Journal of Applied Philosophy

Journal of Applied Philosophy
doi: 10.1111/japp.12112

Consent's Been Framed: When Framing Effects Invalidate Consent and How to Validate It Again

ERIC CHWANG

ABSTRACT *In this article I will argue first that if ignorance poses a problem for valid consent in medical contexts then framing effects do too, and second that the problem posed by framing effects can be solved by eliminating those effects. My position is thus a mean between two mistaken extremes. At one mistaken extreme, framing effects are so trivial that they never impinge on the moral force of consent. This is as mistaken as thinking that ignorance is so trivial that it never impinges on the moral force of consent. At the other mistaken extreme, framing effects are so serious that their existence shows that consent has no independent moral force. This is as mistaken as the idea that ignorance is so serious that its existence shows that consent has no independent moral force. I will argue that, instead of endorsing either of these mistaken extreme views, we should instead endorse a moderate view according to which framing effects sometimes pose a serious challenge for the validity of consent, just as ignorance does, but one which we can solve by eliminating the effect, just as we can solve the problem of ignorance by eliminating it.*

1. Introduction

A patient who consents to a medical intervention when its prognosis is described as 90% chance of survival but who would dissent if it were instead described as 10% chance of mortality is subject to a *framing effect*. As Amos Tversky and Daniel Kahneman introduced the concept, a person is subject to a framing effect when she would express different preferences towards the same option, given the same information about that option, depending only on whether that information is expressed as a gain or a loss.[1] Thus, consumers who are more likely to use credit than cash when the price difference between them is described as a discount (gain) for using cash, but are more likely to use cash when that same difference is described as a surcharge (loss) for using credit, are vulnerable to a framing effect. And, in the initial example, survival suggests gain while mortality suggests loss, but a 90% survival rate is the same as a 10% mortality rate, so using one or the other description merely frames the same information about prognosis differently.

Though academic psychologists have been aware of framing effects for more than thirty years, philosophers have lagged behind in assessing their moral significance. We are starting to catch up, however. In a couple of recent papers, philosophers have examined the impact that framing effects might have on the moral force of consent. Jason Hanna argues that the existence of framing effects suggests that we should attribute less moral significance to consent than we might otherwise.[2] Shlomo Cohen, in contrast, suggests

that doctors may at least sometimes be permitted exploit framing effects when obtaining consent, if in so doing they benefit their patients.[3]

In this article, I will defend the moral significance of consent against both of these claims. My general argumentative strategy will be to show that framing effects pose no more of a problem for the moral force of consent than ignorance does. I will argue first that if ignorance poses a problem for valid consent in medical contexts then framing effects do too, and second that the problem posed by framing effects can be solved by debiasing, or in other words by eliminating those effects.

My position is thus a mean between two mistaken extremes. At one mistaken extreme, defended by Cohen, framing effects are so trivial that they never impinge on the moral force of consent. This is as mistaken as thinking that ignorance is so trivial that it never impinges on the moral force of consent. At the other mistaken extreme, defended by Hanna, framing effects are so serious that their existence shows that consent has no independent moral force. This is as mistaken as the idea that ignorance is so serious that its existence shows that consent has no independent moral force. I will argue that, instead of endorsing either of these mistaken extreme views, we should instead endorse a moderate view according to which framing effects sometimes pose a serious challenge for the validity of consent, just as ignorance does, but one which we can solve by eliminating the effect, just as we can solve the problem of ignorance by eliminating it.

Although my thesis is moderate in the sense just described, there is another sense in which it is quite radical. Medical ethics experienced a revolution in the latter part of the 20th century when it moved from a mere consent standard to an informed consent standard, requiring not just that medical professionals obtain consent but that they obtain informed consent. I will argue that we have just as much reason in medical contexts to eliminate framing effects as we do ignorance. Thus, my thesis is radical in calling for another revolution in medical ethics, from an informed consent standard to an informed *and debiased* consent standard.

