Forced to be free? Increasing patient autonomy by constraining it

Neil Levy

Correspondence to
Associate Professor Neil Levy, Florey Neuroscience Institutes, Royal Parade, The University of Melbourne, Parkville, Victoria 3010, Australia; nllevy@unimelb.edu.au

Oxford Centre for Neuroethics, Suite 8, Littlegate House, 16/17 St Ebbes St, OX1 1PT, UK.

Received 30 August 2011
Revised 18 October 2011
Accepted 16 December 2011

ABSTRACT

It is universally accepted in bioethics that doctors and other medical professionals have an obligation to procure the informed consent of their patients. Informed consent is required because patients have the moral right to autonomy in furthering the pursuit of their most important goals. In the present work, it is argued that evidence from psychology shows that human beings are subject to a number of biases and limitations as reasoners, which can be expected to lower the quality of their decisions and which therefore make it more difficult for them to pursue their most important goals by giving informed consent. It is further argued that patient autonomy is best promoted by constraining the informed consent procedure. By limiting the degree of freedom patients have to choose, the good that informed consent is supposed to protect can be promoted.

INTRODUCTION

Medical ethics as a distinct discipline, with its own norms, institutions and journals, is often said to have emerged out of the confluence of two factors. The first factor is the development of new medical technologies, which raised questions that seemed unprecedented: questions about when distinctively human life begins and when it ends, about the permissibility of using new techniques for creating and sustaining life, about the boundaries between human beings and the world around them. The second factor is a sense of outrage, provoked by a series of medical scandals: Nazi medical experiments, the infamous Tuskegee syphilis studies and so on. If the first development provided much of the material for bioethical debate, the second helped to shape the norms that emerged from this debate. In particular, the laudable urge to avoid repeating the crimes of the past led to the enshrining of patient autonomy as central to bioethics. This, in turn, led to an emphasis on the need for seeking and getting informed consent from patients for every procedure, major and minor.

The centrality of informed consent to bioethics is in some ways quite mysterious, insofar as the aim was to avoid another Tuskegee. Though requiring that the participants give informed consent would indeed prevent such incidents, the wrongs that occurred at Tuskegee were too egregious to make it plausible that informed consent procedures would be a remotely plausible fix for them. Hoping to prevent grave crimes in this kind of manner would be rather like hoping to prevent theft by passing a law requiring thieves to inform the police before committing a crime. It would work if it was adhered to, but there is no reason to think that the people it was aimed at would pay any attention to the requirement. If you have the kind of contemptuous attitudes towards patients exhibited at Tuskegee, you are unlikely to be restrained by informed consent procedures.

That is not to say, however, that there was no (perceived) problem to which emphasising informed consent was a solution. Informed consent was well designed to deal with the problem of everyday paternalism, which was once widespread among doctors who were genuinely seeking to do the best by their patients.[1] Unlike the doctors who were willing to participate in gross violations of human rights, ordinarily paternalistic doctors were open to persuasion that they should seek informed consent. They might be swayed by arguments against paternalism; alternatively, they might agree to abide by the norms of the profession regardless of whether they thought paternalism was justified or not.

We are well rid of the paternalism of the past. Doctors did indeed see themselves as appropriately exercising power over aspects of their patients' lives that they were not justified in claiming (think of involuntary sterilisations in cases in which doctors decided that a woman had had enough children). In this paper, however, I want to argue that the pendulum has swung too far. The picture of the rational individual that underlies the doctrine of informed consent is not psychologically realistic: we cannot expect patients to take on so much of the burden of making choices that will advance their own most significant interests. We ought to allow for, and even require, more in the way of directive counselling, even, sometimes, confrontational counselling. Though we should never ignore patients' wishes, it should be permissible to attempt to change their minds. Mild coercion will not just improve the quality of agents' decisions, it will actually increase their autonomy: since informed consent is justified just insofar as it protects autonomy, modifying informed consent in the way suggested does not represent a limitation on it, a compromise for the sake of other goods such as welfare, but will enable agents to increase their effective autonomy.

INFORMED CONSENT AND LIBERAL INDIVIDUALISM

In promoting informed consent to the central place it occupies in medical ethics, the discipline is in step with central currents in liberal political thought. Liberal thought is characterised, naturally enough, by its emphasis on the liberty of the individual. Normal adult human beings are conceived by liberals as having the right and the capacity to

make choices that advance their own significant projects. The role of the state, with regard to this conception of the individual, is to promote and harmonise the choices of free individuals. The limits of each individual's liberty are defined by the rights of other individuals: each of us is free to choose how to live and how to act, subject to restrictions stemming from respect for other individual's identical rights.

For liberals, the primacy of individual liberty entails a respect for the private sphere. The role of the state is to allow each of us to pursue what Rawls[2] calls our 'conception of the good'; our notion of what kind of life is valuable (a religious life or a hedonistic life, a life of devotion to good works or to knowledge, a life centred around family and so on). Liberals believe that the state must be neutral with regard to conceptions of the good, neither favouring any nor restricting any (so long as they respect the rights of others to pursue their own conceptions of the good). Different liberals offer different justifications of this neutrality: perhaps we cannot confidently judge the worth of rival conceptions of the good, perhaps people have a moral right to be wrong, or perhaps a profusion of what Mill[3] called 'experiments in living' is instrumentally valuable insofar as it allows for the assessment of different conceptions.

