

notizie di

# POLITEIA

RIVISTA DI ETICA E SCELTE PUBBLICHE

Anno XXIV - N. 90 - 2008

## Contents

### UTILITARIANISM: HISTORICAL THEORIES AND CONTEMPORARY DEBATES

Edited by GIANFRANCO PELLEGRINO

- 3 Introduction, by GIANFRANCO PELLEGRINO
- 7 John Stuart Mill's Science of Politics,  
by FREDERICK ROSEN
- 20 Hume, Experience and Value,  
by PETER KAIL
- 31 Utilitarianism and Its 19th-Century Critics,  
by SERGIO CREMASCHI
- 50 Utilitarianism and Contract Theory: Democracy  
in Bentham and Rousseau,  
by ANNAMARIA LOCHE
- 64 Utilitarian Arrogance,  
by JOHN SKORUPSKI
- 75 Rule-consequentialism versus  
Act-consequentialism,  
by BRAD HOOKER
- 86 Utilitarianism as Both Unattainable Ideal and  
Human Morality,  
by TIM MULGAN

Proceedings  
from a  
conference  
on  
Utilitarianism  
in  
Rome, Italy

#### Comitato direttivo

Emilio D'Orazio, Francesco Forte,  
Gian Cesare Romagnoli, Salvatore Veca

#### Comitato di redazione

Carla Bagnoli, Alessandro Balestrino,  
Ian Carter, Daniela Giannetti,  
Maurizio Mori, Luca Parisoli,  
Nicola Pasini, Mario Ricciardi

#### Direzione, redazione, amministrazione e pubblicità

POLITEIA - Via Cosimo del Fante, 13 -  
20122 Milano  
Tel. (02) 583.139.88 - Fax (02) 583.140.72  
E-mail: politeia@fildir.unimi.it  
http://www.politeia-centrostudi.org

#### Abbonamenti/Subscriptions 2008

Un fascicolo € 12,50

Abbonamento annuo:

- Privati € 40,00

- Enti € 50,00

- Sostenitori € 75,00

Arretrati € 12,50

L'abbonamento può essere sottoscritto  
inviando a POLITEIA - Milano  
(specificando la causale)

- assegno bancario non trasferibile

- vaglia postale (Milano - succ. 84)

- bonifico bancario: Intesa San Paolo,  
Ag. 32, Milano: IBAN

IT21 W030 6901 6271 0000 0010 460

intestati a POLITEIA

For subscriptions outside Italy see  
information on final page

#### Editore

Tipolito Subalpina s.r.l.

Via Genova, 57

10090 Cascine Vica - Rivoli

Direttore responsabile:

Emilio D'Orazio

Registrazione presso il Tribunale

di Milano n. 358 del 13.7.1985

Iscrizione R.O.C. n. 4147 del 21.11.2001

Videoimpaginazione e stampa:

AGITsubalpina - Beinasco (TO)

- 98 Some Old Objections to Rule-Consequentialism  
Reconsidered: The Case of Reproductive Freedom,  
by GIANFRANCO PELLEGRINO
- 108 Utilitarianism and Coercion,  
by NIR EYAL
- 124 NOTES FOR CONTRIBUTORS
-

# Utilitarianism and Coercion

NIR EYAL\*

## *1. Commonsense opposition to coercion*

According to commonsense morality, coerciveness makes actions harder to justify. The intuition is that, normally, offering people food is easier to justify than forcibly feeding them. Forced feeding is justified only when it is necessary in order to address a medical emergency and the like, if it is justified at all; offering food is acceptable as a means to many additional ends. The intuition is also that accepting a present is easier to justify than stealing; that consensual sex is dramatically easier to justify than rape; that extracting voluntarily donated kidneys for transplantation is far easier to justify than confiscating them coercively. Intuitions hostile to coercion abound. Much of modern moral and political philosophy takes such intuitions seriously, and is founded on a presumption against coercion.

Commonsense does not view coercion as categorically forbidden; a certain level of coercion is justified all things considered when the consequences are important enough. Nor does it view all forms of coercion as equally difficult to justify. There is no single constant "threshold" against coercion of all forms. Commonsense recognizes a large, perhaps an infinite, collection of thresholds, which specify how difficult to justify are determinate forms of coercion (or, if you will, there is a single threshold against coercion, but what is necessary in order to pass it is different for different forms of coercion). It is easier to justify coercion by certain means (threatening a small fine) than by others (threatening a large fine). It is easier to justify coercion against some people (guilty criminals) than against others (law-abiding innocents). It is easier to justify coercion in certain spheres (the distribution of surplus income) than in others (the distribution of surplus kidneys). Sometimes, the relevant commonsense intuitions demand more than just non-coercion. Prior to participation in risky medical research, for example, our intuitions demand particularly informed and voluntary consent, based on true comprehension of relevant medical facts. Prior to a routine blood draw, non-refusal appears sufficient. Commonsense also bears on the appropriate motivations for noncoercion. Refusing to coerce people out of the sense that coercion would breach their rights or their self-ownership or their dignity or their self-sovereignty is intuitively more commendable than refusing to coerce them only in order to maintain legal cover or to avoid conflict.

Like several other parts of commonsense morality, the large set of intuitions

---

\* Harvard Medical School, Harvard University.

surrounding coercion is often used against act-utilitarianism and against act-consequentialism: on the face of it, justifying a coercive act is not simply a matter of ensuring that it has more good consequences than bad consequences. Intuitively, coercive kidney confiscation is problematic even when its good consequences, saving kidney recipients from reliance on dialysis, are very significant, and quite similar to the good consequences of extracting a voluntarily donated kidney. Intuitively, it is inappropriate to calculate the likely impact on net utility before deciding not to rape someone. Rather, the option of raping should normally not cross the mind of a decent person, not even as a potential means to maximizing net utility. There may seem to be reasons against perpetrating these extreme forms of coercion that go beyond the act's potential impact on utility.