I should articulate an important clarification before continuing. Framing effects, as strictly defined above (expressing different preferences over some options, depending only on whether the same information about those options is expressed as gains or losses) are just one type of bias that might impugn consent. There are a host of others, for example anchoring bias, availability heuristics, confirmation bias, the endowment effect, and the gambler's fallacy, and there are certainly others beyond those yet identified and labelled.[4] It is beyond the scope of this article to discuss all possible cognitive biases; the arguments I give about framing effects might apply to some or even many of these others, but they would need to be discussed individually. By examining the impact of framing effects on consent, I hope to make a start on how ethicists can respond to the threat of cognitive bias, but I do not want to oversell my conclusion, as applying to all forms of cognitive bias. I will, however, extend my analysis a tiny bit, to cover framing effects slightly more broadly construed than Tvkersy and Kahneman's original definition in terms of gains and losses. As I will argue in Section 4, a possible solution to the problem of framing effects, as narrowly defined by Tversky and Kahneman, cannot resolve a similar problem based on a broader definition of framing effects, where the two ways of expressing the same information need not be associated with gain or loss.

This article contains seven sections, including this introductory one. In Section 2, I discuss four familiar points about informed consent, which I will then apply, in subsequent sections, to the problem of framing effects. In Section 3, I argue that, while

framing effects do not typically invalidate consent, they do invalidate consent in medical contexts if ignorance invalidates consent in those contexts. In Section 4, I discuss and reject three possible ways of solving the problem of framing effects: appeal to counterfactual consent, disclosure of all frames, and using a neutral frame. In Section 5, I present my preferred solution to the problem posed by framing effects, that we debias, or eliminate framing effects. In Section 6 I consider some objections to my argument, and Section 7 is the conclusion.

2. Informed Consent

In this section, I will briefly discuss four points surrounding the potential problem that ignorance might pose for the moral force of consent. My discussion here is not intended to be comprehensive; I have selected these particular points for discussion because they will be useful in my examination of analogous issues surrounding framing effects. Before I begin, however, I should discuss an issue of terminology. Following common parlance in bioethics, I will abbreviate 'consent that has sufficient power to make an intervention morally permissible' as *valid* consent, where invalid consent is consent that lacks that moral power.[5] Coerced consent is typically considered invalid, and, in medical contexts, consent in ignorance of risks is also often considered invalid.

My first point is a reminder that *ignorance does not always invalidate consent*. In many contexts, whether a subject's consent to a transaction is valid does not depend on whether the subject knows various important details of that transaction. For example, my consent to buy a bicycle can be valid even if (a) the seller did not inform me of the degree of risk of severe injury in cycling and (b) I would not have bought the bicycle had I known about those risks.[6] On the other hand, the received wisdom in bioethics today is that valid consent in *medical contexts* must be informed of important details. For example, my consent to undergo surgery might be invalid if (a) the doctor did not inform me of the degree of risk of severe injury from the surgery and (b) I would not have agreed to undergo the surgery had I known about those risks. Why should there be a difference?

This question is challenging, and it has not received the attention it deserves. Steven Joffe and Robert Truog have recently addressed the issue obliquely, arguing that the fiduciary role of healthcare professionals as advisors explains why consent must be informed in medical contexts, with the obvious albeit unstated implication that the requirement for informed consent need not carry over to non-advisory contexts.[7] Bicycle vendors have no fiduciary obligations to their customers, whereas medical professionals plausibly do have such obligations to their patients. But why is there this further difference in fiduciary obligations?

I suggest, tentatively, two answers. First, medical decisions typically involve higher stakes than other decisions do, so it is correspondingly more important in medical contexts to promote patient autonomy: for example, typically a lot more is at stake in the decision whether to treat prostate cancer than whether to buy a bicycle. The point is not that terrible outcomes (e.g. death) will befall a patient if and only if her doctors do not intervene. In such cases we might think, paternalistically, that even uninformed consent to the intervention is valid. Rather, the point is that when the stakes are high it is important to respect a patient's autonomous decision, even if that autonomous decision leads to terrible outcomes, and decisions are more autonomous if they are informed than

if they are ignorant. Of course, this point leads to a further question, namely why we should respect someone's autonomous choice if it leads to terrible outcomes. That question is deep and challenging, and I do not here address it; for our purposes it is enough to note that, given this view about the value of autonomy, the requirement for informed consent is more plausible when stakes are higher.

Second, there is a significant information asymmetry in medicine, because (a) medical information is typically complex and fast-changing, to the point that it is unreasonable to expect non-experts to be able to find and evaluate such information competently, while simultaneously (b) medical professionals can and should be expected to develop and maintain expertise. It is a lot easier for a non-expert to find out and evaluate the relative merits of cycling versus taking the bus than to find out and evaluate the relative merits of surgery versus watchful waiting, and it is simultaneously a lot easier for a medical professional to find out and evaluate the relative merits of surgery versus watchful waiting than it is for the typical patient to do those things. For this reason, the cost of mandatory disclosure, which is incurred by medical professionals, is greatly outweighed by the benefit to patients.