Regardless of the justification, liberals hold in common Mill's claim that ' over himself, over his own body and mind, the individual is sovereign'. [3] Mill argues that liberal thought entails state neutrality between conflicting conceptions of the good, and requires a renunciation of state paternalism. That is, the state may not interfere with individuals' actions, even to promote their own conception of the good. This claim is typically justified on epistemic grounds: individuals are best placed to judge for themselves how to pursue their conception of the good.

In rejecting paternalism, medical ethics extends the liberal conception of individual autonomy from the political sphere to the medical. Just as each of us has the right to pursue our own conception of the good without interference from the state, so, it is plausible to think, we have a right to pursue the good life as we see it without interference from medical professionals. Given the importance of health and life to almost all conceptions of the good (perhaps to all reasonable conceptions of the good), the protection of the medical sphere from unwarranted interference seems justified. Medical ethics might adopt Mill's words for its own:

> [T]he only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise, or even right.[3]

The idea that each of us has the right to pursue the good life without interference from others who regard our conception of the good as wrong or immoral is deeply attractive. Nothing in what I shall say here conflicts with it. However, there are grounds for thinking that some degree of paternalism might nevertheless be justified. Though it is unacceptable, on liberal grounds, to promote particular conceptions of the good or interfere with the pursuit of any reasonable conception, there may good grounds for some degree of paternalistic interference with individual choice when this interference can reasonably be expected to promote the pursuit of the good life by that very individual's own lights. We may be justified in interfering with choice when we do so to make people better able to pursue their own conception of the good.

John Rawls[2] famously identified a class of goods he called primary goods. Primary goods are goods that every rational individual can be presumed to want. Possession of a sufficiency of primary goods is almost always useful and never burdensome for the pursuit of any reasonable conception of the good, so no matter what else we want, we should want a sufficiency of primary goods. Primary goods include basic rights and liberties, a sufficient income, freedom of movement and occupation, and so on. The state does not violate its neutrality in ensuring that all citizens have a sufficiency of these goods: on the contrary, it is obliged to provide them if it can, since in doing so it promotes individuals' ability to pursue their conception of the good. Extending Rawls's thought, I will argue that medical professionals are justified in a certain degree of (mild) paternalism insofar as that paternalism can reasonably be expected to promote the primary goods that all reasonable individuals want to possess, or to prevent individuals from taking steps that would interfere with realising their own conception of the good. In making these claims, of course, I take issue with Mill's view that we cannot interfere even for these reasons. Mill's view is plausible, I shall claim, only if human reasoning is well designed to allow us effectively to pursue our conceptions of the good unaided. I shall argue that this view of human reasoning is overly optimistic.

LIMITS ON INDIVIDUAL RATIONALITY

The Enlightenment, out of which liberal political thought grew, was impressed by the power of individual rationality. From Reformation theology, with its emphasis on the individual ability to establish communion with God without the mediation of priests, to the French and American revolutions, central currents of Enlightenment thought emphasised the power of individuals to make decisions for themselves. More than anything else, it was the rise of science that seemed to make this faith in reason plausible. Scientific thought advanced at an ever-quickening pace from the 17th century on, and scientific explanations of natural phenomena became increasingly powerful and encompassing. Enlightenment thinkers saw this success as the result of throwing off chains: chains of deference to Church and tradition, and the chains imposed by despotic authorities. Enlightenment, as Kant saw it, consists in 'man's emergence from [...] the inability to use one's own understanding without the guidance of another'.[4]

The success of science is indeed the most impressive epistemic achievement in human history. However, there is good reason to think that in identifying individual rationality unchained as the driving force of science, the Enlightenment overlooked an equally significant factor: the social organisation of science. Science is the massively successful epistemic enterprise it is, in important part, because it is a distributed enterprise.[5] The distribution of cognitive labour occurs in two ways, one conflictual and the other cooperative. First, scientific claims are tested by researchers (typically groups of researchers: cooperative distribution of cognitive labour occurs within as well as between the units of knowledge production) working independently of one another, who have strong incentives to find fault with the work of rivals as well as to formulate and test hypotheses of their own. Second, researchers take on trust the results of this process, such that a claim that has been independently tested multiple times is very often simply accepted and incorporated into new work. An individual scientist is usually incapable of scientific research on her own: she needs access to the findings of others as well as to the specialised tools

Levy N. J Med Ethics (2012). doi:10.1136/medethics-2011-100207

that others have built (physical tools like fMRI machines or computers, or mathematical tools like tests for significance). In most fields, she also needs to be embedded within a research group to actually do science.

Independent testing helps to ensure that scientific claims are neither fraudulent nor spurious: that data is not faked nor arises from irrelevant factors (say contamination of samples or order effects). Independent testing is also typically hypothesis driven: that is, it aims at attempting to replicate the experiment, and does so in the service of a particular interpretation of the data, thereby aiding in the refinement of the theories that explain the data. This hypothesis-driven methodology, when combined with relations of trust between scientific research groups, enables rapid scientific progress. Researchers do not need to build their theories from the ground up; rather, they aim to add incrementally to the work of their predecessors (science moves so fast that a scientist's predecessors, in this context, may be a group whose work was published last week).

Because science is deeply dependent on a distribution of cognitive labour, it is perilous to infer from the success of science as an epistemic enterprise to the reliability of individual human rationality. It is not because scientists are freed from all constraints that science works as well as it does: it is because scientists accept a range of constraints on what can be said and how it can be said, on what counts as evidence and what should be tested. If anything, the success of science might support the opposite conclusion to the one drawn by Enlightenment thinkers: we might take it to show the epistemic importance of tradition, where 'tradition' is understood as constraining not content but form.

