Nudging and Informed Consent

<table>
<thead>
<tr>
<th>Journal:</th>
<th>American Journal of Bioethics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>UAJB-2012-0316.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Target Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>informed consent, decision making, moral theory, PHILOSOPHY, professional ethics, professional-patient relationship</td>
</tr>
</tbody>
</table>
Nudging and Informed Consent

SHLOMO COHEN

Shlomo Cohen  
Department of Philosophy  
Diller family building  
Ben-Gurion University  
P.O. Box 653, Be’er-Sheva 84105,  
ISRAEL

Email: shlomco@mail.huji.ac.il
Nudging and Informed Consent

ABSTRACT: 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. A growing body of literature attempts to assess the merits of nudging in healthcare. However, this literature deals almost exclusively with health policy, while the question of the potential benefit of nudging for the practice of informed consent has escaped systematic analysis. This paper focuses on this question. While it concedes that nudging could amount to improper exploitation of cognitive weaknesses, it defends its practice in a wide range of other conditions. The conclusion is that, when ethically legitimate, nudging offers an important new paradigm for informed consent, with a special potential to overcome the classical dilemma between paternalistic beneficence and respect for autonomy.

The duty to obtain an informed consent (IC) from patients before any significant medical intervention reflects the central value placed on patient autonomy. The contemporary prominence of this ethical view often supplants the primacy accorded to the value of beneficence (and nonmaleficence) in the history of medicine (Faden and Beauchamp 1986). The potential clash between these two fundamental values, of respect for autonomy and of (paternalistic) promotion of well-being, is a source of some of the toughest dilemmas in medical ethics. These statements are virtually truisms. The recent advent of the idea of “libertarian paternalism” (LP), of interference with individual decision-making so as to make choosers better-off while preserving freedom of choice (Thaler and Sunstein 2003), offers the hope of an ideal synthesis of the two basic values. Attempts have been made in recent years to analyze LP’s fruitfulness in healthcare. While those attempts focused mainly on policy issues, this paper offers a yet undeveloped analysis of its potential contribution to the ethical practice of IC.
“The nudge”

As Richard Thaler and Cass Sunstein presented it, LP operates through the “nudge.” 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 incentives” (Thaler and Sunstein 2008, 6). Rather, nudges exert their influence through “choice architecture” i.e. simply by “organizing the context in which people make decisions” (ibid, 3). LP is paternalistic in that it aims to influence people through means other than rational persuasion to make choices that are perceived as good for them or even that protect them from themselves. Yet LP does not threaten liberty: it neither blocks any choice nor makes it (more than trivially) costlier or more burdensome (in terms of time, money, bureaucracy, etc.). Nudging through changing the default choice is a paradigm example where significant—even dramatic—shifts in public behavior can be achieved while leaving the choice set unchanged and all choices readily accessible (Halpern et al. 2007; Blumenthal-Barby and Burroughs 2012). Crucially, choosing differently from what the nudge suggests must be very easy; and while there is no objective measure to what “easy” is, some norms are commonly agreed upon—for example, that a mouse click is easy, whereas picking up the phone to communicate one’s different choice is already onerous, even if mildly so.

LP sounds like an oxymoron (Vallgarda 2012), but proponents insist it is not (Sunstein and Thaler 2003). “Choice architecture” is the key idea: as long as nudging tinkers only with the structure or environment of choice, we can influence the choices made while remaining loyal to libertarianism’s ideal of not interfering with the chooser’s free actions. If, for example, as a health-promoting measure, we put the healthy foods at eye-level at the beginning of the cafeteria

URL: http://mc.manuscriptcentral.com/uajb
shelf, they will be chosen more often and other items less often, although the freedom to pick any item remains unchanged. Incidentally, this illustrates how the successful practice of choice architecture presupposes knowledge of the empirical science of human decision-making, based on the relevant cognitive and behavioral sciences. Questions on the practice of nudging open vast areas of empirical research as well as theoretical thinking. These are relevant in any field, including bioethics.

The absence of nudge theory in IC

A successful implementation of nudge theory in healthcare could be of awesome importance: many major health problems—surely in the developed world—are (partly) the result of modifiable unhealthy behaviors (Danaei et al. 2009), which can be targets for nudging. Nudge theory has indeed been applied to public health (Lowenstein et al. 2007; Halpern et al. 2007; Brownell et al. 2010; Ménard 2010; Wellesley 2011; Blumenthal-Barby and Burroughs 2012); however, its implications for the ethical theory of IC have not been articulated. Not only is this expansion natural, as nudge theory concerns influencing choice in every context, but one may arguably detect an anomaly in that the theory has been developed in the context of policy when it focuses essentially on individual choice.