The second point I discuss is that *counterfactual consent does not solve the problem of ignorance*. In other words, it is not enough to say, in the absence of informed consent, that the subject *would* have consented *had* she been informed. Rather, what is typically needed is actual consent in the light of actual knowledge. Analogously, the same idea holds when we focus on mere consent: typically *actual* consent to an intervention is needed in order to make that intervention permissible, not just reassurance that the patient *would* have consented *had she been asked*.

Of course, in some unusual situations, for example emergencies, there may be no time to obtain consent. In such cases, counterfactual consent may do some moral work: an emergency intervention might be permissible because the patient would have consented had she been able to make a decision.[8] Likewise, we can imagine perhaps fanciful situations where there is time to get consent, but not enough time to disclose all the relevant information needed to make an informed choice. In such a case, counterfactual consent in light of information disclosure may do some moral work: an intervention in this case might become permissible if both (a) the patient consents and (b) the patient would still have consented had all the relevant information about the intervention been disclosed.

My third point is that *the way to generate valid consent in the context of ignorance is to eliminate the ignorance*. Typically, the easiest way to eliminate ignorance is to disclose information, but that is not the only way. Suppose, fantastically, that a doctor has a hat that confers medical knowledge when worn. Instead of engaging in the tedious activity of educating her patients, this doctor could just ask her patients to don the hat, at which point her patients would know everything there is to know about the procedure. Then their consent, if given, would be valid. There are also less fantastic ways to eliminate ignorance without resorting to disclosure. For example, doctors could hand out pamphlets that contain information, asking their patients to read them, or they could tell their patients where or how to find the information for themselves, for example on the Internet.

The distinction between disclosure and elimination of ignorance raises a related, fourth point: *how much understanding is required for valid consent is a difficult and controversial question*. This question has received a lot of attention, including recently in the

arena of human subjects research because of the so-called *therapeutic misconception*, according to which research participants often believe unjustifiably, and in a way that is difficult to change, that they will benefit from research participation.[9] The question is difficult, and it gets at the very heart of why we care about providing information: if we want to ensure that research subjects make autonomous, well thought-out decisions, perhaps we should require understanding in addition to disclosure.[10] In contrast, if our goal is merely to ensure that we have treated potential research subjects reasonably and given them ample chance to back out, then perhaps we need require no more than disclosure and a good faith effort to obtain understanding.[11]

3. When Framing Effects Invalidate Consent

Now that we have reminded ourselves of a few points surrounding the problem of ignorance, we are ready to discuss the analogous problem of framing effects. I begin by asking when, if ever, consent in spite of framing effects is still valid. Recall the first point I discussed with respect to ignorant consent, that in a wide range of cases valid consent need not even be *informed*. In the example I gave in Section 2, I can consent validly to the purchase of bicycle even if I am ignorant of the risk of harm involved in riding it.

We should expect the same to be true for consent that is vulnerable to framing effects. For example, my consent to the purchase of a bicycle can remain valid even if (a) the vendor reports that 90% of its riders are satisfied and (b) I would not have consented had the vendor instead reported that 10% of its riders are dissatisfied. As a second example, it is perfectly permissible, though perhaps underhanded, for an advertisement to report 'four out of five dentists agree' rather than 'one out of five dentists disagree'. And, to return to an earlier example, my consent to use credit instead of cash can remain valid even if I think of my decision as forgoing a discount rather than as incurring a surcharge. In other words, in a wide range of ordinary cases the moral requirements for valid consent are quite lax: valid consent can be uninformed, and it can be subject to framing effects. This is true even if consent that is vulnerable to framing effects would not be valid in certain other contexts where stricter standards are appropriate.

A more interesting question, then, is whether there are any cases where validity requires the elimination of ignorance but not the analogous elimination of framing effects. We have assumed that consent in medical contexts must be informed in order to be valid, so the question naturally arises for such contexts: does consent to medical care also need to be free from framing effects in order to be valid? Cohen answers this last question in the negative, arguing that medical professionals should be allowed to nudge their patients towards consent by exploiting framing effects. In this section, I will show that Cohen is mistaken, first by arguing positively that the reasons in favour of an informed consent standard in medicine also speak in favour of addressing framing effects and second by rebutting Cohen's arguments to the contrary.[12]

To begin, a patient is less likely to choose autonomously if he is ignorant of relevant facts than if he is aware of them. Likewise, a patient is less likely to choose autonomously if he is incapable of reasoning consistently than if he is so capable. If the stakes of the relevant decision (or perhaps, if the stakes of a typical decision of that type, for example the medical type as contrasted with the cycling type) are sufficiently high, and if preservation of consent's validity requires a resolution of the problem of ignorance, then it also requires a resolution of the problem of framing effects.