On the one hand, then, we can explain the success of our best epistemic enterprises (in important part) by reference to their social organisation. On the other hand, as I shall now show, when we enquire into the powers of individual human reasoners the picture we are presented with is in many ways bleak. Human beings are, under a variety of conditions, systematically bad reasoners, and many of their reasoning faults can be expected to affect the kind of judgements that they make when they are called upon to give informed consent. In what follows, I will outline a small subset of the evidence about the limitations of individual human reasoning, with special emphasis on the kinds of limitations that might be expected to be relevant to situations in which informed consent is sought.

THE FALLIBILITIES OF HUMAN REASON

Myopia for the future

Human beings discount the future at a rate, and according to a function, that is, irrational. It is rational to discount the future to some degree. For instance, it may be irrational for me to save so much of my income for retirement that I suffer real hardship now. That may be irrational because I may not live to enjoy my savings. Less dramatically, inflation and uncertainty regarding the future make it rational for me to prefer AU$1 today to AU $1.05 10 years from now. However the degree to which human beings typically discount the future, at least judging by their revealed preferences, is far greater than is rational. Moreover, the discount function they exhibit is clearly irrational.

Revealed preference theory infers agents' preferences from their behaviour. If we look at revealed preferences, it is clear that agents discount the future more than is rational. For instance, in addition to the millions of Americans who lack health insurance because they cannot afford it, there are millions who can afford it but fail to take it out. This indicates a preference for luxuries now over health later, which seems an irrational preference[6] Certainly, it is a preference that agents regularly later regret. Similarly, many developed countries have a pervasive problem with undersaving for retirement. A survey conducted by the UK Department of Work and Pensions found that 50% of adults between the agents of 25 and 34 did not save for retirement at all, despite the fact that 83% of them agreed that savings were the best way to ensure a comfortable retirement.[7] For these young people, retirement seems inconceivably distant and they therefore cannot motivate themselves to prepare for it when doing so comes at a cost today, even though they understand that it is likely that they will later regret their current behaviour.

The evidence that the function according to which we discount the future is irrational comes from careful studies of people's judgements across time. By examining agents' preferences (revealed or verbally expressed) for goods across time, we can map out the shape of their discount curves. When we do that, we discover that human beings' discount curves are hyperbolic, which is to say that they are highly bowed. This can cause oscillations of preferences across time.[8] A hyperbolic discounter may have the following preferences: at time t she prefers X to Y, and prefers that she continues to prefer X to Y from t right up until the time at which X is available. But at time $t1$, which occurs between time t and the time at which X is available, the same agent may have the opposite preference, preferring Y to X. When X and Y are goods that compete (say eating junk food and maintaining a healthy body weight, or buying shoes and saving for retirement), she may find herself unable to achieve long-term goals. She may continually frustrate her own plans: eating junk or running up credit card bills despite what she resolved this morning. Such an agent clearly experiences a diminution in her autonomy, since she is incapable of exerting her will over her own behaviour.[9]

It is easy to imagine circumstances in which steepness of discounting and hyperbolic discounting affect medical decision making. An agent who discounts the future too steeply may make decisions with regard to interventions that they can reasonably be expected to regret. For instance, an agent who elects not to take a drug in order to avoid burdensome side effects, but at the cost of much worse suffering further down the track, might be said to act irrationally. Admittedly, the conception of 'rationality' invoked here is normative, and not everyone accepts that either steep discounting or hyperbolic discounting is irrational (see Goldin[10] for discussion); however, it is uncontroversial that hyperbolic discounting in particular, may prevent agents' from achieving goals they value (further evidence for the irrationality of hyperbolic discounting will be adduced later in the paper). It is easy to see how hyperbolic discounting may lead agents to act in inconsistent ways, to the detriment of their health. For instance, an agent may go to the doctor's surgery with the intention of getting a blood test, but find she is unable to face the needle when the time comes. In the context of informed consent in particular, steepness of discounting and hyperbolic discounting may lower decision quality. Steepness of discounting might lead a woman who tests positive for the BRCA1 gene to choose not to have a double mastectomy, because she discounts the future; hyperbolic discounting might cause another who consents to the procedure to withdraw her consent when the time for the operation is imminent. In both cases, she might be said to put relatively trivial interests ahead of major interests; in the second, she fails to bring herself to act as she judges she ought to.

Motivated reasoning

There is an enormous range of evidence that human beings are not dispassionate in their assessment of claims. Instead, we defend views to which we are antecedently attached and discount evidence that is, inconsistent with these views. The classic illustration is the work of Lord *et al*.[11] They gave subjects two sets of (fabricated) evidence, one of which supported the view that capital punishment was an effective deterrent and one of which supported the opposite view. The evidence was carefully constructed so that neither set was more persuasive than the other. They differed in that each set used a different methodology: one compared states with and without capital punishment whereas the other compared the same state before and after the introduction of capital punishment (methodologies were switched across conditions, so half the subjects got evidence from interstate comparisons in support of the deterrence claim, while half got intrastate comparisons in support of deterrence). Subjects' prior views about capital punishment predicted their assessment of the methodology: in other words, subjects dismissed the evidential value of the data that conflicted with their prior view. Worse, subjects' attitudes actually hardened after being presented with the data, despite the fact that it was designed to be entirely equivocal.

