi$ -ing—so condition (ii) is not satisfied—yet tries whatever else is feasible or allowed in order to achieve the same goal. For example, in 2013 Jack Dalrymple, the Republican Governor of North Dakota, signed into law a state anti-abortion bill banning abortion after the detection of a foetal heartbeat, meaning, in some cases, a ban as early as 6 weeks into the pregnancy (Eligon and Eckholm 2013). It was considered the most restrictive state law in the US, and it clearly tried to limit the application of the US Supreme Court ruling in the Roe v Wade decision of 1973, according to which abortion must be legal until a foetus is viable (usually 22–24 weeks). The Center for Reproductive Rights, the only abortion clinic in North Dakota, sued to overturn the law in July 2013. And in April 2014 District Judge Daniel Hovland overthrew the law declaring it “invalid and unconstitutional” (Wetzstein 2014). When facing intolerance by stealth, we are really dealing with a version of the external problem of intolerance introduced at the very beginning of this article, when referring to zealots. As I anticipated, this is not what I intend to discuss in this article. Although one must acknowledge that, in practice, the boundary between intolerance by stealth and paternalism can be fuzzy and hard to delineate, especially since their outcomes may be indistinguishable, conceptually the difference is clear, for it lies in the altruistic and benevolent motive that should inform a paternalist attitude. The logical problem on which I shall focus here is the possibility of harmonising paternalism with toleration. In order to clarify such a problem, let me introduce the second definition.

Paternalism is meant to be benevolent, but it may turn out to be violent, if it is not constrained by toleration. At the roots of Western philosophy two major texts highlight such a risk. In the *Republic*, when Plato narrates the myth of the cave, the setting is almost disturbing: the philosopher is back inside the cave and violently forces the other people to leave it, causing pain, for each prisoner is

[...] *compelled* to stand up suddenly and turn his head around and walk and to lift up his eyes to the light, and in doing all this *felt pain* and, because of the dazzle and glitter of the light, was unable to discern the objects whose shadows he formerly saw. Plato, *Republic*, 7.515c [my italics].

The Greek word used by Plato for “compelled” is *ἀναγκάζοντο*, which basically means coercing. Today one could translate it as forcing someone at gunpoint. Fast-forward a few centuries and the same violent paternalism is found in the New Testament. The master wants people to join the party he has organised. They seem to be reluctant to join him, so he forces them:

Then the master told his servant, “Go out to the roads and country lanes and *compel* them to come in, so that my house will be full”. *Luke* 14:23 New International Version [my italics].

The original Greek word is the same: *ἀνάγκασον*. Augustine referred to that passage “to justify forcing heretics into unity, or at least conformity, with the Catholic Church”,<sup>14</sup> and it was later used to justify the Crusades.

---

<sup>14</sup> Renard (2012), p. 92.

Clearly, the relation between toleration and paternalism has not been unproblematic. But what exactly do we mean by paternalism? Here is a definition:

**PATERNALISM:** An agent A  $\psi$ -s (acts, behaves, does or does not do something, etc.) paternalistically towards some  $\phi$ -ing (acting, behaving, doing or not doing something, etc.) by another agent B if and only if

- (a) A is informed about B's  $\phi$ -ing
- (b) A could  $\psi$  to interfere with B's choice to  $\phi$
- (c) A is informed about B's  $\phi$ -ing not improving<sup>15</sup> B's or (inclusive or) C's well-being, where possibly B = C
- (d) A does  $\psi$
- (e) A does  $\psi$  because<sup>16</sup> B's  $\phi$ -ing does not improve B's or C's well-being
- (f) A does  $\psi$  without B's consent.

As they say in the logic textbooks, I leave to the reader the simple exercise of checking that both Plato's and Luke's texts fit the definition. Revising our running example, Alice knows that Bob smokes and she could forbid him to smoke in her house, and since she is concerned that smoking is unhealthy for Bob, she does forbid him, for the sake of his health and against his will.

Compare now TOLERATION and PATERNALISM (as a reminder, a quick overview is available in Table 1 below). The first two conditions about information and power are identical. The difference between (iii) and (c) is not one of truth: we shall assume the more difficult case in which A genuinely disapproves of B's  $\phi$ -ing and her information that B's  $\phi$ -ing does not improve B's or C's well-being is entirely correct. The difference between (iii) and (c) is in their ethical orientation: (iii) is agent-oriented, while (c) is patient-oriented, in the following sense. In TOLERATION, A's assessment of B's  $\phi$ -ing is based on A's ethics and may have nothing to do with B's well-being, even despite A's claims. Whereas in PATERNALISM A's assessment is altruistic, especially because of A's claims, yet B's or C's well-being may have nothing to do with A's ethical assessment of the same  $\phi$ -ing, when other agents are in question, including A.<sup>17</sup> Thus, in our example, Alice may paternalistically forbid her teenager son Bob from smoking for the sake of Bob's health without seeking his

---

<sup>15</sup> Note that, in (ii), "not improving" is used to refer to both "decreasing" and "failing to increase" well-being, especially in terms of welfare, interests, good, and values.

<sup>16</sup> This is weaker than if one were to specify "only because". The weaker formulation seems preferable because the presence of further motivations in A's  $\psi$ -ing do not make A's  $\psi$ -ing any less paternalistic, if it is paternalistic in the first place. In our example, the fact that Alice forbids Bob from smoking in her house because she is also concerned about her own health as a passive smoker does not make her action any less paternalistic, if it is paternalistic, but only paternalistic and, say, self-interested. The advantage of a weaker reading is that, of course, the reader who disagrees may simply reinforce the definition by adding "only".