One might at this point object that framing effects might change what people value, in which case being vulnerable to a framing effect is not a form of inconsistency or irrationality. For example, maybe seeing the word 'survival' makes people risk-seeking, whereas seeing the word 'mortality' makes us risk-averse, where this is interpreted as our having different values depending on what words we have recently seen. According to this objection, neither value is objectively better, and neither is more accurately described as more authentic; it is just that we have different values depending on subtle cues such as which words we have recently seen. The main problem with this objection is that, even if this is how framing effects work, it is quite irrational to change values depending only on what words we have recently seen. In other words, while having different values in different circumstances may be consistent and rational, and therefore compatible with autonomous choice, having different values depending only on how information has been framed is surely irrational, and therefore incompatible with autonomous choice.

The second rationale I mentioned in Section 2 for the requirement of informed consent in medicine is that there is an asymmetry between what we can expect medical professionals to know about medicine and what we can expect patients to know about that topic. On its face, this rationale may seem unable to support a requirement to resolve the problem of framing effects, because we cannot expect medical professionals to know about framing effects. First, cognitive psychology is not their area of expertise, and second even experts are susceptible to framing effects, at least when they are not consciously aware of them.[13]

Still, there is a sound argument from informational asymmetry to the conclusion that we should treat framing effects on a par with ignorance; the argument just needs a lemma, to the effect that, although we do not now expect any medical professionals be experts on framing effects, we should expect at least some of them to be. And clearly, what matters for the obligations of medical professionals is what they should know, not what they in fact know. Of course, this is not to say that the same doctor who will be performing the complicated surgery also needs expertise on framing effects, only that someone on the medical team needs expertise. Likewise, we do not expect that the person operating the MRI machine needs to be able to communicate the risks and benefits of MRIs — often that person is merely a technician — but someone on the team needs to be able to communicate that information to the prospective patient.

I have just offered a positive argument for the claim that, in situations where ignorant consent is invalid (e.g. in medical contexts), consent subject to framing effects is also invalid. In the remainder of this section, I rebut Cohen's arguments to the contrary. Cohen offers several arguments for the thesis that medical professionals may nudge their patients to consent to beneficial interventions. However, his arguments fail with respect to the exploitation of framing effects, even if they succeed for other types of nudges. Some of Cohen's arguments do not even apply to framing effects. For example, Cohen reminds us that some forms of nudging do not even exploit irrationality. For example, mandating a deadline by which patients have to make a decision often spurs those patients to think about their medical problems rather than put them off.[14] This argument does not apply to framing effects, of course, which are paradigmatically irrational.

Cohen also cites Onora O'Neill, suggesting that the purpose of informed consent is to preclude coercion and deception.[15] If so, and since the exploitation of framing effects is neither coercive nor deceptive, we can safely conclude that framing effects pose no

problem for consent.[16] The problem is that this argument begs the question: we want to know whether, in obtaining consent, we should be vigilant about things besides coercion and deception, things potentially including the exploitation of framing effects or other irrationalities. Another example of such an irrationality is drunkenness: if I exploit a drunk person's agreeable nature to have sex with her, her consent is clearly invalid, even if she was neither coerced nor deceived.[17] To stipulate that our only moral concerns are coercion and deception is to be insensitive to the problem rather than to argue against it.[18]

Cohen also endorses 'weak lexical priority' of autonomy over interests, according to which, while autonomy is lexically prior to interests, we choose an interpretation of autonomy that is compatible with promotion of interests whenever we can.[19] The problem is that it is no more an infringement of a patient's autonomy to refrain from exploiting a framing effect than it is to refrain from exploiting ignorance: if a patient's ignorant choice promotes her interests, then we wrongfully infringe on her weakly lexically prior autonomy when we try to eliminate her latent ignorance that led up to that choice. In fact, we should think the exact opposite: on any reasonable conception of autonomy, exploiting ignorance infringes on autonomy, and likewise exploiting framing effects does too.