Since this classic study, there have been many replications of the motivated reasoning effect. In one recent study,[12] subjects were given mock news stories that contained mistakes (eg, they claimed that weapons of mass destruction had been found in Iraq). Some subjects also received information authoritatively correcting the error. They found that subjects who received false information followed by a correction actually believed the false information more than those who received no correction. The effect was greatest on those most partisan: those who wanted to believe that weapons of mass destruction were found (for instance) were left with a stronger belief than ever.

We do not need to imagine circumstances in which motivated cognition affects medical decision making: there are experiments that measure the effects of pre-existing views on these kinds of decisions. One classic study[13] examined how subjects filter out worldview inconsistent information. Subjects were played messages warning of the link between smoking and cancer. The information was difficult to hear because it was accompanied by heavy static, however subjects could shut off the static simply by pressing a button. Smokers pushed the button less often than non-smokers, but when the message was altered so that it disputed the link between cancer and smoking the pattern of responses was reversed. Similarly, Kunda[14] found that subjects who read (fabricated) information about the link between heavy coffee consumption and increased risk of breast cancer disbelieved the information, but only if they were female and heavy coffee drinkers. In other words, response to the article was predicted by whether or not the information conflicted with the subjects' behaviour. We can expect this bias to affect how patients process information with regard to the risks and benefits of treatment options, according to their motivation to engage in behaviours that are risky. This may reduce the quality of the decisions they make in the informed consent procedure, leading them to disbelieve information only because of their biased information processing.

Affective forecasting

There is plentiful evidence that people overestimate the effects of events and changes in circumstances on their level of well-being. They think, for instance, that were they to become disabled their level of well-being, understood as their degree of satisfaction with their lives, would plummet and remain low; conversely they believe that were they to win the lottery their degree of well-being would soar and remain high. The evidence suggests that these predictions are wrong: in fact people tend to adapt to their circumstances. This phenomenon, known as hedonic adaptation, ensures that events and changes in circumstances have smaller effects on our well-being than we expect.

The evidence that becoming disabled does not have the effect on subjective well-being that we expect comes from comparing the judgements of able-bodied people as to how they would feel were they to become disabled with the judgements of people who actually become disabled. At 1 week after experiencing a disability, negative emotions outweigh positive ones, but by as soon as 8 weeks the subjects report a preponderance of positive emotions.[15] The same phenomenon, in the reverse direction, occurs after winning the lottery.[16]

Once again, it is easy to see how this might affect medical decision making. A patient may judge that their quality of life would be unacceptably low were they to undergo an amputation and therefore elect to treat a gangrene infection with antibiotics, despite being told that this course of action carried with it a high probability of catastrophic failure. Or they might opt for high-risk surgery rather than carry the relatively mild burden of requiring twice-daily medication. In these cases, their informed consent would be the product of a mistaken judgement concerning the consequences of the rejected course of action.

Affective recall

We are unreliable at predicting how future events will make us feel, and we are bad at judging how past events actually made us feel. We are not even as reliable as we might think when it comes to judging whether an experience we are having is positive or negative while we are undergoing it, which will make recalling its actual qualities extremely difficult.

Our judgements of the unpleasantness of experiences are overly sensitive to two features of those experiences: their peak intensity and how they end. As a consequence, people may come to prefer undergoing more unpleasant experiences to less, if they differ in how they end. Subjects will recall experience 2 as less unpleasant than experience 1 if experience 2 is identical to experience 1 except that its ending is less unpleasant, regardless of the duration of the experiences. That is, we can turn experience 1 into experience 2 simply by making sure it lasts longer, even when the extra time added on the end is not pleasant.

The classic experiment demonstrating the peak-end rule had subjects listen to a loud unpleasant noise through headphones.[17] They heard 8 s of loud unpleasant noise in experience 1, and 16 s of noise in experience 2; in 2, the first 8 s were identical to the sounds heard in experience 1, but this noise was followed by 8 s of less unpleasant (but still unpleasant) noise. Clearly experience 2 is worse than experience 1: it is experience 1 plus some more unpleasant experience. Yet when subjects were asked which experience they would rather repeat, they opted for 2.

Evidence that subjects are unreliable at judging the nature of their concurrent experiences is also plentiful. We use contextual information to help us to judge the nature of an experience. Schachter and Singer[18] injected their subjects with either norepinephrine, which causes autonomic system arousal, or a placebo. Subjects were then asked to wait with another subject, who was actually a confederate of the experimenters. In one condition, the confederate fooled around while waiting for the experiment (ostensibly) to begin, in the other condition the

confederate expressed anger at the wait. Subjects who had been injected with norepinephrine experienced the arousal caused by the drug, but interpreted it in line with the contextual cues provided by the confederate: as happiness in the first condition and anger in the second.

Again we do not need to imagine cases in which our unreliability at judging the quality of our experiences affects medical decision making. Redelmeier and Kahneman[19] examined the effects of adding an unpleasant experience at the end of a colonoscopy to patients' assessment of the procedure (simply by leaving the scope in place at the end of the procedure). They found that those patients who had the uncomfortable procedure followed by a less uncomfortable and unnecessary wait for the scope to be removed rated the experience as less unpleasant. Moreover, patients in this group tended to be more likely to report for a follow-up procedure than those who got the standard colonoscopy. It seems implausible to suppose that the second procedure was really less unpleasant than the first, since the second procedure was simply the first with an additional unpleasant experience immediately following it. It therefore seems best to understand the judgement as the product of a cognitive illusion.

Insofar as patients are sometimes called upon to make decisions regarding treatments they have experienced before, or which might be expected to lead to the repetition of experiences they have had before, unreliable affective recall may lower the quality of their informed consent.