The transition from public policy to IC cannot be automatic, however. There are both general and specific reasons for this. The former include the following considerations. (1) The meaning of paternalism is quite dissimilar in public policy or legislation and in interpersonal relations. Whether a law or official measure (e.g. regarding consent to organ donation) can be interpreted as legitimately expressing the presumed consent of citizens is a different ethical question from whether an act of persuasion crosses the line of paternalism in the context of
interpersonal communication. Remember also that public choices are not simply aggregations of individual choices but are determined by collective factors too. (2) Justifications of paternalism are different in the two realms. Opposition to even soft paternalism may be justified when the government is the nudger (Thaler and Sunstein 2008, 10; Hausman and Welch 2010), but not necessarily in the different context of a caring relationship between doctor and patient. Conversely, one may view policy nudges as acceptable—say when they address society at large and are designed to benefit all, as in campaigns to reduce littering—but be resentful toward any form of paternalism in face-to-face interaction. (3) Public policies that nudge are often open to criticism on grounds of injustice (Ménard 2010)—of unfairly targeting specific populations or putting certain populations at a disadvantage (for example, the need to indicate in writing one’s rejection of the default option with regard to organ donation discriminates against the illiterate) (Jacob 2006). Normally, such considerations of injustice are not directly relevant to nudging in IC. (4) Various other moral considerations take on different dimensions in public versus individual contexts. For example, utility maximization is often right in public policy, and so measures whose average benefits significantly outweigh their risks may be legitimately promoted. In the IC context, in contrast, where the one-time decision of an individual is concerned, maximin reasoning may make better sense, so that despite greater overall benefits, the risks become the weightier choice-determining factor. This too can affect the moral restrictions on nudging.

The demand for special ethical assessment of nudging in IC is also due to reasons specific to that setting. (1) While in the context of policy the choice architect is the one initiating intervention, in IC he is invited to intervene—prima facie, these different scenarios entail different limits of permissibility. (2) Flipping the default option is arguably the most popular
nudging technique in the context of policy (Halpern et al. 2007). In IC it is often irrelevant, however, since in virtually any case of IC for corporeal intervention, if we could change the default option to an “opt out” one, then we did not need IC in the first place. (3) In public health policy, nudges aim to change people’s lifestyles and health-related habits. A prevalent nudging technique is therefore to incentivize desired behaviors (Blumenthal-Barby and Burroughs 2012). However, offering incentives to the patient so that she chooses in a specific way will normally show lack of respect for her free judgment, undermining the very raison d’être of IC. (4) In public policy, unlike IC, nudging normally attempts to change long-term behavior. This is often doomed to failure, however, due to the long-term effects of counterbalancing influences (Goodwin 2012), and cannot therefore be justified; this consideration does not apply in the mostly ad hoc scenarios of IC.

The above list is obviously far from exhaustive but enough to show that the ethical assessment of nudging in the context of IC deserves special consideration.

What nudging can contribute to IC

IC is the main vehicle expressing the principle of respect for patient autonomy. The explication of that autonomy is debatable, however. With the advent of the idea of patient autonomy, it was (and often still is) pitted dichotomously against the opposite paradigm of paternalism. The dichotomy is problematic since it forces us to choose one of the two important values, beneficence or respect for autonomy (usually the latter), at the expense of the other. The introduction of the idea of “shared decision-making” added a promising third option. Yet this has often been interpreted so as to devolve back into the above dichotomy: the patient brings knowledge of his values, while the physician’s role in this “sharing” is the trivial one of offering
technical knowledge (Eddy 1990). Aside from many other problems (presenting a radically meager idea of medical professionalism; divesting doctor-patient relationship from any depth or even meaning; unrealistically assuming that patients have clear ideas regarding their medical values and what they entail in concrete medical dilemmas; and assuming that a clear-cut distinction between facts and values is possible [in general, and specifically] with regard to medical decisions), such a scheme is in fact no different from the one-sided reliance on patient autonomy in determining treatment. Others presented more balanced and nuanced conceptions of the idea of shared decision-making in IC, where values are discussed and clarified in the doctor-patient interaction, and adapted to the concrete medical situation (Katz 1984; Appelbaum et al. 1987). Emanuel and Emanuel, adding further precision, raised to four the types of IC on the autonomy-paternalism continuum (Emmanuel and Emmanuel, 1992). Alongside the paternalistic model and the unobstructed-autonomy (“informative”) model, they added the “interpretive” and “deliberative” models. In the former, the doctor as counselor helps elucidate relevant patient values to promote patient self-understanding, while in the latter the doctor is more like a teacher who engages the patient in moral deliberation and does not shy from persuasion regarding the most admirable health-related values relevant to the situation. Despite the helpful flexibility that these different models offer to the practice of IC, this proliferation cannot eliminate the fundamental (potential) tension between the doctor’s duty to respect patient choice and to provide optimal care. Ultimately, nothing prevents the reappearance of the dichotomy, forcing a dilemma: if in the deliberative model the patient refuses to listen to the persuasion attempts or reflect on them, the doctor is forced back into the dilemma of either respecting the patient’s unhealthy choice or trying to impose her judgment paternalistically. And if in the interpretative model the doctor provides merely careful, narrow interpretations of the patient’s preferences, she
then merely comments on the patient’s values; but if, on the other hand, she takes freedom to provide more creative interpretations, she will already be imposing her values.