<sup>17</sup> As pointed out by one of the reviewers, A's ethical assessment does not change in the two different situations: what really changes is the assessment of the special circumstances in which the ethical evaluation applies (where A has to grant or deny the practical consequences of the ethical evaluation). In both cases, A judges smoking negatively but sometimes—because of special circumstances - that assessment does not apply to its practical consequences (i.e., the fact of not smoking) but it still applies with regard to its normative dimension (i.e., to consider smoking negatively).

consent, but she may be terminally ill and hence coherently start smoking due to her unconcern about her own health. Finally, note the specification in (c) and (e) that A may behave paternalistically towards B for the sake of B's *or* (inclusive or) C's well-being. It may be that B's  $\phi$ -ing may be fine as far as B's well-being is concerned, yet harmful as far as C is concerned. Imagine the case in which Bob is terminally ill, but Alice paternalistically forbids him from smoking at home because of Bob's daughter Clare, and the negative effects that passive smoking may have on her health. The importance of a third kind of stakeholder will become crucial in the next section, when analysing some forms of paternalism by design.

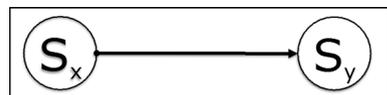
The difference in orientation between (iii) and (c) is interesting, as it shows a clear point of divergence between toleration and paternalism. Yet it is equally clear that this is not the locus where the dilemma emerges. For this, we should rather be looking at the contrast between (iv) and (d–f). These conditions clarify that the same agent A may be asked to respect B's choices and to interfere with them for the sake B's well-being, without B's consent, on the basis of a more privileged epistemic position. Any agent in power faces this dilemma sooner or later. I shall argue that there is a satisfactory way of resolving it, but in order to get there more preparatory work needs to be done. For we need to look first at some varieties of paternalism, and appreciate how many of them fail to cope successfully with the dilemma.

## The Varieties of Paternalism

Paternalism, like toleration, is a second-order way of behaving that qualifies first-order ways of behaving; not unlike “quickly” is a second-order qualification of “walking”. In TOLERATION and PATERNALISM I assumed as much by suggesting an adverbial approach: Alice behaves towards Bob's behaviour *paternalistically* or *tolerantly*. Highlighting such a second-order nature of paternalism helps us devise a simple model through which different forms of paternalism can be compared. For all we need is a first-order change of state in a system (say from Bob not smoking to Bob smoking), represent this more abstractly as a transition of the B system from  $S_x$  to  $S_y$ , and then analyse how such a state transition may be designed by the paternalist agent A. Let us start from the two extreme cases: zero paternalism or total paternalism.

A may not interfere at all with B. If the state transition from  $S_x$  to  $S_y$  is possible it is also allowed, no question asked (see Fig. 1). Stretching the concept a bit, this zero degree of paternalism may be called *anarchism*. Anarchism respects B's freedom at the cost of missing the opportunity to improve B's predicament on good grounds (A knows better) and for the right reason (A acts altruistically, for B's sake). Sometimes such a zero-degree form of paternalism may become hyperprotective, as

Fig. 1 Anarchism



when the state transition from  $S_x$  to  $S_y$  is allowed because it is made as safe as possible. Forms of welfare may be described as being paternalistic in this sense. And the British expression “nanny state” may refer to such overprotective approach by a government or its policies, when the latter interfere unduly with personal choice. In some other cases, they are simply forms of protectionism, as it is the case with mandatory seat belts and air bags. I shall not deal with this form of zero paternalism in the rest of this article because it does not generate a dilemma of toleration, although I shall return to it towards the end of the article, when talking about the best options available when B’s  $\phi$ -ing needs to be modified both paternalistically and tolerantly.

A similar if inverted problem occurs when A completely blocks the transition from  $S_x$  to  $S_y$ . This total degree of paternalism may be called *prohibitionism*. Prohibitionism interferes with B’s choices at the cost of missing the opportunity to tolerate or even foster them (and the ability of making them), for example by allowing B to learn from his mistakes. Note that the asymmetry in scope between anarchism, which is universal (any behaviour), and prohibitionism, which is particular (some behaviour), is only apparent: in principle both are universal. The fact that prohibitionism is always exercised with respect to some specific transition (a particular behaviour is prohibited) is only a contingent accident. In principle, prohibitionism could be applied without any limit to any transition, as anarchism is (Fig. 2).

Neither anarchism nor prohibitionism solves the dilemma of toleration. They merely prevent it from arising by denying that paternalism and toleration may coexist. This is not an objection against their efficacy, but it is a reason to look elsewhere, if what one needs is not the mere prevention of the tension between paternalism and toleration but actually its fruitful resolution in view of a balanced respect for an agent’s choices. For this, one may suggest more compromising approaches, namely *ethics by design* and *nudging*. Both are well known, so a very brief introduction will suffice for the goal of this article.

Ethics by design (van Gorp 2005) is an approach to protecting and fostering what are considered ethical values<sup>18</sup>—such as privacy, security, safety, or sustainability—by embedding them from the beginning into the design specifications of technologies, procedures, practices, or infrastructures. In terms of policies, ethics by design may also be extended to cover the adoption of incentives and disincentives. Finally, ethics by design is strictly related to *value sensitive design* (Friedman 1996; Friedman et al. 2013) and may actually be treated as synonymous with it (van der Hoven and Manders-Huits 2009), while its best-known implementation is *privacy by design* (Hes and Borking 1995; Hustinx 2010). A building that, by design, may be navigated more easily by blind people provides a simple example of ethics by design. In our state-transition model, ethics by design, as a form of paternalism, makes going from  $S_x$  to  $S_y$  more or less difficult for B (Fig. 3 illustrates a case in which the transition is made more difficult). For example, Alice may remove all the ashtrays in her house, in order to make Bob’s smoking there a bit more difficult (e.g. socially awkward).