It would be very easy to multiply examples of the pathologies to which human reasoning is subject, even limiting ourselves only to biases that can reasonably be expected adversely to affect medical decision making. We are subject to a variety of pathologies when we attempt to assess probabilities: base rate neglect leads us to overlook how typical an event actually is, the representativeness heuristic and saliency effects cause us to be overimpressed by cases that come to mind easily. The confirmation bias leads us to look for evidence that supports a claim and overlook evidence that conflicts with it; framing effects cause us to make different judgements with regard to identical cases, depending on how the cases are described and so on. I will say just a few words about one other phenomenon: the resource dependence of good decision making. When we are under cognitive load (stressed, tired, multitasking) or when our cognitive resources are depleted for some other reason (the most significant source of depletion seems to be recent calls on these resources), all the heuristics and biases are exacerbated.[20] We are more subject to base rate neglect, to motivated reasoning, more susceptible to irrelevant framing or trivial features of our surroundings and so on. This is obviously very important in the context of medical decision making, for two reasons. First, patients may be asked to make a series of decision. When they do so, they can be expected to suffer decision fatigue and a consequent decline in the quality of their judgements. Second, and more pervasively, almost by definition the context in which informed consent is sought is a stressful one. The cognitive resources of patients can be expected to be at a low ebb in these circumstances: because they may be overwhelmed with information and because (obviously) the decision is a significant one, which will be found stressful by all patients.

TOWARDS PSYCHOLOGICAL REALISM IN MEDICAL ETHICS

The motivations for making informed consent central to medical ethics were laudable, but as it currently conceived, it rests on implicit assumptions with regard to the capacities of normal human beings that may be unrealistic. At least insofar as the doctrine rested upon the supposition that normal human beings can be expected, unaided and in stressful and novel contexts, to make choices that contribute to the achievement of their own most cherished ends, it seems to be in trouble. Of course, individual autonomy is, very plausibly, important enough that we ought to promote it even if it predictably imposes costs on some individuals, but the evidence is accumulating that these costs are far higher than most people imagine. It may be possible to avoid much of this cost without unduly infringing on autonomy. Indeed, it may be possible to avoid these costs while actually increasing autonomy.

The simplest way to avoid these costs, it might be thought, is by ensuring that patients are given the chance to reassess their decisions. We may think that though patients may initially be overwhelmed by the stress of the decision facing them, if we allow them to reflect again, perhaps over the course of several days (when this is practicable) and change their minds, they will be able to avoid some of these pathologies. Indeed, this kind of strategy may be a part of the solution to the problems outlined above, but there is reason to think that it will be of limited usefulness by itself. There is a large literature on what happens to our judgements after we have made a decision. As a consequence of making the decision, our attitudes to the alternatives change: we come to think the option we have chosen is far superior to the options we have rejected. The relevant mechanism here seems to be cognitive dissonance: because we are aware of the attractive features of those options we have decided not to pursue, we experience dissonance, and dissonance can be resolved by changing our judgements. This phenomenon is known as the spreading of alternatives: alternatives that were initially thought of more or less equal (or at any rate, not very dissimilar) value come to be thought as very unequal after one is chosen. We tend to come to see the option we have chosen as very much better than those we have rejected.[21] Lieberman *et al*[22] showed that the spreading of alternatives does not require that subjects recall the option chosen. This evidence suggests that if patients do not change their minds about their initial judgements, this may not be because the initial judgement was the one that was really most in accord with their values.

Alternatively, we might look at ways in which we can prevent patients being subject to the biases and other pathologies in the initial context of choice by teaching them techniques to avoid it. This is known as debiasing in the psychological literature. There have been some successes with the application of debiasing. For instance, prompting subjects to conduct symmetrical memory searches seems to mitigate the effects of confirmation bias to some degree.[23] In general, however, debiasing has not proved very effective, and is unlikely to be of very much use in the context of the informed consent procedure. For one thing, there are very many biases to correct for. For another, the context in which these decisions are made, with its inevitable stresses, is far from conducive to the application of these strategies, which are typically cognitively demanding. Though debiasing could be a part of the solution, it comes nowhere near to solving the problems outlined above on its own.

Debiasing is an attractive strategy because it avoids placing any pressure on patients. Given its limitations, however, we have good reason to look beyond debiasing, towards measures that are somewhat more coercive (inasmuch as they involve confronting and placing pressure on patients, while leaving the final decision in their hands). Many philosophers would balk at this suggestion, because they believe that putting pressure on patients infringes on their autonomy. This kind of worry has caused some thinkers to look for non-coercive alternatives. In

a recent article constructing a parallel argument for what he calls 'institutional prosthetics' (the promotion of good social and individual choices by designing institutions so that the biases inherent in human psychology dovetail with the options that promote human goods), JD Trout[24] argues that interventions that harness biases are minimally intrusive, and therefore are not paternalistic. He offers a test for whether an intervention is paternalistic: does it conflict with what would be chosen by fully informed and unbiased decision maker? While I endorse institutional prosthetics, and suggest that they may be helpful with regard to the kinds of political and social problems Trout has in mind and with regard to medical decision making (for instance, presenting statistics in a frequency format may assist patients in some circumstances), I do not think that they have a large role to play with regard to informed consent. The range of possible situations in which informed consent must be sought is too great for institutional nudges to help much. Moreover, I doubt that institutional prosthetics offer a superior solution to the more coercive measure I propose, even from the point of view of avoiding paternalism alone. It is not obvious, first, that institutional prosthetics pass Trout's test for absence of paternalism; furthermore, the test is not one that tracks a morally important property.