To what extent are commonsense intuitions on coercion correct? Suppose the bulk of these intuitions are incorrect, as many act-utilitarians say they are. Should we then "liberate" ourselves from them, to use Peter Unger's term, or may we be justified in keeping them? Could it be justified, or even obligatory, to keep and to foster these intuitions nevertheless?

This article seeks to provide a new answer to these questions. It starts out by criticizing some utilitarian attempts to justify commonsense morality's opposition to coercion. In particular, section II criticizes attempts to justify commonsense intuitions that invoke especially pragmatic and material considerations. Section III criticizes Richard Hare's attempt to patch up holes in these pragmatic utilitarian justifications in defense of commonsense opposition to coercion. Section IV criticizes utilitarian justifications that rely on the contribution of autonomy to the good life. Section V moves to criticize central deontological justifications for commonsense morality's opposition to coercion. The final section concludes that commonsense intuitions in support of thresholds against coercion are probably incorrect; but it also proposes a new utilitarian argument for maintaining and fostering these intuitions. This argument emphasizes the "dignitary harms" that often issue from coercive acts. It says that the most efficient way for an individual to prevent dignitary harms is usually to maintain and foster these intuitions.

Throughout, I focus on an extreme form of coercion: on coercive medical intervention in the bodies of fully competent and innocent decision makers who in non-emergent circumstances clearly refuse that invasive intervention. In this area, I feel, commonsense condemnation for coercion tends to be robust. Most people seem to agree that such coercive action wrongs the patient, whether it purports to benefit her (as in paternalistic medical care) or to benefit others (as in medical research).

## *2. Pragmatic utilitarian justifications*

Some utilitarian attempts to justify commonsense opposition to coercion focus on simple, pragmatic harms that may ensue from coercion. It is said, for example, that a

person usually knows better than others what is good for her, and cares about it more than others do. Therefore, coercion that purports to benefit the person rarely does so. Moreover, agents who take the liberty to coerce someone for her own or for a third party's sake are often tempted to coerce her harmfully without good enough outcomes. Finally, deciding for oneself, without external coercion, often boosts the quality of one's future decision making. Such arguments are usually associated with John Stuart Mill's attempt to ground liberalism in utilitarianism<sup>1</sup>.

In particular, it is said that coercive medical intervention on refusing competent patients and research subjects exposes them to doctors' professional bias and abuse. Doctors do make mistakes, even when they believe that "Doctor knows best". Patients often know their own medical and non-medical goals better, and they typically mind these goals much more than doctors and medical investigators do. When patients do not take care of themselves well enough, treating them as though they did can encourage them to do so. Coercive care often antagonizes patients who then cease cooperating with an assigned medical team or inform other patients who may not seek care for fear of unwanted medical intervention.

Such pragmatic utilitarian considerations are clearly part of the overall case for commonsense opposition to coercion. However, they cannot fully vindicate commonsense thresholds. There will always be some exceptional circumstances in which the bad outcomes that a determinate pragmatic consideration adduces against coercion remain unlikely, and do not transpire. For example, even if people typically know their own interests best and care about them the most, sometimes they do not, and it can be foreseeable in advance that a particular person will not. In the context of coercive medical care, for example, sometimes even relatives and patient advocates confirm that the doctor knows better the medical facts in the case and minds clinical outcomes more, and that no central project or religious belief of the patient counts against the proposed treatment<sup>2</sup>. The fact that patients typically know best and care the most about themselves allows that, in certain known circumstances, typically the doctor knows best and cares the most.

In addition, sometimes, the patient is asleep or unconscious, it is late at night, and a highly discreet doctor can surreptitiously subject the patient to non-detectable intervention that the patient had earlier refused, without anyone realizing it. The patient's and the public's trust and compliance would not be affected. On such occasions, none of the determinate worries driving pragmatic utilitarian justifications seems fully in place<sup>3</sup>. Nevertheless, the intuition is that such circumstances do not justify highly intrusive coercive care: intuitively, whether someone is looking, whether the doctor is discreet and so forth, should not make very much difference. Commonsense remains fairly indifferent to such prospects. Its counsel against the relevant forms of coercion stands even in many of them. Our intuitions against coercion come in the form of thresholds.

Put differently, our intuitions are more "homogenous" than they would have to be if their only justification were pragmatic utilitarian. Intuitive thresholds against coercive forms of conduct do not suddenly become weak when the room is empty,

the patient is asleep, and so forth. Certain circumstances that affect these pragmatic considerations a lot seem, intuitively, to have little if any normative significance.

A similar problem affects attempts to justify other commonsense thresholds in pragmatic utilitarian terms: while *typically*, doing harm is materially more harmful than allowing harm; intending harm is materially more harmful than merely foreseeing it, and so forth, this is not always the case. Pragmatic utilitarian concern about material harm cannot fully justify commonsense morality's fairly homogenous opposition to harming, to intentional harm, and to other act-types. Many ethicists take this to be sufficient reason to reject act-utilitarianism: proof that act-types carry inherent (dis)value that is not simply the (dis)value of irregular consequences.

### 3. *Hare's move*

Utilitarian philosopher Richard Hare tries to defend utilitarianism from this familiar worry. Hare concedes that pragmatic utilitarian considerations do not always count against coercion. But typically, he insists, they do count against it: more often than not, coercion diminishes net utility. Hare then adds that calculating what to do on a case-by-case basis is usually slower and more open to abuse – less utility-maximizing – than acting fast, automatically, from spontaneous reluctance to coerce. Having general reluctance to coerce thus tends to have the best results. An act-utilitarian will therefore seek to internalize general reluctance to coerce. This pragmatic value, Hare concludes, sufficiently justifies our opposition to coercion. It makes that opposition very fruitful. Hare accordingly lists coercion among the forms of conduct that we should educate ourselves to oppose<sup>4</sup>.