---

<sup>18</sup> In Floridi (2013a) I discuss them as features of a society’s ethical infrastructure.

Fig. 2 Prohibitionism

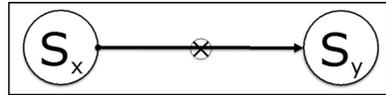
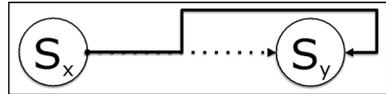


Fig. 3 Ethics by design



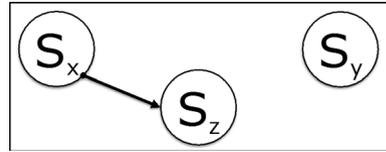
Nudging rose to fame as an approach to non-forced compliance in 2008, with the publication of *Nudge: Improving Decisions About Health, Wealth and Happiness*, the popular book by Thaler and Sunstein (2008, 2009). Since then, administrations, especially in the US and UK, have adopted it as part of their policies to improve public health and related fields, such as organ donation and nutrition habits. One difficulty in analysing nudging is that it is a rather flexible notion, which can be adapted to describe quite a range of different strategies. Thaler and Sunstein define a nudge in terms of a negation, by emphasising how it *does not forbid* a choice: “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler and Sunstein 2008) p. 6. However, I agree with Selinger and Whyte (2012) that such a notion of “nudging” may be used too broadly and sometimes inconsistently in the literature. Di Nucci (2013), for example, distinguishes between “hard” and “soft” nudges. The problem seems to be the focus on “altering people’s behaviour” without forcing them to act in a specified way. This is a rather weak condition, which can easily lead to ambiguity: how strong can the nudge be before it becomes a shove so strong to eliminate *de facto* any alternative option? In this article, I will not try to answer this question. Instead, I shall focus on the design of the environment within which people make decisions (Floridi 2011). So I shall distinguish between two kinds of nudging, one *structural*, and the other *informational*.

Structural nudging is ontological, for it changes the nature of the actual courses of actions available to an agent. Modifying the order in which food is presented in a school cafeteria is a textbook case of a structural nudge. When made increasingly compelling—by hiding unhealthy food, putting it out of reach, or making it too expensive, for example—this sort of nudge can lead to *de facto* forced compliance.

Informational nudging is epistemological, for it changes the nature of information to which an agent is exposed in order to obtain a goal. Labelling with clear colours the more or less healthy properties of food is an informational nudge. It is much more difficult to turn informational nudging into *de facto* forced compliance. I shall say more on informational nudging in the next section. In the rest of this section I shall focus only on structural nudging.

Similarly to ethics by design, structural nudging seeks to shape agents’ environments and their available courses of action. When it is not just another form of value-sensitive design—and it can easily be so—it is a strategy that modifies choices or behaviours by providing alternative options. This is easier to grasp if we refer to our state-transition

Fig. 4 Structural nudging



model. Instead of making going from  $S_x$  to  $S_y$  more or less difficult for B, nudging makes it easier for B to move from  $S_x$  to another state different from  $S_y$ , call it  $S_z$  (see Fig. 4). For example, Alice may offer to Bob an electronic cigarette, thus nudging him towards a less unhealthy choice and perhaps quitting altogether.

The main problem with both ethics by design and structural nudging approaches is that the more they succeed in shaping B's behaviour—the better they implement forms of paternalism—the less they succeed in respecting B's choices (and, when available, the more or less autonomous decisions behind them). When they are pursued quietly, without making B aware of the process, they can become dangerous forms of illiberal manipulation, depending on which power is in charge. By progressively designing B's space of options (choice architecture) and firmly steering B towards specific behaviours, A may end up taking away any ability to choose on B's side, who becomes a mere pawn in the hands of the designer. This is why the debate about nudging in health care tends to concern the compatibility of nudging with informed consent and the preservation of the ability to choose.<sup>19</sup> Two textbook examples help to clarify the point.

In terms of ethics by design, speed bumps are a kind of traffic calming measure designed to slow down vehicles and improve safety. They may seem like a good idea, but they involve a physical alteration of the road, which is permanent and leaves no real choice to the driver. This means that emergency vehicles, such as a medical ambulance, a police car, or a fire engine, must also slow down, even when responding to an emergency.

In terms of nudging, an important case study is provided by organ donation. This is one of the central issues in health policy debates because there is a growing shortage of transplantable organs in the world. An influential way of raising the number of registered organ donors is by shifting the design of a choice architecture from *explicit consent* (opt-in) to *presumed consent* (opt-out). In countries with a presumed consent system, the default is that anyone who does not explicitly refuse to donate is considered a donor, and since people tend not to opt-out, the result is that organ donation consent is vastly higher than in countries with an opt-in system. Austria and Germany are two comparable countries, but the former relies on presumed consent and has a consent rate of 99 %, whereas the latter relies on explicit consent and has a consent rate of just 12 % (Thaler 2009).

---

<sup>19</sup> Unsurprisingly, the debate does not focus on the ethical and political issue of toleration. For a review of some main ethical issues related to nudging in health care see Blumenthal-Barby and Burroughs (2012). Cohen (2013) assess positively the merits of nudging in health care, with further support from Brooks (2013) and Saghai (2013). Their broad view is that nudging can be made compatible with respect for patients' informed consent. Huang and Baum (2012), Bell et al. (2013), and Sagoff (2013) are much more critical. For a strong defence of paternalism in public health see Verweij and van den Hoven (2012).