It is not obvious that institutional prosthetics pass Trout's test, because there may be reasonable disagreements about what it means to be 'fully informed'. Is an agent fully informed if she is presented with all the relevant information in a format she is capable of grasping? Or does being fully informed require that she actually grasp the information? Or is the standard even more demanding: perhaps she is fully informed only if the information actually alters her beliefs and other propositional attitudes in the normatively correct way? To put the point in the context of informed consent, is an agent fully informed if she is (say) told about the phenomenon of hedonic adaptation, if she actually understands hedonic adaptation, or if she alters her attitudes towards her decision to take full account of hedonic adaptation? An institutional prosthetic, should one be available, that reliably leads agents to take full account of hedonic adaptation may reasonably be judged paternalistic, on the basis of Trout's test.

We can and should avoid the entire issue, by simply setting aside the question whether paternalism is involved in favour of focusing on the moral goods that antipaternalistic measures are supposed to protect. We should not fetishise antipaternalistic measures like the informed consent procedure. As Beauchamp[25] reminds us, informed consent has an ethical rationale: it is designed to respect the autonomy of individuals. If we can redesign the informed consent procedure so that it is sensitive to the evidence regarding the fallibilities of human reasoning without compromising autonomy (perhaps even while increasing it), it would be unethical not to do so. Even if interventions fail Trout's antipaternalism test, they are ethically permissible if they do not violate the goods that antipaternalistic measures are designed to protect.

Since the role of informed consent is to protect and promote the autonomy of individuals, we can best approach the question of redesigning it by reference to the concept of autonomy. Unfortunately, there is no agreed-upon definition of autonomy in the philosophical literature; worse, there are ongoing controversies about central aspects of it. However, there is substantial agreement on core features of autonomy, by reference to which we can guide our reconstruction of informed consent.

The core of autonomy (as its etymology suggests) is self-rule: the autonomous individual is not ruled by another person, or in thrall to any institution or government. However, absence of rule by others is not sufficient for autonomy. The autonomous agent must actually rule, or be capable of ruling, herself. That is, she must be capable of shaping her life as she wants: in accordance with her values, her projects and her conception of the good. It is for this reason that, say, addiction impairs autonomy: the addict may not be ruled by another, but she has difficulty in shaping her life in accordance with her overarching values.[26]

Informed consent procedures are justified insofar as they protect autonomy: that is, insofar as they conduce to allowing agents to shape their lives in accordance with their own values. Once we recognise this, we also ought to recognise the moral urgency of reforming informed consent to take the fallibilities of human reasoning into account. Though some will see in the proposals I will advance a threat to autonomy, I will argue that just the opposite is true: the constraints I will suggest (or at least something along the lines to be proposed) can be expected to increase autonomy.

The many problems with human reasoning threaten our autonomy in two ways (some threats work one way, some another, and some might work in either or both). Either they cause cognitive illusions, causing us to misapply our values, or they cause our actions to be driven by attitudes that, while in some sense ours, should not be identified with our values.

Most of the evidence outlined above concerns cognitive illusions. When an agent misjudges which of two experiences is more pleasant (or less unpleasant), when it seems to her that an argument is fallacious only because (unbeknownst to her) she is motivated to reject it, when she responds in a certain way to a case due to strictly irrelevant aspects of the way it is framed and so on, she is subject to cognitive illusions. A cognitive illusion can be understood as analogous to a visual illusion: just as a visual illusion can cause us to make erroneous judgements by causing us to misperceive aspects of the situation in which we find ourselves, so a cognitive illusion can cause us to make erroneous judgements by causing us to misperceive features of our circumstances. When we are subject to cognitive illusions we may act in accordance with our own values and our own conception of the good, but we misapply them because the world is not as we take it to be. When an agent is subject to cognitive illusions, she is not ruled by another, but she fails nevertheless to rule herself: her actions cannot be expected to advance her goals in the ways she thinks they will.

The second way in which the fallibilities of human reason threaten autonomy is by bringing us to act in ways that do not reflect our values by causing attitudes of ours that we do not endorse to play a crucial role in our behaviour. Hyperbolic discounting might be understood along these lines. Hyperbolic discounters, recall, are subject to preference reversals: though she prefers good 1 to good 2 at almost all times, when the opportunity to consume good 2 is imminent her preferences reverse. When an agent's discount curves cross, there need not be any fact regarding which she is mistaken. She may recall very clearly that, and why, she usually prefers good 1 to good 2. She may even be well aware that she can expect to regret consuming good 2 (when good 2 competes with 1: eating fast food or smoking with health, for instance). She does not seem, therefore, to be subject to any illusion. However, her autonomy is compromised. Autonomy is a diachronic property of agents: an agent rules herself when she is able to exert her will across time.[9] The agent subject to preference reversals is impaired in her autonomy because she cannot do this. Instead, she continually finds herself frustrating her own ends. She cannot effectively pursue health, say, because she regularly fails to go to the gym or refrain from

smoking. She cannot even effectively pursue pleasure, because she regularly spends money on gym memberships that she could have spent on holidays and throws out the packet of cigarettes she bought only minutes before.