Hare's powerful utilitarian defense of commonsense opposition to coercion could have successfully supported developing somewhat broad reluctance to coerce. But it cannot successfully support the carpet reluctance that commonsense opposition to at least some forms of coercion demands. Hare fails to vindicate convincingly what I have called the homogenous nature of these commonsense intuitions.

Let me illustrate. Consider a physician who readily adheres to the relatively simple rule (*Let the patient decide*). Clearly, she acts more in line with commonsense opposition to coercion than a colleague who readily adheres to the somewhat more complex rule (*Let the patient decide, unless everyone except the patient agrees that your treatment plan would best serve the patient's own goals and good, you are recognized as especially caring, you are discreet, it's late at night, and nobody is looking*). Commonsense demands homogenous, not chequered opposition to coercion. It commends adherence to the simple rule, not to the relatively complex one<sup>5</sup>. However, adherence to the relatively complex rule would benefit many patients in need of medical intervention, and it would *not* do great harm to medical results, to public trust in doctors, and so forth: the "*unless...*" clause outlines circumstances where harm to medical results, to public trust etc. is unlikely. The

pragmatic utilitarian value of adherence to the somewhat complex rule is thus higher than that of adherence to the simple rule.

Hare pre-empts a similar worry. He says that the relatively complex rule is so complex that it is costly or impossible to memorize and instill consistently in one's practices, and attempting to follow it usually fails to maximize utility. Utilitarianism orders us to seek, instead, to adhere to the simple rule, which is easier to translate effectively into action<sup>6</sup>.

Hare's response is elegant, but ultimately unsuccessful. Although the relatively complex rule is slightly more complex than the simple rule, applying either is already very complex. Therefore, the difference between them is relatively small: small in comparison with the absolute complexity of acting on either. To demonstrate the absolute complexity of acting on either rule, consider that, to apply commonsense opposition to coercion, one must be able to tell when an act constitutes coercion. One must master the difference between threatening on the one hand and offering and warning on the other. One must also be able to comprehend enough to translate other intricate notions into action. The definition of coercion is immensely complicated, and the focus of extensive philosophical writing. Furthermore, if my earlier reconstruction of commonsense opposition to coercion is right, one also must know when intervention is highly intrusive and otherwise the sort of intervention coercing which (in particular ways, spheres and so forth) is especially difficult to justify. Stating these criteria in full would take many pages. Evidently, we are able to memorize and act on immensely complex rules. The patina of simplicity in the surface structure of simple-*looking* rules often hides an inordinate amount of information, which we must remember and process in order to adhere to those rules. Usually we cannot spell out that information, but our fairly consistent and coordinated behavior shows that we do. In comparison with this level of complexity, the added complexity of the relatively "complex" rule (a three line "*unless...*" clause) is quite minimal. The many potential benefits of adherence to the "complex" rule are arguably more significant<sup>7</sup>.

A further limitation of Hare's pragmatically-guided defense of utilitarianism is that it may recommend "doing the right thing for the wrong reasons". More accurately, it may recommend doing what intuitively seems right, but not necessarily out of motives that intuitively seem right. Suppose that a doctor heeds Hare's advice and develops the relevant behavioral tendency. She becomes reluctant to impose medical intervention on competent adults and to carry out other coercive forms. She feels guilt or shame when, occasionally, in order to save lives or out of self-interest, she does impose medical intervention, and that helps her maintain her general reluctance to impose it. However, this particular doctor believes that the only reason that she should not impose medical intervention is that doing so might damage future patients' willingness to use the medical system – a pragmatic utilitarian consideration mentioned above. In her view, which she readily shares with patients, the reason to avoid imposing care on them has nothing whatever to do with their own rights over their bodies, their self-sovereignty, and so forth. Patients, she

explains, lack any such rights; the only problem with coercive care is that the public is stupid (as she says) and that, oblivious to the benefits of forced protection, the public will stay away from hospitals if it suspected that care might be imposed there. To keep the public coming in to receive care, she has successfully developed strong carpet reluctance to impose care (to which her past record attests).

The intuition is, I believe, that something about this doctor's approach is ethically astray. The doctor seems to be acting right, but out of a deficient motivation. She does not believe or mention at all that the patient may have a right to choose for their own, to control their body on their own, to exercise a capacity for autonomy, to enjoy the dignified status of a full adult who is not a piece of the doctor's property. Hare's very pragmatic defense of opposition to coercion does not guard against this prospect. It would find no fault with this doctor's approach. The doctor is reluctant to act coercively, and the guilt that she would feel if she acted coercively strengthens her reluctance. She develops these tendencies because coercion usually diminishes net utility in pragmatic terms, say, by eroding the public's cooperation with doctors. From Hare's pragmatic utilitarian standpoint, nothing is amiss.

For these two reasons, even self-imposed beneficial inclinations à la Richard Hare do not enable pragmatic utilitarian considerations to justify commonsense opposition to coercion. Pragmatic utilitarian considerations do not take us beyond adherence to the relatively complex rule, and they do not justify commonsense judgments on suitable background motivations.

#### ***4. Autonomy-based utilitarian justifications***

A different attempt to reconcile commonsense opposition to coercion with utilitarianism emphasizes other, less pragmatic harms that are more inherent to coercion. A possible move is as follows. Autonomy or self-governance is a central component of any good life. Coercion violates autonomy, and in that way it undermines people's good. Unwanted medical treatment, for example, tends in that way to undermine an important objective interest of an adult patient, in leading her life as she sees fit and in promoting her own goals<sup>8</sup>. Thus, the argument goes, coercion diminishes utility even when it purports to enhance it. This argument is also attributable to Mill, and contemporary writers continue to develop it<sup>9</sup>.