At this point, one may wonder whether there is anything wrong with ethics by design or structural nudging. As someone who approves of organ donation and is weary of human inertia and biases, I tend to be sympathetic towards the adoption of a choice architecture based on presumed consent. However, it may represent a problem for some Muslims, for example, who may be uncertain about whether organ donation is forbidden according to their religious tradition.<sup>20</sup> More generally, it can subtly constrain one's ability to choose in a way that may be difficult to withstand. Organ donation may not be very controversial, but what if someone starts nudging agents in a direction they would not approve of? A third example brings the point home more vividly. When taxi drivers in New York adopted credit card readers, they also embraced a presumed consent choice architecture. The checkout screen shows a 20 % tip as apparently the smallest option. It can be modified, but doing so is not easy, requires time, and may be socially embarrassing when the difference is only a few dollars. Passengers have been nudged. "20 % became the new norm overnight. It's a strategy known as anchoring, subtly or not so subtly establishing a new standard by shifting the choices you present"(Park 2014).

None of the previous alternatives appear to provide a good balance between paternalism, toleration, and individual choices. When they seem acceptable, the reasoning presupposes a high degree of trust in those who determine the values that are being embedded in the design or who control the type, degree, and direction of the structural nudging exercised. Is there a better option than just trusting those in power to know better and always to do the right thing?

## **Pro-ethical Design and the Resolution of the Dilemma of Toleration**

Ethics by design and structural nudging, as described above, are unsatisfactory because they tend to shape the actual options available to an agent, and rely on human inertia and biases as the pragmatic motivation for action. By contrast, what I shall call *pro-ethical design* is an approach to choice architecture that shapes only the *information* about the actual options available, relying on human intelligence and explicit interests as the pragmatic motivation for action. It may be seen as a more tolerant version of informational nudging that not only is respectful of agents' ability to choose—irrespective of whether this is interpreted as autonomous or heteronomous—but actually foster their abilities to decide about their actions more critically. To put it more theoretically, ethics by design and structural nudging shape the system and hence tend to disempower agents, whereas pro-ethical design and informational nudging shape the model, that is, the level of abstraction at which the system is epistemically accessed, and hence force agents to "empower themselves" through their informational choices. The difference may be explained more intuitively by expanding on the previous examples.

Speed cameras are another kind of traffic calming measure designed to slow down vehicles and improve safety. However, they are a case of pro-ethical design

---

<sup>20</sup> Organ donation is accepted by the majority of Islamic religious authorities, but some conflicting opinions have caused uncertainty among Muslims, see Moazam (2006) and Padela and Curlin (2013).

because they work as warnings: they shape not the physical but the informational environment within which drivers have to exercise their intelligent evaluation. They send a clear message (this is why they *should*<sup>21</sup> be big, bright, and located very visibly) but still leave it to drivers to choose whether to disregard the speed limit and accept to pay the ticket if they wish, for example in the case of an emergency.<sup>22</sup> Similarly, *mandated choice* is a choice architecture supported by a pro-ethical design approach. It is neither opt-in nor opt-out by default (both options shape choices), but requires people who wish to achieve a particular goal (e.g. renewing a driving license) to choose whether they are willing to engage in a particular action or make a particular choice (e.g. be an organ donor or not) before moving forward with their course of action, which remains available and unmodified, no matter what choice is made, as long as it is made. This is the sense in which pro-ethical design shapes only the informational space.

In our state-transition model, pro-ethical design does not presuppose autonomy but it is comparable to a tolerant version of *libertarian paternalism*.<sup>23</sup> It makes going from  $S_x$  to  $S_y$  conditional on going through  $S_z$ , where  $S_z$  is now an informational variable (see Fig. 5). This is the sense in which pro-ethical design modifies the level of abstraction of the choice architecture.<sup>24</sup> Further revising the running example, Alice knows, and is concerned, that Bob smokes, does not forbid him from doing so in her house but, for the sake of his health and against his will, she reminds him of how unhealthy it is and the serious risks it involves, asking him whether he may not reconsider his choice. More realistically, pro-ethical design is the strategy that orients the legislation on so-called cigarette package health warnings.

Before seeing how pro-ethical design can help resolve the dilemma of toleration, let me dispose of two practical objections and introduce an equally practical advantage.

The first practical objection is that pro-ethical design is always less successful than ethics by design or structural nudging. An extended study in the Netherlands (van Dalen and Henkens 2014) has recently provided empirical evidence to the contrary.<sup>25</sup> The researchers compared three defaults in organ donation systems: mandated choice, presumed consent (opt-out), and explicit consent (opt-in). The result was that explicit consent is the least successful, as one may expect, but that

---

<sup>21</sup> I am aware that in some countries, e.g. Italy and (I am told) Australia, they are not, where they may be conceived more as invisible threats or even sources of revenue. This is not the point under discussion here, but, if pressed, I would argue that this is a mistake in choice architecture, hence the italicized “should” in the text.

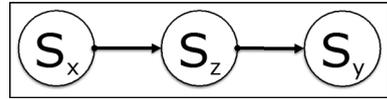
<sup>22</sup> There is a distinction here between information about harm—e.g. smoking causes cancer—and information about harm that is backed by a threat—e.g. a ticket.

<sup>23</sup> The expression is introduced by Thaler and Sunstein (2008), but in their work it also covers what I have defined here as structural nudging, and this is problematic from the point of view of toleration and respect for freedom, as I argued above. Menard (2010) evaluates libertarian paternalism from an ethical point of view.

<sup>24</sup> I owe this clarification to an enlightening conversation with Massimo Durante during a graduate seminar at the Faculty of Law of the University of Turin in which I presented an earlier version of this article.

<sup>25</sup> For earlier evidence in the US see Spital (1995).