**Nudge theory offers a new paradigm, purporting to overcome this intractable dilemma.**

Once the patient has expressed a clear choice, the dilemma may be ultimately unavoidable; but if the doctor can prevent the very formation of an unhealthy preference—never through breaching the requirements of proper disclosure, only through choosing its modes—then it can be evaded. An example would be the doctor who secures the patient’s consent to an important intervention by presenting its success rates as opposed to the complementary failure rates, that is by choosing one of two equally precise modes of information disclosure (McNeil et al. 1982). Introducing nudge consideration into IC presents a new (fifth) model, where the task is not anymore that of weighing autonomy against beneficence ethically but rather of finding a true synthesis between them. The question becomes subtler: not *when*, if ever, to impose the doctor’s view of best healthcare but *how* to interfere so as to preempt the gap between patient preference and medical recommendation. As we will see, such interference (choice architecture) is in fact frequently inescapable.

**The challenge**

The idea that nudging is an autonomy-preserving intervention faces an immediate and deep challenge. The criticism is that nudging takes advantage of people’s flawed methods of reasoning or decision heuristics (as indeed the example above suggests); it is thus manipulative rather than respectful toward their autonomy. Autonomy is preserved either by non-interference, or, when there is good reason, interference must be through rational persuasion alone. In contrast, “exploitation of imperfections in human judgment and decision-making...[is] prima facie as
threatening to liberty, broadly understood, as overt coercion” (Hausman and Welch 2010, 130). In fact, it is argued, overt coercion may be better than fostering irrationality, as it leaves more decision capacity in the hands of the chooser, albeit from a restricted choice set. (This is parallel to a well-known argument that deception may be less respectful than violence.) Nudging is hence allegedly autonomy-thwarting: since it influences people by illegitimately perpetuating bad reasoning, it ends up supplanting the patient’s will with that of the doctor.

Universally, IC requires patient competence. Once this is satisfied, however, IC theory does not commonly condition the duty of respect on the relative validity of the chooser’s reasoning methods. While the general question this observation raises is well beyond our scope, the challenge posed by the above criticism is of real relevance to this examination. In response, the following discussion will first present cases where nudging does not encourage irrationality and where therefore the force of the criticism is neutralized; subsequently, it will offer arguments for nudging even when the criticism is not essentially neutralized. These two lines of reasoning will provide strong ethical justification for incorporating nudging in IC.

**Nudging without irrationality**

**Nudges often do not elicit irrational decision-making.** In such cases the criticism fails to apply. Below are examples.

**Adding information on the caloric content of the food items in the menu nudges people to consume leaner, healthier foods.** Providing relevant information is surely not a threat to rationality—quite the contrary. Such information is considered a “nudge” only because it is not normally provided and is introduced specifically to guide behavior. The doctor too may nudge in a similar way. It was found (Lipkus et al. 2005) that women’s concern about contracting breast
cancer increases when they learn about their comparative risk (as distinguished from their absolute risk). Greater concern, in turn, increases consent to medical evaluation. If, accordingly, in cases where absolute risk clearly mandates screening, the doctor nudges simply by adding information: instead of strictly communicating to the woman her lifetime risk of breast cancer, adding a comment on the lower risk of the average woman, then the nudge can hardly be considered to promote irrationality.

Offering information can be a choice-steering nudge also in a different way. There is rarely an objective measure to the appropriate amount of information the doctor should provide; there is rather a spectrum of reasonableness. Now, within that spectrum, the doctor may legitimately offer the larger amount of relevant information, anticipating it may test the particular patient’s patience to listen. A typical reaction by the patient could then be to forgo his right to hear more and request the first option presented, provided it is sufficiently good. By explicitly declining further information he allows the doctor to not even present those options she thinks are better not chosen. Now this may seem to justify the criticism of promoting irrationality, but it does not. In contradistinction to Rational Choice Theory’s assumption of optimal knowledge, decisions in life are normally taken under conditions where it is not cost-beneficial to attempt to acquire all relevant information. It is pragmatically-rational under such conditions to engage in “satisficing,” i.e. choosing the first option that crosses some threshold of acceptability (Simon 1956). The doctor thus nudges the patient into a perfectly common (and theoretically-endorsed) form of rational decision-making, while securing consent to the option she believes is the patient’s best interest. Decisions arrived at rationally and without coercion are not autonomy-thwarting.
Another example: By appearing trustworthy and projecting optimism the doctor can induce optimism in the patient. This may be the nudge necessary to secure consent to treatment. Now optimism deviates by definition from realistic assessment, yet is sometimes indispensable for initiating important advancement that can come only through a self-fulfilling prophecy. One needs hardly belabor the point of how important optimism can be for the sick specifically. Since one cannot consciously decide to be optimistic, it is sometimes rational to endorse in retrospective reflection the doctor’s nudge as the inevitable ironic way to achieve a therapeutic result which requires an optimistic attitude but that the patient cannot self-induce. In this example too nudging does not induce irrationality; it rather involves a valid alternative rationale (that of irony). Its rationality is shown by its being the only way for the patient to achieve his desired purpose.