Lastly, Cohen suggests that a patient's preference to undergo the procedure is her actual preference, even if that preference was produced by framing effects. The preference against the procedure, in contrast, is merely counterfactual. Cohen then points out that hypothetical consent is not real consent.[20] The problem with this argument is that we want neither counterfactual consent nor actual yet defective consent; we want consent that is both actual and free of defects. We can see that this is true because, as with the previous argument, this one would also support the validity of ignorant consent: a patient who gives uninformed consent still gives actual consent, and the fact that she would have dissented had she been fully informed shows only that her informed dissent is merely hypothetical.

4. Three Failed Solutions

In the previous section, I argued that if ignorance poses a problem for consent in medical contexts then framing effects do too. In this section, I consider three possible solutions to the problem of framing effects, and I explain why each one fails. The first two fail outright, and while the third can solve the original problem it cannot solve a more severe variant of it. The search for a solution is important because the possibility of a solution bears on Hanna's two main conclusions, first that consent vulnerable to framing effects is invalid and second (and admittedly more tentatively) that consent never has independent moral force. Hanna argues for those conclusions from the premise that we cannot solve the problem of framing effects.[21] Thus, if we can solve that problem, Hanna's conclusions will be undercut. An indirect goal in this section is to motivate the *plausibility* of Hanna's sceptical conclusions, even though I will ultimately argue, in the next section, that they are unwarranted.

The first failed proposal is an appeal to counterfactual consent. According to this proposal, we can be sure of consent's validity, even in the face of vulnerability to framing effects, if the subject still *would* have consented had she *not* been vulnerable to those

framing effects. After all, what it means for a subject to cease being vulnerable to framing effects is that her decision (consent or dissent) ceases to vary across frames. Then, we are permitted to intervene on a subject when, though her decision actually *does* vary depending on the frame, had her decision been invariant, that invariant decision would have been to consent.

It should not be surprising that this proposal fails, as the analogous proposal for the problem of ignorance also fails, as I discussed in my second point about informed consent. One problem with counterfactual consent is that, because it appeals to the truth of a counterfactual, we cannot operationalize it. In other words, we cannot tell, concretely, when a subject would have consented had she been immune to framing effects. We can easily tell when someone does or does not consent, but it is difficult to tell whether that person would have consented in the counterfactual scenario that she is immune to framing effects.

A more fundamental problem with this counterfactual solution is that, just as in the case of ignorance, we care about actual consent, not counterfactual consent. In the case of ignorance, recall, we are not satisfied with the alleged justification, 'yes, her actual consent is ignorant, but she still *would* have consented *had* she been informed'. No, we want actual information disclosure and actual consent in light of that actual disclosure. Of course, there may be unusual cases where there is not sufficient time to get actual consent, or actual informed consent, or actual consent free of framing effects, and in these cases hypothetical consent may have some role to play, but in general we want actual informed consent and actual consent free of framing effects, rather than their counterfactual cousins.

A second proposal, favoured by Ruth Faden and Tom Beauchamp, is that we simply disclose all the relevant frames.[22] For example, doctors might convey prognosis information both in terms of survival and in terms of mortality. The main problem with this suggestion is that disclosing all relevant frames need not elicit the subject's genuine preference. Perhaps, for example, a subject is genuinely pessimistic, but she will be overly (from her point of view, anyway) optimistic if the word 'survival' is used at all. Or perhaps, following a suggestion made independently by Hanna and by Frank G. Miller and Luke Gelinas, patients still cling to the initially presented description of prognosis.[23] Suppose, for example, that a patient consents when told that the prognosis is '90% survival, 10% mortality', but she would dissent if she were instead told that the prognosis is '10% mortality, 90% survival', because in each case she focuses on the first way in which the information is presented. If so, the patient is still under the sway of a framing effect, except now the relevant frames are not the simpler '90% survival' and '10% mortality' but rather the more complex '90% survival, 10% mortality' and '10% mortality, 90% survival', where whether a frame expresses information as a gain or a loss is determined by which word comes first, 'survival' or 'mortality'.