**Fig. 5** Pro-ethical design  
(informational nudging)



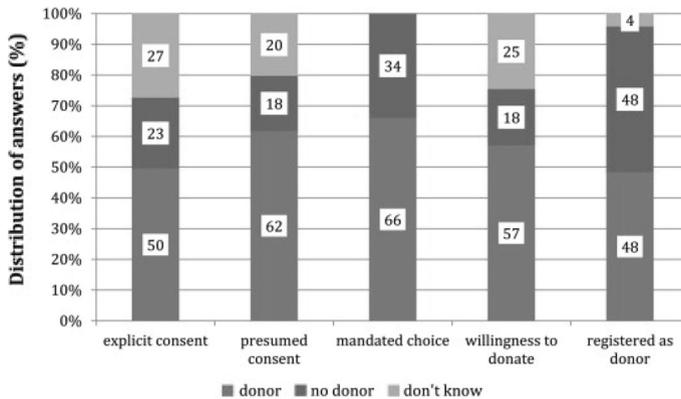
mandated choice, which grounds a pro-ethical design strategy, is as effective and perhaps even more effective than presumed consent at ensuring that individuals do register as donors. This because mandated choice zeroes the number of individuals who remain undecided (see the third column in Fig. 6).

So it seems that pro-ethical design can be at least as pragmatically effective as ethics by design or structural nudging, at least in some cases (more on this qualification in the next section).

The second practical objection is that the analysis developed so far seems affected by an ambiguity, namely whether various nudging policies are reviewed and approved of by the affected parties. If so, it is unclear how paternalistic a policy really is, at least if paternalism is defined as intervention against the person's will, decisions, or choices. If such policies are kept secret or (worse) instituted against the expressed wishes of a majority then it is unclear that such policies are democratic. The answer to this objection is that it is based on a misunderstanding. First, recall that both A and B are two variables: they could be interpreted as two single agents (Alice and Bob, as I have simplified in the examples), but also as an individual and a group or vice versa (as in Plato), or as two groups of agents (as in Luke and Augustine). I specified earlier that A and B are different and that A has power over B. If A and B belong to the same group and this group has democratically accepted the nudging in question, then we are not facing the dilemma that is the topic of this article. So either the objection is correct but does not apply, or it applies but then it is incorrect.

Finally, the practical advantage of pro-ethical design concerns the agents' "ownership" of their choices. Ethics by design and structural nudging, by shaping the courses of actions rather than the informational interface, run the risk of ethically desensitising, de-skilling, and de-responsabilising agents, who may end up being merely herded, mindlessly and non-responsibly, towards some pre-established options chosen by the designers of the environment. On the contrary, pro-ethical design is meant to educate agents to make their own (autonomous, if the reader endorses this qualification) critical choices and assume explicit responsibilities. It fosters the ability to choose (freedom or autonomy, again, if the reader intends to adopt a more loaded framework of analysis) rather than weakening it. For this reason, it has the further advantage that, after making a decision, agents tend to change their original preferences and align them a posteriori<sup>26</sup> more closely with the choice that they actually made, both morally and psychologically. In the latter case, this is known in neuroscience as choice-induced preference modification (Egan

<sup>26</sup> In the course of a discussion of this paper a critic argued that this is not "a posteriori" but "post hoc". Since "a posteriori" may simply mean "epistemically post hoc" (as opposed to, for example, "causally post hoc"), I remain unconvinced, for it seems to me that what is at stake here is *both* an epistemic state *and* a chronological order in its change. However, this is a subtle point and the reader should feel free to replace "a posteriori" with the more inclusive and generic "post hoc" if this seems preferable.



**Fig. 6** “Organ donation under different scenarios and actual registration as donor, the Netherlands”.  
 Source: van Dalen and Henkens (2014), p. 139

et al. 2010; Izuma and Murayama 2013). “Humans tend to modify their attitudes to align such attitudes with past actions. For example, after choosing between similarly valued alternatives, people rate the selected option as better than they originally did, and the rejected option as worse” (Sharot et al. 2009), p. 3760. In other words, we (unconsciously) choose to be happy with what we have chosen: the process of making a choice itself induces preferences, even when preferences cannot guide choices (e.g. in blind, two-choice tasks), further reinforcing a sense of coherence and ownership of our own decisions, quite independently of the whole debate about the autonomy or heteronomy of the choices in question.

Finally, pro-ethical design is also morally preferable because it can resolve the dilemma of toleration, as I shall argue in the next section.

## Tolerant Paternalism

We saw that pro-ethical design supports an informational version of paternalism, which interferes only with B’s ignorance. Ignorance about a topic is here defined as the absence of both a question and the corresponding correct answer about a topic. Consider the following example. Suppose Bob is *informed* that “Paris is the capital of France”. This can be rephrased as saying that Bob is informed that “Paris” is the correct answer to the question “Which city is the capital of France?”. Now, Bob is *uncertain* if he holds only the question without having the answer. And he is *ignorant* if he holds neither the question nor the answer.<sup>27</sup> Removing Bob’s ignorance about *x* means making him face the relevant question about *x*. Pro-ethical design forces B to acknowledge the presence of a question and to answer it. And since the question is a personal choice not a test, the answer is a decision that is entirely up to B, who is forced to make up his mind about the topic selected by the

<sup>27</sup> When Bob holds the question but the wrong answer then this is not a second kind of ignorance but rather a case of misinformation and it is not in question here.