Unfortunately, this autonomy-based justification fails to vindicate homogenous thresholds. Autonomy-breaching coercive acts sometimes promote net autonomy, and any utility that autonomy may generate. In the medical context, for example, they may do so by potentially improving a population's clinical outcomes. Imposing medical intervention on a person can improve others' clinical outcomes a lot, say, by enabling transplantation or research to develop medicine for their diseases. It can also improve her own clinical outcomes a lot, say, by enabling medically indicated surgery. I wish to argue that improved clinical outcomes often

maximize net autonomy.

Autonomy means different things for different writers. It may mean having a rich capability set. It may mean having the capacity for rational decision-making. It may mean having the freedom to form, pursue and revise a conception of the good. It may mean independence from others. It may mean freedom from coercion. However, improved clinical outcomes can promote autonomy on any of these conceptions. Obviously, disease and disability can undermine capabilities. Diseases can also undermine mental capacities to decide rationally and to form, pursue and revise plans. They can render a patient physically weak and dependent and a likelier victim of ongoing pressure and duress. Coercion that boosts clinical outcomes can therefore boost net-autonomy.

Insofar as autonomy is an important component of the good life, which utilitarians often seek to maximize, coercion that boosts clinical outcomes can foreseeably boost good, and utilitarians would have to encourage it. It is true that *such* coercion initially consumes a measure of personal autonomy – since it is coercive – but it sometimes generates more autonomy than it consumes. When, overall, coercion maximizes autonomy, and it is beneficial in more pragmatic, material terms as well, utilitarianism recommends coercion hands down. But commonsense morality does not make special exceptions for all such cases. Our intuitions are more homogenous than they would have to be if coercion were recommended on many occasions in which nobody is looking and so forth *and* coercive care would boost clinical outcomes enough to generate more autonomy than it consumes. Hence, coercion cannot be problematic only due to pragmatic and/or autonomy-based utilitarian considerations. There must be more to a successful justification of commonsense thresholds against coercion than these two types of utilitarian consideration.

### ***5. Deontological alternatives***

In their search of a more general justification, some contemporary bioethicists and philosophers turn to deontology. Many suggest that commonsense morality's hostility towards coercion fundamentally rests on our self-ownership, on Kant's formula of universality, or on respect for autonomy (often associated with Kant's formula of humanity). Let me comment briefly on these three deontological alternatives.

First, it is common to cite the value of self-ownership as the basis for opposition to coercion<sup>10</sup>. If we own ourselves, then others need our approval for legitimate intervention, and coercion ignores that need for approval. However, as Jerry Cohen has shown, self-ownership makes our approval necessary simply because, in English, self-ownership means having a right not to be coerced: hardly a suitably independent justification for opposition to coercion<sup>11</sup>.

A somewhat better defense of commonsense opposition to coercion rests on Kant's first formulation of the categorical imperative: on the so-called formula of universality. This defense comes in different stripes. For example, some philosophers argue that it is inconsistent to will for everyone to coerce others all the time, since this would involve willing to be coerced to act against one's will. Others argue that it is unreasonable to *agree* to a moral contract that involves permission for everyone to coerce others even if they *disagree*. Still others interpret the formula of universality differently. I shall not venture a special response to each of these professed grounds. Instead, I point to a consideration that counts against the bulk of them. We started out by saying that there is no one constant threshold against coercion: commonsense morality's level of opposition to a particular form of coercion depends on the kind of coercion involved, on the risk and the impact of the coercive act, on the sphere it affects, and on still other factors. Actions involving coercion of minor form are seen as justified all things considered when the consequences of these actions are fairly good. Not so for actions involving severe coercion. They are harder to justify, and usually they are wrong all things considered. The formula of universality, by contrast, tends to yield a single constraint or a steady threshold. Either an act could be (gainfully) universalized, or it could not: either it is consistent to will certain scenarios, or it is inconsistent to will them. Either it is reasonable to agree to certain contracts, or it is not. The output of the first formula is too black and white to ground commonsense opposition to coercion, where there are many shades of grey<sup>12</sup>.

A third deontological ground for commonsense opposition to coercion is respect for autonomy<sup>13</sup>, often associated with Kant's injunction to treat persons with respect<sup>14</sup>. The rough idea here is that persons are dignified ends in themselves in virtue of being autonomous. Coercion treats a person as a lesser being, as though she lacked the capacity for autonomy or the dignity associated with autonomy. That is why coercion is disrespectful, and hence wrong. Some Kantians add that some forms of coercion are more disrespectful than others, thus seeking to explain the variety of threshold levels against coercion<sup>15</sup>.

The argument usually consists in elaborate proofs that everyone's autonomy has supreme value, or that we have reason to treat it as if it did; this is typically the end of the argument. The consideration I have just mentioned is thought to clarify why coercion is disrespectful and thus difficult to justify. Authors assume, without arguing, that the relevant forms of coercion treat or represent people as though they lacked autonomy or as though their autonomy did not matter. They seem to understand "respect for autonomy" as involving refrain from acts that express a message that disrespects autonomous beings: the message that someone lacks full autonomy or that her autonomy doesn't matter<sup>16</sup>. Their argument is that coercion always expresses this disrespectful message, and that, in the light of the dignity of autonomy, coercion is always at least somewhat difficult to justify.