A third proposal is that we describe the intervention neutrally, in other words without framing it either as a gain (survival) or a loss (mortality). The problem with the second proposal was that it still uses words that connote gains and words that connote losses, and subjects might be vulnerable to framing effects just because such words are used, even if both gain-connoting and loss-connoting words are used. According to this third strategy, then, framing effects show that one or more ways of expressing information are biased and therefore should be avoided. Thus, our goal is to find a neutral, unbiased way of communicating that information instead. For example, instead of giving prognosis

information linguistically as either survival or mortality rates, perhaps we can present that information via a pie chart, with 90% filled in as survival, the other 10% as mortality. This strategy evades the problem of framing effects by appealing to a potentially neutral way of expressing information, the visual one, instead of the allegedly biased linguistic ones.[24]

One problem with this view is that it might not be possible to find neutral ways of presenting all the information we need to disclose. But let us waive this problem, as there is a more fundamental problem. The fundamental problem with the strategy of finding a neutral way of presenting information is that this strategy is still vulnerable to a more generalised problem of framing effects, where frames need not evoke the prospects of gains or losses. According to this more generalised notion of framing effects, a person is subject to a framing effect when she would express different preferences towards the same option, given the same information about that option, *depending only on how that information is expressed*. In contrast, recall that my original definition of framing effects is exactly like the one just stated, except that its replacement for the italicised clause is more restrictive: 'depending only on whether that information is expressed as a gain or a loss'.

On this more liberal conception of a frame, where frames need not invoke the prospect of a gain or a loss, conveying prognosis visually rather than linguistically merely presents that same information using a third frame. Yes, this third visual frame can be labelled 'neutral' in that it will not be construed as a gain or as a loss (and thus would not count as a frame on the original, more restricted way of thinking about framing effects), but that does not show that it is better able to elicit the subject's genuine preference, and so in the sense of 'neutral' meaning 'free of distortion' we have no reason to think that this third frame is any more neutral than the first two.[25] More generally, there is no reason to think that any particular frame — whether linguistic, visual, or anything else — is necessarily privileged as better able to elicit the subject's genuine wishes. It is not that the mortality frame always leads to irrational decisions because we are overly concerned to avoid losses, nor that the survival frame always leads to irrational decisions because we are insufficiently concerned to achieve gains, so that if we could only get a neutral frame then we could elicit the subject's authentic preference. Some subjects might identify most closely with loss aversion, others with gain seeking, and others in between, and we cannot hope to elicit the genuine preferences of all of these subjects by appeal to a single allegedly neutral frame.

Thus, consider two patients who would dissent to the medical intervention under the '10% mortality frame' but who would consent under the neutral visual pie-chart frame. However we identify authentic preferences, it is possible that only one of those patients authentically prefers to undergo the intervention. As another example, consider two pie-charts, one with survival coloured blue and mortality red, the other with the colours reversed. A patient might consent if presented with one pie-chart but dissent if presented with the other one, perhaps because he is more inclined to focus on information presented in red than in blue, but we cannot hope to claim that one version of the pie-chart is more neutral than the other. More generally, a patient might unconsciously be more or less attentive to different colours along a gradient, but even if so it is not clear that any particular colour is *neutral*; all we can say in this situation is that one colour is more attention-grabbing than another, not that one colour, let alone a pair of colours, is neutral with respect to its attention-grabbing effect.

5. Debiasing

The problem with the last failed solution — that we express information neutrally — is that even alleged neutrality does not guarantee that we capture what subjects genuinely want. A subject who consents under some frames but not under others is still vulnerable to a framing effect, even if, among the frames under which the subject consents are some allegedly neutral ones (e.g. a pie-chart). What we want instead, if we can get it, is some way to ensure that the subject's decision is invariant across all frames, so that we do not have to defend some particular frame as privileged or neutral. In other words, we want to *eliminate the framing effect*. In the jargon of cognitive psychology, eliminating a framing effect *debiases* its subjects. That debiasing is the best response to the problem of framing effects can be seen by reference to the analogous problem of ignorance, as I discussed as my third point on that topic: the solution to the problem of ignorance is the elimination of that ignorance; likewise, the solution to the problem of framing effects is the elimination of those effects.

Recall that, by definition, a person is subject to a framing effect if, though she *in fact* consents to the procedure under one frame, she *would* not consent if the information *were* presented under an alternate frame. In other words, framing effects themselves are defined counterfactually. Then to eliminate the framing effect, we must by definition make the following sort of counterfactual true: the subject in fact consents under one frame but *would still* consent *had* alternate frames been used. For example, a subject who consents under the '90% survival' frame and also would have consented under the '10% mortality' frame as well as the visual pie-chart frame is not biased in the relevant sense; this person's consent to the procedure with that prognosis is invulnerable to that particular framing effect.