designer of the choice architecture and accept the responsibility of his choice. Since pro-ethical design operates at the informational level, it avoids falling into *pragmatic paternalism*, which interferes with B's choices directly, not just with his informational state, for it does not prescribe at all (not even in a mildly nudging way) which answer B may give or which decision he may have to take. By adopting a pro-ethical design strategy, A can therefore refrain from interfering directly with B's ability to behave as he wishes, while still seeking to improve B's or C's well-being, by presenting B with the right kind of informational questions to answer. In more Kantian terms, it forces agents to exercise their autonomy consciously, to embrace a sort of ethical *sapere aude*, by making them dare to take autonomous decisions about informational choices and bear their responsibilities. It is still a form of paternalism, because it does all this without B's consent (B did not ask to be forced to make the informational choice in question) and without asking B what kind of choices he is forced to make up his mind about. In other words, A forces B to play a Wittgensteinian language-game with significant pragmatic consequences for future actions. This ends up expanding the scope of responsibilities of the agents in question and hence their actual ability to choose. Here is a more formal definition:

TOLERANT PATERNALISM: An agent A  $\psi$ -s (acts, behaves, does or does not do something, etc.) *both* tolerantly *and* paternalistically towards some  $\phi$ -ing (acting, behaving, doing or not doing something, etc.) by another agent B if and only if

- (1) A is informed about B's  $\phi$ -ing
- (2) A could  $\psi$  to interfere with B's choice to  $\phi$
- (3) A is informed about B's  $\phi$ -ing not improving B's or (inclusive or) C's well-being, where possibly B = C
- (4) A refrains from  $\psi$ -ing
- (5) A could  $\chi$  to make B's  $\phi$ -ing conditional on B making a choice about  $\gamma$
- (6) A does  $\chi$  because B making a choice about  $\gamma$  could improve B's or (inclusive or) C's well-being, where possibly B = C
- (7) A does  $\chi$  without B's consent.

Table 1 summarises the three definitions.

The first two conditions are identical to the corresponding conditions (i)-(ii) and (a)-(b) in the definitions of TOLERATION and PATERNALISM, and represent the common roots of all three positions: information and power (Table 1). Conditions (3) and (c) are where TOLERANT PATERNALISM is identical to PATERNALISM and differs from TOLERATION: the same comments made above about the agent—vs. the patient-oriented nature of (iii) vs. (c) apply. But note that now condition (4) is identical to (iv) and differs from (d). This is where TOLERANT PATERNALISM is identical to TOLERATION and differs from PATERNALISM. This potential tension is resolved by adding conditions (5)–(7), which combine TOLERATION with an informational version of paternalism different from PATERNALISM. The advantage is that A refrains from interfering with B's choice to pursue his course of action, but makes the latter

**Table 1** The definitions of toleration, paternalism, and tolerant paternalism

Toleration	Paternalism	Tolerant paternalism
(i) A is informed about B's $\phi$ -ing	(a) A is informed about B's $\phi$ -ing	(1) A is informed about B's $\phi$ -ing
(ii) A could $\phi$ to interfere with B's choice to $\phi$	(b) A could $\psi$ to interfere with B's choice to $\phi$	(2) A could $\psi$ to interfere with B's choice to $\phi$
(iii) A disapproves of B's $\phi$ -ing	(c) A is informed about B's $\phi$ -ing not improving B's or C's well-being, where possibly B = C	(3) A is informed about B's $\phi$ -ing not improving B's or C's well-being, where possibly B = C
(iv) A refrains from $\phi$ -ing	(d) A does $\psi$	(4) A refrains from $\psi$ -ing
	(e) A does $\psi$ because B's $\phi$ -ing does not improve B's or C's well-being	(5) A could $\chi$ to make B's $\phi$ -ing conditional on B making a choice about $\gamma$
	(f) A does $\psi$ without B's consent	(6) A does $\chi$ because B making a choice about $\gamma$ could improve B's or C's well-being, where possibly B = C
		(7) A does $\chi$ without B's consent

conditional on B making an explicit and autonomous choice about a topic (which may be different from that course of action, more on this presently), in a way that is independent of, and that possibly violates B's consent (B may not have consented to be forced to make a choice about that topic).

I anticipated in the previous section that it is important to add considerations about a third stakeholder C, who is potentially different from A and B. This should now be clear: TOLERANT PATERNALISM shares with nudging-based PATERNALISM the possibility of forcing B to make a choice, as a necessary condition for B to achieve a particular goal, about something that does not have to concern either B or the goal itself. Pro-ethical design can be used to request agents to make up their minds and take decisions about issues that are unrelated to the goal they wish to pursue, and that affect not just themselves but also others, and sometimes only others, as in the case of deciding whether to become an organ donor after death. Here is another simple example. ATM machines have been designed for some time to request users to decide whether they wish to make a donation before being allowed to check their balance or take out cash. Both the decision and the goal concern money, but the donation already concerns people different from those making it ( $B \neq C$ ). Nothing prevents a choice architect from using ATM machines to require users to make up their mind about other topics, e.g. whether they wish to become organ donors. This is similar to the kind of nudging strategies that make registering for, or renewing, a driving licence online conditional upon deciding whether to become an organ donor

(Eslambolchilar and Rogers 2013). In short, pro-ethical design can inherit from structural nudging some of its most positive effects, while still safeguarding toleration and respect for individual preferences, including the wish not to be implicitly forced into a position, and fostering the development of a more critical understanding of one's own choices.