We also finding ourselves acting in ways that do not reflect our values when we are under cognitive load, tired or stressed. Under these conditions, our behaviour tends to be unduly influenced by environmental stimuli, especially temptations, and also by our implicit attitudes; attitudes that we have either as consequence of our enculturation or perhaps innately.[19] These implicit attitudes are genuinely ours, in some sense, but they are not our values. When they have a content that diverges from the content of her conscious values, they are neither endorsed by her nor does she take herself to have reasons to act upon them (often she does not even take herself to have reason to act upon them at the precise moment they cause her behaviour: implicit attitudes typically cause by behaviour in ways that bypass our capacities for reflection upon what we are doing; sometimes, they cause us to misperceive the circumstances we confront and thereby cause cognitive illusions). They cause behaviour that conflicts with agents' plans and with their conception of the good and thus undermine their autonomy.

Since the fallibilities of human reasoning threaten to undermine autonomy, but the purpose of securing informed consent is to protect and promote autonomy, we have good reason to redesign the informed consent procedure in ways that help to avoid these fallibilities, even if the redesign reduces the scope for individual decision making in the procedure. We fetishise the procedure if we insist that the scope of decision making must be as broad as possible, even at the cost of a decrease in autonomy.

It would be a clear infringement of a competent patient's autonomy to have their decisions made by doctors, or anyone else to whom they have not granted this power, in accordance with values that are the doctors, or the hospitals, or what have you, and not the patient's. That would be rule by another: heteronomy. However it does not infringe the patient's autonomy if steps are taken to ensure that their decisions (a) reflect their own conception of the good and (b) promote the primary goods that all agents can be reasonably expected to want no matter what their conception of the good; this may remain true even if, left to his or her own devices, the patient would make different choices due to cognitive illusions or other influences.

What steps, concretely, ought we to take in reforming the informed consent procedure? Here I shall put forward some tentative suggestions: though I am confident that the evidence presented above demonstrates that we need to rethink informed consent, the question how this is to be done requires input from a variety of perspectives. Further, the suggestions I shall offer will be relatively abstract: they would require further work to bring them to the point of implementation (see box 1 for further details and suggestions).

Given the range of evidence that bears on how reasoning processes can be distorted, there is a case for introducing informed consent specialists. These specialists would receive special training in human reasoning and would be taught to be on the look out for the major pathologies so far identified. They might also be expected to communicate some of this information to patients. For instance, when they suspect that faulty affective forecasting might be distorting a patient's judgements (for example, in refusing an amputation), they might draw the phenomenon of hedonic adaptation to the patient's attention. There is some evidence that doing so tends to make people's affective predictions more realistic.[27] Informed consent specialists might even express their disagreement with a patient's choice, saying that they will come to regret it. They might point out when the choice conflicts with primary goods that the patient can be expected to value, such as health, absence of suffering and length of life. Perhaps the patient should be asked, or required, to meet with people who have found themselves in circumstances like the one she is in; those who made the choice she has made and those who made a conflicting choice.

Of course it is possible that a particular patient has a highly idiosyncratic conception of the good: perhaps some kind of religious view to which they are deeply committed, which leads them to place little value on a good that most people view as primary. In cases like that, when they arise (and assuming the patient is competent), the patient's choice should be respected.

Harder cases arise when a patient is known, with reasonable certainty, to have a conception of the good with which their choice conflicts, directly or indirectly. Should we respect choices when we have strong grounds for believing that they are made as a result of a distortion of reasoning, but the patient remains obstinate in their choice despite directive counselling of the sort envisaged above? I think we should, though there are steps we can take, with regard to institutional design, that make it less likely that patients will persist with such choices. We might institute mandatory cooling off periods after informed consent, during which the patient is given the opportunity to change their mind. Procedures that are especially likely to give rise to later regrets might require longer waiting periods and more counselling.

It should be noted that there are no panaceas for the pathologies of human irrationality. Any strategy aimed at leading agents to make better decisions short of coercion will itself fall prey to the very problems that it tries to solve: agents will irrationally discount the advice of counsellors or the psychological evidence that is, adduced; they will take themselves to be exceptions to the claims made and so on. The strategies advocated here attempt to pull off a balancing act: respect patients' autonomy by leaving the final choice in their hands, since the conception of the good to be advanced is (almost always) theirs, and not ours, but at the same time raise the quality of their decisions by reducing the extent to which they are subject to cognitive illusions and to which they make choices that they can be expected to regret. No strategy that leaves the final choice in patients' hands entirely avoids the pathologies of human irrationality, but the kinds of strategies suggested can be expected to reduce their power.

CONCLUSIONS

In the *Social Contract*, Rousseau argued that we may force people to be free. Interpreting this claim is difficult; insofar, however, as Rousseau meant to advance a particular conception of the good, to which individuals must subordinate their desires, his view was illiberal. The reforms to the informed consent procedure that I put forward do not force people to be free in this, illiberal, way. They do not promote a particular conception of the good. They are designed to promote individual patient's own conception of the good, whatever it might be. They do this by prompting patients to choose behaviours that conduce to that conception of the good or which are conducive towards the primary goods that patients can reasonably be expected to want, no matter what else they want. Further, they do not force patients to accept anything: though they are designed to require patients to rethink their choices, they leave the final decision up to the patient.