Maybe the most straightforward nudge that involves no irrationality is that of “active decision” (or “active choice”). The intervention simply consists of specifying a deadline by which the patient is required to report her preference. Since “active decision” forces a decision-maker to think about a problem, it is a type of paternalism; but since it leaves all choices equally open, it qualifies as LP (Carroll et al. 2005). “Active decision” was shown to prevent formidable losses in saving rates (ibid.), but it can have even more dramatic effects in preventing medical catastrophes secondary to patients’ procrastination of decision-making regarding treatment. Moreover, when patients hesitate but are not allowed to procrastinate, they are more likely to seek and adopt the physician’s advice. In this way too the nudge of “active decision” can end up helping patients without restricting their choice set and without stirring irrationality.
To the extent that nudge-influenced decision-making is rational—in whatever sense—the criticism that nudging exploits cognitive flaws does not hold and cannot support opposition to nudging on grounds of respect for autonomy.

Nudging on the balance

Other methods of deliberation are indeed epistemically or rationally flawed. If success in nudging depends on taking advantage of such methods, then the criticism against nudging is coherent. This, however, is not sufficient to make it morally convincing.

In nudging scenarios the doctor relies on deliberative capacities ubiquitous among people. “Respect for autonomy” cannot automatically receive the radical interpretation of a duty to communicate only in ways that secure optimal patient reasoning. Assuming responsibility over improving patient’s self-consciousness regarding their thinking practices may be one worry too many. Remember that usually the doctor does not initiate the interaction, and so the duty involved may unreasonably amount to the stricter one of non-cooperation with a person’s suboptimal reasoning; and this is already well down the slippery slope toward de facto treating almost everybody as incompetent for decision-making. How excessive all this is can also be appreciated once we look beyond the theoretical bubble of medical ethics and remember that in most social transactions consent need not even be informed to be valid (Miller and Wertheimer 2010). It is wholly appropriate that in bioethics we expect higher standards, but these can be nicely accommodated by, for example, a theory like Onora O’Neill’s, where the function of IC is to prevent patients from being coerced or deceived (O’Neill 2002). In the standard examples of nudging, patients are obviously not coerced but surely not deceived either, even if faulty reasoning is relied upon. (Recall again the example of steering preference by framing outcome
statistics in terms of success instead of failure, which—far from deceiving—conveys accurate, non-partial information.) Moreover, attempting to improve or correct the deliberation processes of competent patients can easily be seen as overprotective and unacceptably paternalistic as such.

A stark example is provided by “the Gettier problem of informed consent,” where the attempt to ensure the highest standards of autonomy demands verifying the reliance of consent on the information provided, which *ipso facto* imposes a new form of paternalistic intrusiveness (Cohen 2011). If we accept this general line of reasoning, then nudging can be wholly morally legitimate even where it does take advantage of flawed deliberation practices.

Alternatively, we may hold that responsibility over patients’ reasoning modes indeed falls within the purview of the duty to respect autonomy, but that nudging is nonetheless supported by powerful arguments. The first is that even granting such responsibility, it is still reasonable to hold that it is relatively weak, so that if the doctor must choose between improving the reasoning or the health of patients, it is certainly (some would say: trivially) his duty *qua* doctor to choose the latter. This may not always be true, discretion is indispensable when balancing conflicting values, but such discretion seems to fall squarely within the fiduciary role of the doctor (its agentic function in particular (Joffe and Truog 2010)). By default, at any event, nudging is ethical.