## **Conclusion: Tolerant Paternalism as a Default Policy**

Pro-ethical design has many advantages, and it can provide a fair balance between toleration, paternalistic benevolence, and individual choices, but it is not a panacea. There are cases in which human rationality is counterbalanced by other interests, such as “consumer surplus” in the case of smoking, which refers to the lost pleasure smokers experience when they stop smoking (Song et al. 2014). There are cases in which forcing agents to choose even if they are not (and may not be expected to be) well informed about either alternative seems to be unfair or too demanding.<sup>28</sup> And there are cases in which what is at stake is too important, and mistakes are irreversible or too severe, not to force agents to take the course of action that is known to be right. A car still allows one to drive, even if a warning system is activated because the seatbelt is not fastened. And an air bag makes no difference to one's driving style or options, only to one's potential safety should something go wrong, i.e. counterfactually. However, the same car may simply not start unless the clutch is in neutral, leaving no choice to the driver but to comply, because the alternative may be too risky. It follows that there should be a variety of flexible approaches to choice architecture, depending on what is at stake and the goals one wants to achieve. As technologies develop and our information societies become increasingly dependent on complex digital systems, the temptation of shaping the available options and the courses of actions for entire segments of the population through a variety of hard and soft structures or codes (Lessig 1999) may have to be resisted more strongly. Policy makers should evaluate carefully whether the risk of allowing people to make the wrong choice or no choice at all requires less liberal measures. When paternalism needs to be exercised (recall that a zero paternalism option may also be available), that is, when A needs to affect B's  $\phi$ -ing, the default option, albeit overridable, should be in favour of a tolerant paternalism and hence of a pro-ethical design whenever possible. Ethics by design, structural nudging, and opt-out choice architectures should not be the preferred approach but rather a fallback when pro-ethical design, informational nudging and mandated choice architectures fail to be a satisfactory solution. Tolerant paternalism should always be favoured whenever possible, for the price paid to forfeit it is a limitation in individuals' ability to choose (independently of whether this is an instance of freedom or not) that may be unnecessarily intolerant, democratically dangerous, pedagogically counter-productive, and morally de-responsabilising.

---

<sup>28</sup> I am grateful to Stephen Clarke for having called my attention to this important point. I intend to deal with the tension between mandate choice and informed consent in a separate paper because this is an issue too significant and rich in consequences to be treated in this context.

**Acknowledgments** I discussed several versions of this article on many occasions and I am deeply indebted for their feedback to many colleagues and members of the audiences. In particular, I wish to thank Andrew A. Adams, Pompeu Casanovas, Stephen Clarke, Ralph Schroeder and above all Massimo Durante and Ugo Pagallo, for their insightful comments and the opportunities they kindly provided to discuss the ideas presented in this article. The following meetings were especially useful for discussing and improving the final version: an invited talk at The Ethical and Legal Aspects of Digital Security Special Workshop on Digital Security and Data Protection (Paris, 29 November 2013); a presentation discussed during a Master Class Seminar at the Oxford Internet Institute (Oxford, 10 June 2014); an invited talk at the University of Turin, Faculty of Law, Ph.D. program (Turin, 26–27 June 2014); and a keynote at the Asian Privacy Scholars Network, 4th International Conference, Meiji University (Tokyo, 10–11 July 2014). Three anonymous reviewers offered remarkably detailed, long and insightful comments on a previous draft, and this version is much better thanks to them. I am very grateful for their time and scholarship. David Sutcliffe carefully copyedited and commented on the last draft of this paper, improving it substantially.

## References

- Bayer, R., & Beauchamp, D. E. (2007). *Public health ethics: Theory, policy, and practice*. Oxford; New York: Oxford University Press.
- Bell, E., Dubljevic, V., & Racine, E. (2013). Nudging without ethical fudging: Clarifying physician obligations to avoid ethical compromise. *The American Journal of Bioethics*, 13(6), 18–19.
- Blumenthal-Barby, J. S., & Burroughs, H. (2012). Seeking better health care outcomes: The ethics of using the “nudge”. *The American Journal of Bioethics*, 12(2), 1–10.
- Brock, D. W. (1988). Paternalism and autonomy. *Ethics*, 98(3), 550–565.
- Brooks, T. (2013). Should we nudge informed consent? *The American Journal of Bioethics*, 13(6), 22–23.
- Cohen, A. J. (2004). What toleration is. *Ethics*, 115(1), 68–95.
- Cohen, S. (2013). Nudging and informed consent. *The American Journal of Bioethics*, 13(6), 3–11.
- Di Nucci, E. (2013). Habits, nudges, and consent. *The American Journal of Bioethics*, 13(6), 27–29.
- Dunn, J. (1979). *Western political theory in the face of the future, themes in the social sciences*. Cambridge; New York: Cambridge University Press.
- Dworkin, G. (2010). “Paternalism.” *The Stanford Encyclopedia of Philosophy* available online.
- Egan, L. C., Bloom, P., & Santos, L. R. (2010). Choice-induced preferences in the absence of choice: Evidence from a blind two choice paradigm with young children and capuchin monkeys. *Journal of Experimental Social Psychology*, 46(1), 204–207.
- Eligon, J., & Eckholm, E. (2013). *New laws ban most abortions in North Dakota*. New York: New York Times.
- Eslambolchilar, P., & Rogers, Y. (2013). Theme issue on persuasion, influence, nudge, or coercion (PINC). *Personal and Ubiquitous Computing*, 17(6), 1171–1172.
- Euchner, W. (1969). *Naturrecht und Politik bei John Locke, Kritische Studien zur Politikwissenschaft*. Frankfurt am Main: Europäosche Verlagsanstalt.
- Feinberg, J. (1984). *The moral limits of the criminal law* (Vol. 4). New York: Oxford University Press.
- Filmer, R. (1949). *Patriarcha and other political works*. Oxford: Basil Blackwell.
- Floridi, L. (2008). Trends in the philosophy of information. In P. Adriaans & J. F. A. K. van Benthem (Eds.), *Handbook of philosophy of information* (pp. 113–131). Amsterdam; London: North-Holland.
- Floridi, L. (2009). The information society and its philosophy. *The Information Society*, 25(3), 153–158.
- Floridi, L. (2011). A defence of constructionism: philosophy as conceptual engineering. *Metaphilosophy*, 42(3), 282–304.
- Floridi, L. (2013a). Distributed morality in an information society. *Science and Engineering Ethics*, 19(3), 727–743.
- Floridi, L. (2013b). *The ethics of information*. Oxford: Oxford University Press.
- Floridi, L. (2015). Toleration and the design of norms. *Science and Engineering Ethics*, 21(5), 1095–1123.
- Friedman, B. (1996). Value-sensitive design. *Interactions*, 3(6), 16–23.
- Friedman, B., Kahn, P., Jr, Borning, A., & Hultdtgren, A. (2013). Value sensitive design and information systems. In N. Doorn, D. Schuurbiens, I. van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 55–95). New York: Springer.
- Hes, R., & Borking, J. J. F. M. (1995). *Privacy-enhancing technologies: The path to anonymity*. Registratiekamer.