My response is that coercion that incurs intuitive opposition do not always express that disrespectful message: not, in any case, more than refraining from them

would. For example, we already mentioned that intrusive medical intervention often promotes net autonomy by foreseeably boosting the patient's, or other people's, clinical outcomes and autonomy. Furthermore, it can protect clinical outcomes and autonomy from potential deterioration, thereby protecting autonomy. Thus, coercive and intrusive medical intervention can promote or protect overall autonomy. While interference with autonomy may tend to express disrespect for autonomy, surely failure to promote or at least to protect autonomy has the same tendency. Such neglect also tends to express that autonomy is nonexistent or unimportant. Put differently, either forcing intrusive medical intervention or not forcing it tends in one way to express disrespect for autonomy. Neither option is free of disrespectful tendencies. Either tends in one way to express that someone lacks autonomy or that her autonomy is unimportant-unworthy of promotion or protection (in the one case) or unworthy of noninterference (in the other). The need to avoid expressing disrespect for autonomy does not determine whether or not to impose highly intrusive care.

It might seem as though promotion and protection are unsuitable ways to express recognition of autonomy. Thomas Hill says that the value of autonomy is not a value to be promoted; it is a value that should inspire awe<sup>17</sup>. But what is wrong with translating one's awe into protection and promotion more than into noninterference? Perhaps awe and respect invite noninterference more than they invite protection and promotion, only as a psychological matter. But surely a mere psychological tendency does not constitute moral justification. Perhaps awe and respect invite noninterference more than they invite protection and promotion by their very definitions. For example, respect and awe may be definitionally linked to deference, which is definitionally linked to noninterference and to protection and promotion. However, if that were the case, then the fact that autonomy commands supreme value would not clarify that autonomy should be treated with awe and respect, and not with "shmwaw" and "shmespect", which are my terms for similar things, only involving definitional ties to promotion and protection, not to noninterference. Put differently, that the value of autonomy reigns supreme generates moral reason to promote autonomy, if necessary, through coercion, just as much (or just as little) as it generates moral reason to show awe and deference for it. Nothing in the notion of autonomy and in the fact that it has supreme value seems to decide between valuing autonomy in one way or the other<sup>18</sup>.

I do not wish to deny, of course, that commonsense morality opposes coercion. Nor do I wish to deny that arguments like the ones that Kantians marshal against coercion – that it violates autonomous persons, their dignity and their rights – intuitively feel in place. But on reflection these arguments are found to be invalid. Their conclusion is not borne out by their premises. Nor do I wish to argue that coercion becomes intuitively permissible once we note this invalidity. Our opposition to coercion has deep causal roots, perhaps in our sociobiological makeup, and it is not easily overcome. The appearance of utter disrespect does *not* fade away when a doctor precedes coercive intrusive treatment on a competent adult by explaining: "I shall now force intrusive intervention on you despite your refusal, but

please do not take it personally. I believe that you have autonomy, and that autonomy is very important. By coercing you to undergo treatment, I am promoting autonomy more than I am violating it". But there is a difference between a strong appearance of disrespect and the actual reality of disrespect. We have not found a valid argument showing that extreme forms of coercion always involve actual disrespect for autonomy.

### *6. A new utilitarian justification*

Our explorations so far have led us to an inconvenient, but I believe unavoidable conclusion regarding commonsense morality's opposition to coercion. There are probably no homogenous thresholds against the forms of coercion that commonsense finds problematic. Neither pragmatic utilitarian nor autonomy-based utilitarian nor deontological grounds for such thresholds are successful. It is hard to see what other grounds could have greater success. It therefore seems that central commonsense intuitions surrounding coercion are false. They uphold homogenous thresholds exist. The truth is that it does make a lot of moral difference whether someone is in the room, whether the patient is conscious, and so forth.

Nevertheless, I shall now argue, it remains justified for an individual to conform to the bulk of commonsense intuitions on coercion in her conduct and even to believe in their truth. She is even morally justified in trusting those intuitions that specify that self-ownership, autonomy and so forth are appropriate grounds and motives for noncoercion – the very intuitions that we argued against above. The justification for honoring this large section of commonsense morality despite its falsity is, I shall suggest, act-utilitarian or otherwise act-consequentialist. The limitations of other utilitarian justifications we examined notwithstanding, utilitarianism offers the best determinate justification for honoring commonsense opposition to coercion.

This new utilitarian justification borrows a crucial move from Hare. It also focuses on a category of harm that, no less crucially, Hare neglected; and this focus is what allows it to generate commonsensical results that Hare's framework as he fleshed it out does not successfully generate. The new utilitarian justification for commonsense opposition to coercion starts by pointing out a fact we already observed. Incorrect as it may be, the perception that certain forms of coercion are disrespectful of autonomy is both prevalent and persistent. For example, we just mentioned that people would typically perceive highly intrusive coercive care as evincing utter disrespect toward the patient as an autonomous person, no matter what the doctor explained to convince them otherwise. The new utilitarian justification is therefore concerned about such medical intrusion. As mentioned, these perceptions are probably false: such intrusion does not always express disrespect. But these perceptions typically generate a special kind of disutility. The

perception that one has been treated disrespectfully can generate “dignitary” harms like a painful sense of humiliation, decline in one’s level of subjective self-respect and in other good things that causally hinge on a healthy self-respect, like a positive mood and willingness to pursue plans<sup>19</sup>. If a person feels disrespected as an autonomous being, her ability or her willingness to exercise autonomy may decline – as Mill argued<sup>20</sup>. Moreover, if someone is *repeatedly* treated in a way that is broadly perceived as disrespectful, the respect of third parties who observe how she is being treated could take a blow. She may ultimately lose their concern and friendship. She may also encounter social exclusion, active stigma, discrimination, neglectful treatment or, following an exceptional, relentless process of dehumanization, even violence. Finally, imagine a society where mutual treatment or the treatment of minorities is regularly perceived as disrespectful of autonomous persons. In such a society, the culture of respect for autonomy, which is highly beneficial overall, is clearly at risk. Utilitarianism tells us to minimize harm, including dignitary harm. Toward that end, I now argue, utilitarianism tells individuals to cultivate their trust in and their adherence to commonsense opposition to coercion. Each individual should cultivate her tendency to believe that certain act-types, which in the foreseeable future most others will continue to perceive as disrespectful (no matter how much she tries to convince them otherwise) are indeed disrespectful. She should also cultivate her tendency to believe others’ interpretation of what makes these act-types disrespectful. Once an individual cultivates these tendencies, she will often adhere to commonsense thresholds quite homogeneously, out of commonsense beliefs about the correct reasons to adhere to them.