There is an important second argument in favor of nudging even where it takes advantage of faulty reasoning. The criticism against nudging is based on the view that autonomy has priority over beneficence. Now we can say that the idea of respect for autonomy understood in terms of negative liberty militates against external interference with people’s decisions and rather accepts them as they are, whatever heuristics they involve (and so nudging is allowed); the idea of respect for autonomy understood in terms of positive liberty, on the other hand, may condone
interference to improve deliberative rationality (and so nudging may become inapplicable). Granting for argument’s sake the lexical priority of the duty of respect for autonomy over beneficence, it is nonetheless reasonable to claim that when we must choose between two different conceptions of autonomy to respect, the moral requirement is to choose the one more compatible with beneficence. We may call this “the principle of weak lexical priority” as applied to respect for autonomy. Its general form is: When there are two (or more) acceptable options for the prioritized norm, choose the one more compatible with the lower-ranking norm(s)! I believe this missing principle offers an important tool for discussions of autonomy in bioethics.

The thrust of the above arguments is reinforced by one further consideration. While a conclusion reached through flawed deliberation is epistemically inferior to one reached through sound deliberation, once it generates a preference, that preference is not similarly inferior. The reason is that any “superior” preference—the result of better deliberation—is a counterfactual that is not in fact the subject’s preference. While normative assessment of modes of reasoning can be done in the abstract, the point of respect for individual choice is that it ordinarily refers to people’s actual preferences. Virtual preferences can give rise to “hypothetical consent,” whose virtuality precludes its being considered true consent. (Hypothetical consent can serve as a helpful default when a patient cannot give consent, as when comatose, but not when a competent patient in fact expresses a different preference.) IC expressing badly-generated preferences can hence still be fully valid, if only the generation of the preference passes a threshold of acceptability. (This may be taken to express—in allusion to Kant—a special sense of “the primacy of the practical over the theoretical.”) Accordingly, when nudging that exploits flawed deliberation is justified by whichever reason (as discussed above and below), the IC it generates is as good as any. Non-ideal reasoning can yield fully legitimate preferences.
Nudging and ethical manipulation

If nudging does not exploit flaws of rationality, or if it does but this is none of the doctor’s business, or if, alternatively, while it is the doctor’s business, he has overriding ethical reasons to nudge, then nudging is not ethically wrong; indeed, since it promotes well-being, it is ethically good. However, despite that multi-layered defense, in the important case where the justification for nudging is the priority of the duty to heal over protecting the quality of patient reasoning, the full defense of nudging is in fact much more intricate. We must determine the extension or scope of the pro-nudging default. In any case of balancing two opposing valid claims, careful weighing is mandatory. In our case this involves judging the relative permissibility of the manipulation involved: while soft manipulations tend to be unproblematic, more blatant forms can be wrong.

Manipulation is classified between persuasion and coercion (Faden and Beauchamp 1986, chapter 10). While the former is legitimate, the latter is not, and so manipulation, falling between those two, is often understood to be half-way coercive, half-way wrong. As a recent paper on nudging contends, “manipulation bypasses the exercise of autonomy” (Blumenthal-Barby and Burroughs 2012, 5); it is thus supposedly problematic in principle and in need of special justification whenever legitimate. However, manipulation’s midway position should not be taken to mean that as such it is always moderately blameworthy, but rather that it sometimes is and sometimes is not. (A moment’s reflection reveals that there is virtually no substantive communication without manipulation in some trivial senses; if only because there is no objective shape to most illocutionary acts and any choice of presentation has some purpose.) This fact applies in particular to nudging (when it involves flaws in rationality). Accordingly, this section
will present the main criteria for evaluating the relative permissibility of nudging from the point of view of the ethical assessment of manipulation.

(1) The mode of communication is the first criterion. A common view, as mentioned, is that influencing choice through any means other than rational persuasion undermines autonomy and is therefore wrong manipulation. This is an extreme position; the idea that people express their “true will” only through submitting every idea or proposal to rational deliberation is hard to defend. Generally speaking, good, sound judgment involves many elements that transcend rationality. It is often impossible to judge without assistance from the sentiments (Pizzaro 2000); in other instances, judgments cannot be made without some recourse to non-rational intuition—at times it is even logically impossible (trust, for instance, is strictly-speaking rational only if based on trustworthiness, but there are good reasons to believe that strict adherence to this condition would eliminate the possibility of trust (Becker 1996)); many of our descriptive concepts, and hence the judgments based on them, include a built-in value element (“thick concepts”), which is often partially non-rational (Williams 1985); visual and other sensual representations shape our judgments in mundane ways that transcend reason-giving but are considered legitimate everywhere; and so on. If all these are overwhelmingly normative, yet all transcend “rational deliberation” in the strict sense and are therefore manipulative (per definition) when used to influence, then surely manipulation is often normative. Importantly, this applies to the nudge too. Some modes of influencing that may fall under the formal definition of the nudge are indeed wrong (e.g. subliminal advertising), others are not—careful evaluation, not wholesale rejection, is required.