- Huang, C. J., & Baum, M. L. (2012). Nudge ethics: Just a game of billiards? *The American Journal of Bioethics*, 12(2), 22–24.
- Husak, D. N. (1981). Paternalism and autonomy. *Philosophy & Public Affairs*, 10(1), 27–46.
- Hustinx, P. (2010). Privacy by design: Delivering the promises. *Identity in the Information Society*, 3(2), 253–255.
- Izuma, K., & Murayama, K. (2013). Choice-induced preference change in the free-choice paradigm: A critical methodological review. *Frontiers in Psychology*, 4, 1–12.
- Lessig, L. (1999). *Code: And other laws of cyberspace*. New York: Basic Books.
- Locke, J. (1689). *A letter concerning toleration*. Translated from the Latin of John Locke by W. Popple: London.
- Locke, J. (1991). *A letter concerning toleration*. Edited by John Horton and Susan Mendus for the Routledge philosophers in focus series. London: Routledge.
- Locke, J. (2002). *Essays on the law of nature: The Latin text with a translation, introduction and notes, together with transcripts of Locke's shorthand in his journal for 1676*. Oxford: Clarendon.
- Machle, A.-H., & Geyer-Kordesch, J. (2002). *Historical and philosophical perspectives on biomedical ethics: From paternalism to autonomy?*. Aldershot: Ashgate.
- Menard, J.-F. (2010). A 'nudge' for public health ethics: Libertarian paternalism as a framework for ethical analysis of public health interventions? *Public Health Ethics*, 3(3), 229–238.
- Moazam, F. (2006). *Bioethics and organ transplantation in a Muslim society: A study in culture, ethnography, and religion*. Bloomington, IN: Indiana University Press.
- Nys, T., Denier, Y., & Vandeveldt, T. (2007). *Autonomy & paternalism: Reflections on the theory and practice of health care, Ethical perspectives monograph series*. Leuven; Dudley, MA: Peeters.
- Padela, A. I., & Curlin, F. A. (2013). Religion and disparities: Considering the influences of Islam on the health of American Muslims. *Journal of Religion and Health*, 52(4), 1333–1345.
- Park, M. Y. (2014). *How tipping is changing in the digital age*. The Huffington Post.
- Renard, J. (2012). *Fighting words: Religion, violence, and the interpretation of sacred texts*. Berkeley, CA; London: University of California Press.
- Rothstein, M. A. (2014). Autonomy and paternalism in health policy: Currents in contemporary bioethics. *The Journal of Law, Medicine & Ethics*, 42(4), 590–594.
- Saghai, Y. (2013). Salvaging the concept of nudge. *Journal of Medical Ethics*, 39(8), 487–493.
- Sagoff, M. (2013). Trust versus paternalism. *The American Journal of Bioethics*, 13(6), 20–21.
- Scanlon, T. M. (2003). *The difficulty of tolerance: Essays in political philosophy*. Cambridge: Cambridge University Press.
- Scoccia, D. (1990). Paternalism and respect for autonomy. *Ethics*, 100(2), 318–334.
- Selinger, E., & Whyte, K. P. (2012). What counts as a nudge? *The American Journal of Bioethics*, 12(2), 11–12.
- Sharot, T., De Martino, B., & Dolan, R. J. (2009). How choice reveals and shapes expected hedonic outcome. *The Journal of Neuroscience*, 29(12), 3760–3765.
- Song, A. V., Brown, P., & Glantz, S. A. (2014). When health policy and empirical evidence collide: The case of cigarette package warning labels and economic consumer surplus. *American Journal of Public Health*, 104(2), 42–51.
- Spital, A. (1995). Mandated choice: A plan to increase public commitment to organ donation. *JAMA*, 273(6), 504–506.
- Thaler, R. H. (2009). *Opting in vs opting out*. New York: The New York Times.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness* (Rev. and expanded ed.). New York: Penguin Books.
- van Dalen, H. P., & Henkens, K. (2014). Comparing the effects of defaults in organ donation systems. *Social Science and Medicine*, 106, 137–142.
- van der Hoven, J., & Manders-Huits, N. (2009). Value-sensitive design. In *A companion to the philosophy of technology* (pp. 477–480). Wiley-Blackwell.
- van Gorp, Anke. (2005). *Ethical issues in engineering design; safety and sustainability*. Tilburg: 3TU Ethics.
- VanDeVeer, D. (1986). *Paternalistic intervention: The moral bounds of benevolence, Studies in moral, political, and legal philosophy*. Princeton, NJ: Princeton University Press.
- Verweij, M., & van den Hoven, M. (2012). Nudges in public health: Paternalism is paramount. *The American Journal of Bioethics*, 12(2), 16–17.

- von Leyden, W. (1982). *Hobbes and Locke: The politics of freedom and obligation* (p. 1981). London: Macmillan.
- Wetzstein, C. (2014). *Pro-life setback: Judge voids N. Dakota's 'heartbeat' abortion law*. Washington: The Washington Times.