Note that this successful counterfactual solution differs subtly yet crucially from the earlier rejected counterfactual solution, because it appeals to a different counterfactual scenario. The failed counterfactual solution was: *if the subject were not subject to a framing effect, she would still consent*. The correct counterfactual solution is: *if any alternate frame were used, the subject would still consent*. The difference between these two solutions is that the correct one suffices to show, by definition, that the subject is *in fact* invulnerable to a framing effect, whereas the failed solution shows only that if the subject *were* invulnerable to the effect, she would consent.

I have just suggested that the way to solve the problem of framing effects is to eliminate those effects, and I have shown that, by definition, the only way to eliminate a framing effect is to make a certain counterfactual — namely, that the subject still would have consented had any alternate frame been used — true. My work is not nearly done, however. That is because we want to know, not just on an abstract level that we solve the problem by eliminating the effect, and not just on a slightly less but still very abstract level that we eliminate the effect by making a certain counterfactual true, but how precisely to make that counterfactual true. To be told that we eliminate the effect when we make a certain counterfactual true is not much help; we want suggestions that can be easily operationalized.

Consider, first, a fanciful hat that confers perfect rationality on its wearer. By 'perfect rationality' I mean that it turns its wearer from a mere mortal into someone who recognises all equivalent frames and whose preferences are consistent across equivalent frames. If doctors had such hats, they could solve the problem of framing effects by

asking their patients to don the hat before telling them about prognosis and then asking for consent. Obviously, we have no such hats, but, as with the analogous case of ignorance, the point of supposing that we do is to make a theoretical point clear: we solve the problem of framing effects by eliminating those effects. How, in particular, we eliminate the effects is a separate and further question.

Given that we have no rationality-conferring hats, what real world methods might we adopt to eliminate framing effects? The most obvious solution, again on analogy with the problem of ignorance, is to eliminate those effects via *disclosure*. Here, however, we are not concerned to disclose facts about the intervention in question; rather, we disclose facts about framing effects and how they can make preferences irrational. If we call the former sort of facts *information*, because those facts are information about the intervention, then we might call the latter sort of facts *meta-information*, because they are facts about information, rather than facts about interventions. Thus, for example, 'the intervention is associated with 90% survival' conveys information about an intervention's prognosis, but '90% survival is equivalent to 10% mortality, and people are vulnerable to framing effects depending on which word is used' does not convey information about prognosis; it instead tells us about a framing effect that arises based on how information about prognosis is presented. Thus, just as one lesson of the problem of ignorance is that it might be important to disclose information, a lesson of the problem of framing effects is that it might be important to disclose meta-information.

What sort of meta-information might suffice to eliminate framing effects? The obvious answer is that we disclose *that there is a framing effect*, where this will typically include disclosure of particularly salient frames, and the fact that they are equivalent, as part of the explanation of the effect. To return to our initial example, the hope is that, by telling patients both that there is a framing effect, those patients will no longer be vulnerable to the effect. Patients who are no longer vulnerable to framing effects can then give valid consent.

Of course, this suggestion is vulnerable to a sceptical challenge: what if people persist in their framing effects even after being told that there is a framing effect? After all, evidence suggests that other biases often cannot be eliminated merely by making them explicit.[26] Several points are worth mentioning here. First, a recent study suggests that even the cursory recommendation to 'think like a scientist' reduces framing effects, so it is not overly optimistic to hope that a full explanation of the effect might be even more effective in eliminating it.[27]

Second, and relatedly, my suggestion here is not that a brief and cursory disclosure might suffice to eliminate framing effects but rather that a lengthy and thorough explanation might. Of course, we might worry that requiring a lengthy and thorough explanation of framing effects is too burdensome, but of course a similar worry arises for the problem of ignorance: when disclosing information about the differences between chemotherapy and radiation, for example, a brief and cursory disclosure is often insufficient, and only a lengthy and thorough explanation will suffice.

Third, recall my fourth and final point on the problem of ignorance, that whether we should require mere disclosure versus genuine understanding is a difficult and controversial question. Research subjects are often vulnerable to the therapeutic misconception, meaning that they persist in believing that they will benefit from research even when they are explicitly told that they will not. In such cases, maybe we should require only a

good faith effort to disclose information about the trial, rather than also requiring understanding. Similarly here: if patients persist in being affected by framing effects even after being educated and warned about them, then maybe all that we should require of medical professionals is a good faith effort to disclose, rather than also requiring understanding.