The criticism against the nudge’s exploitation of flawed reasoning echoes most closely the biblical injunction against “placing a stumbling block before the blind” (Leviticus 19, 14);
this is probably the core of its intuitive appeal. But (beyond the significant fact that that
injunction refers to harming, while the nudge is beneficent) as explained, what is commonly
considered normative judgment is much more permissive than the facile equation of less-than-
fully-rational with wrongly-manipulative suggests. Hence, the question of when manipulation—
and therefore the nudge—is in itself problematic and when not could only receive definitive
answer under the umbrella of a (missing) comprehensive ethical theory of communication. First
steps in this direction were taken by Manson and O’Neill (2007); here I can but articulate
relevant parameters for such a theory.

(2) In their criticism of nudging, Hausman and Welch (2010, 133) write: “Even when
unshaped choices would have been just as strongly influenced by deliberative flaws, calculated
shaping of choices still imposes the will of one agent on another.” In response, we must
recognize a continuum of ethical relevance with regard to this “imposition.” We should ask: Is it
reasonable to hold that one chooses without coercion and that (ex hypothesi) the quality of one’s
deliberation remains unchanged by outside intervention, but that the situation is morally
problematic just because the environment one operates in was shaped by another? Every social
environment was shaped by others. One important criterion is therefore the extent to which a
nudge exerts its influence on the environmental circumstances of choice as distinct from the
chooser itself. This criterion may be understood as the relative purity of the choice architecture.
The purer it is, i.e. the more indirect is its influence on the act of choice, the further it will drift
from our intuitive idea of paternalism and the more innocent the manipulation will be.

(3) This uncoupling of influence and effect is closely related to another, neglected,
parameter, which we may call “disproportionate impact” in human communication. When a
system displays sensitive dependence on initial conditions, then very small differences in those
conditions can produce significant changes in outcomes; in other contexts, such dependence is best known as “the butterfly effect.” Now a system of a nudge and its consequences (the actual choice) can exhibit “disproportionate impact” to a higher or lower degree. The wider the gap of magnitude between intervention and effect—when very subtle changes in behavior, in framing or in the setting of the circumstances of choice can leverage differences in choice—the less ethically problematic the nudge, or any manipulation, becomes. (Evaluation of manipulations according to the proportion between magnitudes of intervention and impact and its meaning for IC deserve scrutiny; these important topics remain beyond our scope, however.) It is important to note that extant theories of autonomy often do not discuss intersubjective considerations and certainly are too thin to account for parameters such as the level of indirectness of intervention or of the proportionality between intervention and impact. Again, judgment of nudging that incorporates such considerations could be carried out substantially only by articulating an ethics of human conversation as a foundation for the ethics of autonomy and of IC.

The critic of the nudge may argue that the problem is not in the mere fact that the chooser’s environment was constructed by another, but specifically that it was shaped in view of the event of his choice. However, to the extent that the intervention is negligible, the question of its intentionality becomes moot. Otherwise, anyone’s mere thoughts could constitute another’s unfreedom, which is absurd (even when we call that thought-bearer “choice architect”).

(4) The previous criterion of the relative negligibility of the act of influence is complemented by another: the extent to which an intervention actually changes choosers’ preferences. If on top of being non-coercive in nature (per definition), a nudge also does not in fact change the patient’s preference, then the intervention is non-intrusive both in its nature and its consequences and the grounds for the criticism from autonomy diminish significantly—
indeed, they may become insubstantial. Now the deepest insight of behavioral economics is that people’s preferences are often inchoate to the point that they do not really exist before being presented with a specific context of choice (Slovic 1995). A simple illustration is the phenomenon of “loss aversion,” where the negative value people assign to a given loss is larger in absolute terms than the positive value they assign to the identical gain—a clear deviation from the basic idea of preference underlying the rationality of the theory of expected utility. “What this means is that people do not assign specific values to objects,” as Thaler and Sunstein (2008, 33) succinctly put it. Under these conditions Rational Choice Theory cannot get off the ground.

Now to the extent that people do not have preferences outside given contexts of choice, the doctor can intervene in a beneficent manner without changing—let alone overriding—patients’ existing preferences. Hence, in limit cases, influencing (manipulating) people’s preferences is arguably blameless.

Moreover, the clarity and definitiveness with which people hold specific preferences vary. Default options and other nudges exert stronger effects on choices about which people lack strong preferences to guide their decisions (Slovic 1995). In addition, there is evidence that framing effects are less likely in situations with higher perceived ethicality (Bateman et al. 2002). Importantly, these findings indicate that nudging is more possible (successful) the more it is justified, and this offers extra assurance to the legitimacy of exercising the nudge.