Many people in bioethics worry that informed consent procedures leave too little in the hands of patients. They worry

Box 1 Enhancing autonomy by constraining informed consent

- Constraints designed to enhance autonomy may be divided into two broad classes depending on the kind of pathology of human reasoning they target: correcting cognitive illusions and ensuring that decisions are driven by states of the agent with which she appropriately identifies. Under the first heading, we may include informed consent specialists, with training in the psychology of reasoning. These specialists have the job of detecting cognitive illusions in patients and informing them that they are likely to be at work in their decision making. They may indicate what decision they believe an agent who is not subject to the illusions would make, and perhaps even attempt to persuade the agent by encouraging them to speak with people who had earlier made decisions that were and were not influenced by the illusion.
- Under the second heading are included measures to ensure that the patient is neither unduly stressed nor fatigued when she makes the decision. They may also include measures to attempt to dissuade patients from changing their mind as a result of hyperbolic discounting. The final decision must never be taken out of the patient's hands: even if she has consented to a procedure, she must retain the option of withdrawing her consent. However, it may be permissible to place obstacles in the way of her withdrawing her consent (especially if the obstacles are themselves consented to). For instance, we can require that a patient who can be expected to change her mind as a procedure becomes imminent take a long series of steps to withdraw her consent: perhaps attending several counselling sessions. Less coercively still, the option to withdraw should not be made salient. Cooling off periods may also sometimes be appropriate.

that patients may have inadequate understanding of the information given to them, might receive too little information and might be unduly pressured by doctors. The perspective I am advancing here, though it does not entail that these worries are never warranted, comes from the opposite direction. Since we know that human beings, unaided, are subject to a dizzying variety of pathologies of reasoning, I hold that we ought not to expect patients to make crucial decisions unaided. Rather they should be helped and supported to make good decisions, and sometimes this help should come in the form of confrontation. We should tell patients when we think their decisions are distorted by cognitive illusions or when they are misapplying their values. We should do these things in the service of promoting their values and their conception of the good. To refrain from doing these things is not to respect autonomy, it is to decrease it.

Acknowledgements This paper has been considerably improved by the thoughtful comments of three reviewers for the *Journal of Medical Ethics*. The author is grateful to the Australian Research Council for funding the research leading to this paper.

Funding Australian Research Council.

Competing interests None.

Contributors NL is the sole author of this paper and conducted all research leading up to it.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Goldman A.** The refutation of medical paternalism. In: Arras JD, Steinbock B, eds. *Ethical Issues Modern Medicine*. Mountain View, Calif: Mayfield Publishing, 1983:58—66.
2. **Rawls J.** *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
3. **Mill JS.** *On Liberty*. London: Penguin Books, 1985. [1859].
4. **Kant I.** An answer to the question: 'What is enlightenment?'. *Political Writings*. Cambridge: Cambridge University Press, 1991:54—60. [1784].
5. **Kitcher P.** *The Advancement of Science*. New York: Oxford University Press, 1993.
6. **Frank RH.** Not insured, and not worried. *The New York Times* 6 October 1999.
7. **Adams I.** Fewer than half of people saving for retirement. *Guardian* 3 April 2010.
8. **Ainslie G.** *Breakdown of Will*. Cambridge: Cambridge University Press, 2001.
9. **Levy N.** Autonomy and addiction. *Can J Philos* 2006;**36**:427—48.
10. **Goldin J.** Making decisions about the future: the Discounted-utility Mode. *Mind Matters* 2007;**2**:49—56.
11. **Lord CG,** Ross L, Lepper MR. Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *J Pers Soc Psychol* 1979;**37**:2098—109.
12. **Nyhan B,** Reifle J. When corrections fail: the persistence of political misperceptions. *Polit Behav* 2010;**32**:303—30.
13. **Brock TC,** Balloun JL. Behavioral receptivity to dissonant information. *J Pers Soc Psychol* 1967;**6**:413—28.
14. **Kunda Z.** Motivation and inference: self-serving generation and evaluation of evidence. *J Pers Soc Psychol* 1987;**53**:636—47.
15. **Lucas RE,** Clark AE, Georgellis Y, *et al*. Reexamining adaptation and the set point model of happiness: reactions to changes in marital status. *J Pers Soc Psychol* 2003;**84**:527—39.
16. **Brickman P,** Coates D, Janoff-Bulman R. Lottery winners and accident victims: is happiness relative? *J Pers Soc Psychol* 1978;**36**:917—27.
17. **Schreiber CA,** Kahneman D. Determinants of the remembered utility of aversive sounds. *J Exp Psychol Gen* 2000;**129**:27—42.
18. **Schachter S,** Singer JE. Cognitive, social, and physiological determinants of emotional state. *Psychol Rev* 1962;**69**:379—99.
19. **Redelmeier DA,** Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 1996;**66**:3—8.
20. **Levy N.** Resisting 'weakness of the will'. *Philos Phenomenol Res* 2011;**82**:134—55.
21. **Brehm JW.** Postdecision changes in desirability of alternatives. *J Abnorm Psychol* 1956;**52**:384—9.
22. **Lieberman MD,** Ochsner KN, Gilbert DT, *et al*. Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychol Sci* 2001;**12**:135—40.
23. **Lilienfeld SO,** Ammirati T, Landfield K. Giving debiasing away: can psychological research on correcting cognitive errors promote human welfare? *Perspect Psychol Sci* 2009;**4**:390—8.
24. **Trout JD.** Paternalism and cognitive bias. *Law Philos* 2005;**24**:393—434.
25. **Beauchamp TL.** Informed consent: its history, meaning, and present challenges. *Camb Q Healthc Ethics* 2011;**20**:515—23.
26. **Levy N.** Autonomy, responsibility and the oscillation of preference. In: Carter A, Hall W, Illes J, eds. *Addiction Neuroethics*. Elsevier. forthcoming. London: Elsevier, 2012:139—51.
27. **Ubel PA,** Loewenstein G, Jepson C. Disability and sunshine: can hedonic predictions be improved by drawing attention to focusing illusions or emotional adaptation? *J Exp Psychol Appl* 2005;**11**:111—23.