Let me elaborate. We already know that utilitarianism gives us reason to cultivate adherence to relatively complex rules (rules with a longish “*unless...*” clause that makes exceptions for factors that commonsense deems irrelevant, like whether someone is looking). That coercive actions tend to generate dignitary harm somewhat confines the set of circumstances in which coercive conduct produces more benefit than harm. But that consideration still gives factors like whether someone is looking a lot of weight. Crucially, however, utilitarianism also tells each to cultivate beliefs that minimize harm. For example, utilitarianism tells each to preserve and foster her natural tendency to believe that intensive medical intervention on an autonomous adult is utterly disrespectful because it violates autonomy. If we feel respect for persons, as we ought to anyhow, this belief will bolster our tendency to adhere to the complex rule. It thus helps each to minimize the dignitary harm that she causes; and, as I explain below, it does so without counter-productively generating dignitary harm through, for instance, feigned explanations that only cause offense. In that respect, cultivating this belief has utilitarian value and justification. This belief then affects behavioral patterns. It somewhat “evens out” the agent’s behavioral tendencies until they become quite “homogenous.” The agent becomes reluctant to perpetrate the relevant forms of coercion, and not only when the patient is conscious or someone is looking. The agent now thinks and, accordingly, acts like a Kantian; the justification for her personal makeover remains utilitarian.

Utilitarianism, in other words, recommends developing a character trait usually associated with Kantianism: homogenous refrain from coercion and other intrusive conduct, propelled by the belief that autonomous people's second-personal claims to respect are what makes such actions wrong. A "re-programmed" utilitarian will believe that it is respect, and not utility, which mandates that she treat people in those ways. From a utilitarian point of view, she will have cultivated a false belief, but admirably so. Cultivating a belief that one knows is false is necessarily wrong by the lights of many moralities. But utilitarianism does not find basic fault with it, not when it manipulates people in utility-enhancing ways<sup>21</sup>.

It may seem as though cultivating such "Kantian" beliefs is unnecessary in order to minimize dignitary harms: that refrain from acts that look disrespectful is enough. Can't the utilitarian doctor, for example, simply refrain from coercive medical intervention and explain to patients that the reason that she does not coerce them is that coercion would have likely offended them? The doctor could explain, for example, "I actually believe that there is nothing truly disrespectful about coercive care, which, after all, benefits patients and often protects their autonomy. But patients usually perceive coercive care as disrespectful. Therefore, to protect you from incorrectly feeling offended, I will not coerce you to undergo what I believe would have been beneficial to you". But to say so would usually offend patients in its own right. It appears condescending, neglectful, and disrespectful. It would generate dignitary harm. It does *not* maximize utility. Can't the doctor hide her real motivation, then? For example, she could conceivably lie to patients about the reason that they are not coerced to undergo unwanted care. She could say (as doctors sometimes do), "I don't coerce you to undergo this because you are an autonomous being". But to say so *while believing otherwise* runs a big dignitary risk. If the doctor's involuntary facial gestures, for instance, betray that deceitful scheme, most patients would feel doubly-offended. The doctor's approach would seem especially condescending, manipulative, and utterly unsuitable to autonomous adults. Dignitary harm could surge. A person can be pretty sure that she will cause pretty severe dignitary harm in the course of the many years in which she enacts such a deceitful scheme.

The simplest, safest way to minimize dignitary harm is to preserve the "Kantian" character trait – a trait that comes naturally to most of us. When occasionally an individual comes momentarily to doubt what usually none of us doubts – the link between respect for autonomy and coercion – utilitarianism tells her not to pursue that doubt further. Instead, it recommends trying to let go: slipping back into dogmatic slumber. To insist to run against the current and deliberate, either openly or secretly, as a "utilitarian", although others deliberate as "Kantians", could have caused severe dignitary harm to others. It could offend, humiliate, exclude and so forth on many everyday occasions. Utilitarianism therefore tells us to protect and to cultivate these character traits. The deliberative procedures that they involve maximize utility. Foster other deliberative tendencies, and you will violate the utilitarian dictum. Utilitarianism judges us based on whether we maximize utility. It

is indifferent as to the means, and that includes whether the deliberations that brought about utility maximization were false, from a utilitarian standpoint. Utilitarianism fully supports cultivating utility-maximizing Kantian beliefs.

If coercion did not *typically* diminish utility even in pragmatic terms and in terms of undermining personal autonomy, then perhaps this concern about dignitary harms, which encompass only one type of harm, would not suffice as basis for cultivating a “Kantian” character. However, the pragmatic and the autonomy-based utilitarian considerations noted above already make coercion *typically* wrong. Moreover, Richard Hare is right in holding that these considerations support following either the simple rule or the relatively complex rule. What dignitary harms do is to explain, better than Hare did, why we should oppose coercion in a homogenous fashion: why we should cultivate character traits that involve following the simple rule and not the relatively complex rule. Dignitary harms also explain why we should act in this way out of determinate beliefs and motivations, and not others. Someone who avoids coercion only out of concern about a potential fall-out with patients that would compromise third party care is still likely to spread much offense and other dignitary harms on an everyday basis. Once she cultivates the right beliefs and motivations, ones that minimize dignitary harm, they inform her thoughts and behavior even when no one is looking.