(5) Not only need nudging not damage autonomy but it can enhance it. For example, hanging mirrors around eating areas nudges people to eat more responsibly. To the extent that this prevents akratic behavior and so resonates with people’s wills, it can reasonably be viewed as autonomy-enhancing. In medicine, nudging can similarly use reminders which assist people in making choices compatible with their own goals. May and colleagues (2010) demonstrated that
showing to smokers video simulations of them suffering a myocardial infarction significantly helped smoking cessation. If a similar nudge is used, say, in the context of acquiring IC for enrollment in a smoking cessation program, then such “virtual reminders” would be similarly autonomy-enhancing. Other nudges could remove other obstacles to autonomous judgment. As is well-known, medical jargon can provoke anxiety in patients (Taylor et al. 2011). Irrational anxiety tends to cloud good judgment. If therefore the doctor interested in securing IC frames information so as to convey the same message while avoiding certain medical terms, and if this prevents anxiety in the patient who subsequently gives consent, then the successful nudge is autonomy-enhancing. (This must be distinguished from not using medical terms that the patient cannot understand, which is mandated by the rules of adequate disclosure and does not count as a nudge.) The more a manipulation is autonomy-enhancing, the less vulnerable it is to the criticism from autonomy.

Patients, like everybody else, exhibit a wide range of decision-making biases (Redelmeier et al. 1993). When such biases are successfully offset by a nudge, then again, nudging can be autonomy-enhancing—indeed, it even has an obvious rationality-promoting effect. A couple of examples will illustrate this. One of many ways in which memories have been shown to be distorted is “duration neglect,” where the duration of a past experience tends to be not as well represented as its peak intensity (Fredrickson and Kahneman 1993). Now imagine that two good treatments, A and B, are available to a given patient, but that A is overall better. Although A is much shorter, it involves greater discomfort, and so the patient is called upon to choose between the treatments. She indeed chooses A due to its much smaller potential to disrupt normal life. If she confirms her preference at the end of treatment but upon returning for a second treatment is suddenly indecisive between A and B, the doctor may well suspect an instance of the “duration
neglect” bias and nudge her by, for example, offering A as default. If this counteracts the patient’s judgment bias, then to that extent it enhances autonomy. (Notice that even if it didn’t counteract judgment bias and enhanced autonomy, it would still be a useful nudge, since A is the best treatment.) Another example: “Preference for Status Quo” is a biased decision-making heuristic; it can cause patients to stick irrationally with current treatment, even when clearly superior treatment becomes available. If the doctor properly frames the information regarding old versus new treatments in a way that nudges in favor of the new, and if the new is therefore indeed chosen, the nudge is likely to have an autonomy-enhancing effect.

Paternalism often aims to promote people’s “true will” by knowing better than them what they really want. There is also a well-known paradox in using coercion to help people achieve self-government. But autonomy-enhancing nudging does not fall into those traps: it is non-coercive (per definition) and we shall not refer to it as autonomy-enhancing unless it uses its suggestive power to steer people only toward self-proclaimed goals. Indeed, even authors hostile to nudging admit that it can be used “to facilitate autonomous decision making by freeing individuals from other irrelevant influences” (Hausman and Welch 2010, 134). The nudge’s capacity to enhance autonomy is similarly recognized in its designation as “asymmetric paternalism” (Camerer et al. 2003): the asymmetry is manifested in that nudging measures do not restrict the freedom of those who already behave with optimal rationality—and hence autonomously in that respect—while they can rectify the conduct precisely of those who do not so behave.

(6) Finally, any ethical criticism against nudging becomes pointless the more choice architecture (choosing some influencing strategy) is inescapable. For example: people are more likely to consent when the costs involved—pain, for instance—are not incurred at the time of
decision (O’Donoghue and Rabin 1999). This phenomenon can be successfully exploited by the
doctor who seeks an IC not immediately before a painful procedure but on the previous visit.
Since it is inevitable to obtain the IC at some time, the doctor cannot be criticized for practicing
choice architecture. Another example: when presented with a list of options, people cling to the
last option with higher probabilities (compared to, say, the third before last). If the doctor uses
this rule to improve chances for IC, then criticism is beside the point, since he cannot avoid
presenting choices in some order. Yet another example: upon arrival to his office every morning
the surgeon chooses between wearing scrubs or a white coat. If he knows that being in scrubs
(say) improves chances of consent (supposedly by increasing patients’ trust), then he can hardly
be criticized for his choice architecture, since surely he has to wear something.

It is sometimes claimed that there is no value-neutral presentation of facts (Kirklin 2007).
Be this as it may, notice that the point about inescapability illustrates the difference between
nudging in the context of public policy versus in IC: as mentioned above, while implementing
specific policies is normally optional, fashioning IC in some manner is truly inescapable once the
doctor’s help has been sought.

To the extent that choice architecture is inescapable, and therefore blameless, nudging is
in fact morally required. This is not only because it promotes the duty of beneficence, but
because since no neutral option is present, if we do not choose the beneficent intervention, we
ipso facto end up maleficent (nudging then provokes the “nocebo effect of informed consent”
(Cohen 2012)).

This review of criteria is surely partial, but it shows that not all manipulations are
ethically equal. If a manipulation fulfills all these criteria: it uses modes of communication
generally considered normative, its interference is flimsy in nature, it does not act on people but
on environments, it does not change pre-existing preferences or it supports action compatible
with such preferences, and if it is inescapable, then moral denunciation becomes hard to justify
and is arguably out of touch with the reality of human communication.

Our examination of parameters of manipulation originated from the need to balance the
doctor’s duty not to manipulate cognitive imperfections and the duty to improve patient well-
being. The more the above criteria hold, the more secure is the default in favor of nudging to
improve well-being.

**Conclusion**

By means of the nudge, LP aspires to influence patients’ choices in health-promoting ways while
respecting their autonomy. The criticisms leveled against nudging are often too sweeping and
overlook important distinctions. A recent paper, possibly the first to deal specifically with
nudging in IC, claims that “libertarian paternalism entails a weakening of the standard
requirements of informed consent” (Ploug, et al. 2012). Here I attempted to show that in some
instances such criticism against nudging simply fails to apply and that in many others it is
unconvincing—this latter is strongly dependent on a an analysis of the spectrum of legitimacy of
manipulation within an ethics of human communication. Since nudging is welfare-promoting, if
it is otherwise ethically legitimate, then it is ethically good, combining the advantages of
beneficence and respect for autonomy.

That the nudge is an important tool does not mean it is always the best. Shared decision-
making, in its best forms, is a good model for IC. When it is likely to promote fruitful discussion
and yield new insights in the patient, it is probably the best practice. This will happen when
conditions of trust, openness, and rational thinking prevail; then, shared decision-making is

URL: http://mc.manuscriptcentral.com/uajb
likely to assist patient reflection on her preferences and enhance autonomy. At other times, as when the patient is unreasonably stubborn or when a therapeutic relationship is difficult to establish, nudging may remain as the best option for IC. When the patient’s unhealthy choice is mainly the result of unexamined ideas or lifestyle, then mutual deliberation may be the preferred model; when it is due to akrasia, then nudging may work best.

The wise implementation of nudging can and should be part of the art of medicine. As it is with art, expecting a manual of how and when to nudge is unrealistic. Lists of types of nudges do exist (Thaler and Sunstein 2008; Blumenthal-Barby and Burroughs 2012), but the real challenge is to develop proficiency in applying the right kind of nudge in the appropriate circumstances. For instance, offering a default option may be unproductive for a patient who procrastinates decision-making; in such a case, “active decision” is warranted. On the other hand, default options are generally better at preserving the autonomy of patients with higher medical literacy, and fixed defaults are similarly better the more group preferences are homogeneous. Such rules-of-thumb have nowhere been richly elaborated; systematic thinking on this matter is clearly needed.

The judicious use of nudging could also benefit from research on attitudes toward nudging in the population in general and among patients in particular. Might the use of nudging damage trust between physician and patient? Relevant research will be welcome, although a number of reasons suggest this fear is unwarranted. Due to their gentle nature, nudges tend to blend imperceptibly in the normal subtleties of human communication (a fortiori in the oft-tortured communication between doctor and patient). Moreover, when nudges target irrational heuristics, they are imperceptible almost by definition, since ceteris paribus if people were aware that their heuristic is not rational, they would not use it. Yet even if detectable, it should be
reemphasized how benign the milder, gentler forms of manipulation often are: everyone, including the most decent and truthful, sometimes choose certain words rather than others or attenuate modes of expression in various ways to advance conversational purposes. Not only is this not deception, but it is practically inevitable if speech is not to undermine basic politeness, privacy, human bonds, or public peace too often and in myriad ways; and not only is this not a setback to human interaction, but it is unwise to never so act. Since the average person arguably comprehends all of this, the gentler, harmless kinds of manipulation seem to be widely considered normative. It would therefore be rare for a patient to react adversely to a gentle nudge, all the more so when it is prudently practiced in the course of a caring relationship.

The field of decision-making biases and their status in human communication ought to be applied to IC. This paper offered but an initial presentation of a needed discussion. In 2009, the American National Institute of Health prioritized research to advance the Science of Behavior Change (https://commonfund.nih.gov/behaviorchange/). This serious approach should also be applied to the theory and practice of IC.
REFERENCES


URL: http://mc.manuscriptcentral.com/uajb


Ploug, T. et al. 2012. To nudge or not to nudge: Cancer screening programmes and the limits of libertarian paternalism. *Journal of Epidemiology and Community Health* doi:10.1136/jech-2012-201194


URL: http://mc.manuscriptcentral.com/uajb