Utilitarian attention to dignitary harms is, then, key toward justifying homogenous opposition to coercion. Obviously, utilitarianism takes into account all harms and benefits, including traditionally neglected dignitary harms. This aspect of utilitarianism is what crowns it as the normative theory that best accounts for our reasons to honor an important section of commonsense morality: the section that surrounds coercion.

Similar considerations apply to other sections of commonsense that we associate with respect for persons. We perceive intentional harm as more disrespectful of ends in themselves than merely foreseeable harm to those ends. That perception boosts the utilitarian case for remaining reluctant to perpetrate intentional harm. It boosts it even if the perception is false (as I believe it is). That we perceive discrimination as failure to see people as distinct persons boosts the utilitarian case for reluctance to discriminate. That paying a dollar price for a poor man’s surplus kidney is perceived as disrespectful of a priceless being boosts the utilitarian case for reluctance to do so. Similar considerations apply to additional sections of commonsense morality. The very fact that people have certain moral beliefs and tendencies can sometimes be a source of harms or benefits. Minimizing dignitary harm can require that we cultivate personal traits that are largely in line with some of these beliefs: with deeply entrenched, widespread beliefs on how to show respect for full-blown persons. A sophisticated form of act-utilitarianism justifies trustful adherence to sections of commonsense morality that govern respect for persons. Ironically, Kantian approach to respect for persons fails to justify these constraints. The centrality of these sections of commonsense morality counts for, and not against, utilitarianism.

## Notes

<sup>1</sup> Mill, 1989, Ch. IV.

<sup>2</sup> Even patients' ultimate goals and religious commitments can be clearer to insightful doctors and psychologists who talk to these patients and their families than to these patients. I also believe that it can be bad for patients to fulfill their own ultimate goals (for example, an ultimate goal to spend life counting blades of grass or in painful service of an inexistent God). However, the present argument does not rest on this claim. Even if patients always knew their own ultimate goals best, and they always benefited from fulfilling them, doctors' judgment on what medical means can fulfill these goals would often be superior. It could benefit these patients more than doctors' heeding these patients' less learned views on the medical means that best fulfill these goals.

<sup>3</sup> Admittedly, coercing the patient to receive intrusive care would waste an opportunity to train the patient to make good decisions, but there will be other opportunities; forced interventions can boost the patient's clinical outcomes a lot.

<sup>4</sup> Hare, 1981, pp. 35-64; Hare, 1984 (where Hare comments on coercion). In a similar vein, Hare's account of what is wrong with slavery (Hare, 1986) is that the denial of "freedom" tends to produce material harm. Two caveats are worth noting. First, Hare focuses on coercive act-types in general, whereas for me, only most coercive act-types are intuitively problematic; second, Hare focuses on absolute refusal to perform such acts, whereas I focus on limited reluctance (sometimes strong and other times weak) to perform them. But these differences are immaterial to the present discussion.

<sup>5</sup> Similarly, it commends adherence to (*Don't murder*) more than to (*Don't murder unless everyone agrees that something clearly stands to be gained from murder, and no one will ever know that murder took place*), and so forth.

<sup>6</sup> Hare, 1984, pp. 122-3, 128-9; Hare, 1981, pp.35-6, 41.

<sup>7</sup> Hare also mentions that abiding by consequence-sensitive rules requires extensive and costly speculation about long-term effects. But the "unless..." clause in our context is not consequence-sensitive.

<sup>8</sup> We could have added here that an empowered, autonomous patient will often overcome disease at relative ease, or enjoy other psychological benefits. But when autonomy-violations remain secret, they cannot affect the victim psychologically.

<sup>9</sup> Mill, 1989, Ch. III; Raz, 1986; Hurka, 1996, Ch. 11.

<sup>10</sup> Nozick, 1986, p. 172; Taylor, 2004.

<sup>11</sup> Cohen, 1995, Chs. 9, 10.

<sup>12</sup> For a related argument, see Metz, 2002b, pp. 281-3. Kant took the formula of universality to lay down absolute demands. Contemporary Kantians are not absolutist, but it is still hard to see how *not* to interpret the formula of universality in black and white, regarding the existence or inexistence of a (finite) threshold.

<sup>13</sup> In bioethics, for example, Gerald Dworkin, Dan Brock, Allen Buchanan, Alan Donagan, and other leading philosophers and bioethicists use "respect for autonomy," "promoting autonomy," and "patient self-determination" interchangeably for the most fundamental basis for the medical principle of informed consent. The most influential textbook in the field titles the chapter on informed consent "Respect for Autonomy." See Beauchamp & Childress 2001, ch. 3, as well as pp. 63-4; Dworkin, 1988, pp. 110, 111, 113; Brock, 1994, pp. 31-2; Brock 1999, p. 523; Buchanan & Brock, 1989, pp. 36ff.; Donagan, 1977, p. 314; and Faden and Beauchamp, 1986, p. 8.

<sup>14</sup> Allen Wood writes that "coercion... obviously violate[s] the formula of humanity] because [it achieves its] end precisely by frustrating or circumventing another person's rational agency and thereby treat[s] the rational nature of the person with obvious disrespect. Wood 2006, 153. Wood emphasizes that respect for humanity is a matter of what we "express". Likewise, philosophical attacks on paternalistic coercion and coercive state perfectionism often rest on the value of respect for persons. See Herman, 1984, 600-1, 602; Nagel, 1987, p. 223 n. 8; O'Neill, 1985, pp. 264-5; Korsgaard, 1986, p. 335; Nozick, 1986, pp. 32-33; Darwall, 2002, p. 15; Darwall, 2006, pp. 128, 141-2; Anderson, pp. 1999, 301-2, 330-3; Hodson, 1983, pp. 46; Raz, 1986, p. 423; Margalit, 1996, pp. 224, 238, 259; Feinberg, 1986, pp. 23-

24, 27; VanDeVeer, 1986, pp. 111ff.; Shiffrin, 2000, pp. 207, 213, 215, 218, 220, 225, 231, 237; Taylor, 2004, p. 74; and Metz, 2002a.

<sup>15</sup> See Metz, 2002b, pp. 289f.

<sup>16</sup> “Respect for autonomy” may alternatively designate something like the nonviolation of a right, called “autonomy,” not to be coerced. But this just transforms the question whether coercion is justified, its potential benefits for the victim and others notwithstanding, into an awfully similar question: whether there is a right not to be coerced, potential benefits notwithstanding. This justification does not go far enough in addressing the initial request for justification.

<sup>17</sup> Hill, 2000, p. 140; Hill, 2002, p. 233.

<sup>18</sup> For proper Kantians, the first formula of the Categorical Imperative may suggest that noncoercion is more urgent than promotion and protection: we could not generalize a demand for promoting or for protecting anything, whereas we could do so for the negative demand not to coerce. However, earlier I explained why the first formula constitutes a problematic basis for commonsense opposition to coercion.

<sup>19</sup> For philosophical endorsement of similar ideas, see, e.g., Rawls, 1999, p. 386; Margalit, 1996.

<sup>20</sup> Mill’s fundamental argument does not condition this decline in personal autonomy on the truth of the perception that one’s autonomy is being disrespected. The argument rests on the existence, strength, and prevalence of that perception, false as it may be.

<sup>21</sup> It is true that having false beliefs, for example, being deluded about into believing one has true friends, can be somewhat bad for us. That utilitarian consideration counts against the proposed cultivation of false beliefs. However, this consideration pertains mainly to beliefs about central personal commitments, and mainly to areas where other people realize that one is deluded. I would insist that having a false belief about the forms of conduct that are disrespectful is an abstract matter that does not objectively blight a person’s life.

## References

- Anderson, E. (1999), “What is the Point of Equality?”, *Ethics*, 109, pp. 287-337.
- Beauchamp, T. L., and Childress, J. F. (2001), *Principles of Biomedical Ethics*, 5th ed., Oxford: Oxford UP.
- Brock, D. W. (1994), *Life and Death: Philosophical Essays in Biomedical Ethics*, Cambridge: Cambridge UP.
- (1999), “A Critique of Three Objections to Physician-Assisted Suicide”, *Ethics*, 109, pp. 519-47.
- Buchanan, A. E., & Brock, D. W. (1989), *Deciding for Others: The Ethics of Surrogate Decision-Making*, Cambridge: Cambridge UP.
- Cohen, G. A. (1995), *Self-Ownership, Freedom, and Equality*, Cambridge: Cambridge UP.
- Darwall, S. (2002), *Welfare and Rational Care*, Princeton, NJ: Princeton UP.
- (2006), *The Second-Person Standpoint: Morality, Respect, and Accountability*, Cambridge, MA: Harvard UP.
- Dworkin, G. (1988), *The Theory and Practice of Autonomy*, Cambridge: Cambridge UP.

- Faden, R. R., & Beauchamp, T. L. (1986), *A History and Theory of Informed Consent*, New York: Oxford UP.
- Feinberg, J. (1986), *Harm to Self*, Oxford: Oxford UP.
- Hare, R. M. (1981), *Moral Thinking*, Oxford: Oxford UP.
- (1984), “Rights, Utility, and Universalization: Reply to J. L. Mackey”. In R. G. Frey, ed., *Utility and Rights*, Minneapolis: University of Minnesota Press: 106-20.
- (1986), “What is Wrong with Slavery?”. In P. Singer, ed., *Applied Ethics*, Oxford: Oxford UP, pp. 165-84.
- Herman, B. (1984), “Mutual Aid and Respect for Persons”, *Ethics*, 94, pp. 577-602.
- Hill, T. E. Jr (2000), *Respect, Pluralism, and Justice*, Oxford: Oxford UP.
- (2002), *Human Welfare and Moral Worth—Kantian Perspectives*, Oxford: Clarendon Press.
- Hodson, J. D. (1983), *The Ethics of Legal Coercion*, Dordrecht: Reidel.
- Hurka, T. (1996), *Perfectionism*, New York: Oxford UP.
- Korsgaard, C. (1986), “The Right to Lie: Kant on Dealing with Evil”, *Philosophy and Public Affairs*, 15, pp. 325-49.
- Margalit, A. (1996), *The Decent Society*, Cambridge, MA: Harvard UP.
- Metz, T. (2002a), “Respect for Persons and Perfectionist Politics”, *Philosophy and Public Affairs*, 30, pp. 417-42.
- Metz, T. (2002b), “The Reasonable and the Moral”, *Social Theory and Practice*, 28, pp. 277-301.
- Mill, J. S. (1989), “On Liberty”, in his *‘On Liberty’ and Other Writings*, ed. by S. Collini, Cambridge: Cambridge UP.
- Nagel, T. (1987), “Moral Conflict and Political Legitimacy”, *Philosophy and Public Affairs*, 16, pp. 215-40.
- Nozick, R. (1986), *Anarchy, State, and Utopia*. Oxford: Blackwell.
- O’Neill, O. (1985), “Between Consenting Adults”, *Philosophy and Public Affairs*, 14, pp. 252-77.
- Rawls, J. (1999), *A Theory of Justice*, revised edition, Cambridge, MA: Harvard UP.
- Shiffrin, S. V. (2000), “Paternalism, Unconscionability Doctrine, and Accommodation”, *Philosophy & Public Affairs*, 29, pp. 205-50.
- Taylor, R. S. (2004), “A Kantian Defense of Self-Ownership”, *Journal of Political Philosophy*, 12 (1), pp. 65-78.
- VanDeVeer, D. (1986), *Paternalistic Intervention: The Moral Bounds of Benevolence*, Princeton, NJ: Princeton UP.