I am intentionally cautious in using the qualifier 'maybe' because my goal here is not to settle how much understanding is required for valid consent. Rather, my point is that the mere fact that people might be incapable of understanding what is told to them no more supports or exacerbates the problem of framing effects than it does the problem of ignorance. I do not want to take sides on the difficult debate of just how much understanding is required for valid consent, but the reasons for taking one side or another in that debate are the same as that in the analogous debate on framing effects. If we think that the point of consent is to obtain fully rational, autonomous decisions, then a patient who is incapable of understanding framing effects is thereby incompetent, and his consent is morally inert, in the same way that incompetence generally renders consent morally inert. However, if we think that the point of consent is to ensure that patients have a reasonable chance to opt out, then the mere good faith effort to explain the effect can validate consent, just as the effort to explain the therapeutic misconception in research cases might validate consent even for subjects who persist in that misconception. Either way, framing effects raise no new problem for the moral force of consent beyond those already raised by ignorance.

Fourth, just as there are ways to eliminate ignorance without disclosing information, so also are there ways to eliminate framing effects without disclosing meta-information. While such non-disclosure methods might be extremely atypical in the case of ignorance, because disclosure is so effective, they might turn out to be the default method of eliminating framing effects, depending on empirical facts about how best to eliminate those effects. As it happens, a recent study suggests that framing effects can be eliminated in medical decisions by having subjects fill out a simple questionnaire wherein they must list the advantages and disadvantages of their treatment options, as well as the information that was most relevant in their particular choice.[28] Stepping back, though, my main commitment is not to any particular method of debiasing, but rather just to the more general thesis that the proper response to the problem of framing effects is to debias, or in other words to eliminate those framing effects. How best to operationalize the commitment to debias is an empirical question, on which we need further research.

6. Objections

In this section I respond to three related objections to my thesis that we solve the problem of framing effects by debiasing. The sequence of objections begins with the worry that different methods of debiasing might generate different decisions. For example, a patient might consent under all frames if told to think like a scientist but dissent under all frames if told to think like an insurance agent, perhaps because, while both injunctions encourage consistency, thinking of insurance might prime patients into being more risk averse. If so, then in order to capture what the patient genuinely wants it is not sufficient that we debias; rather, we must debias *in the right way*. But what is this right way?

Notice that this objection does not claim that debiasing fails to resolve the problem of framing effects so much as insist that debiasing may lead to further problems. But this is objectionable only if (a) it actually happens, and (b) the second problem is insoluble. In this case, both (a) and (b) are dubious. On (a), whether some ineliminable feature of different debiasing techniques leads to different decisions is of course an empirical matter, but there is as yet no evidence to suggest that they do. On (b), there is every reason to think that we can solve the second bias problem in the same way that we solve the first one, by (second order) debiasing. One obvious way to do this is via disclosure (of meta-meta-information!), to the effect that different ways of debiasing themselves are biased, but there might be other techniques.

Of course, at this stage we might worry about something like an infinite regress, that anything we do might contaminate or bias our patients in some way or other. This leads to a second objection, that it is impossible to eliminate every bias. Instead of trying to debias people, even doctors should just embrace the fact that people are vulnerable to framing effects and try to use those effects for good.[29] This objection goes too far, as can be seen by comparison to informed consent: it is impossible to eliminate all the ignorance relevant to a given intervention, and yet that is no problem for the demand that consent be informed, when such a demand is appropriate. There may be a massive amount of relevant information about an intervention, as is typical in clinical care, for example, but doctors lack the time and resources to disclose all of it, just as patients lack the time and resources to absorb it all. We cannot expect patients to learn the basics of chemistry in order to understand fully the different effects of two drugs, for example, and more generally one does not have to be a board certified oncologist to make a sufficiently informed decision about cancer treatment options. Instead, we can demand only that medical professionals disclose the most important information, simplifying when necessary. Similarly in the case of framing effects, even if we are all always subject to a great number of framing effects, we can demand only that potential interveners debias their subjects from the most important such effects, simplifying when necessary.

A related third objection insists that framing effects, and indeed biases in general, are so subtle and pervasive that we cannot even hope to *identify*, let alone eliminate, them all. This again suggests that consent is typically invalid, though we may be unable to point to the source of the invalidity. Like the last one, this objection is too quick to abandon an important moral concept on flimsy evidence, as we can see from an analogously flimsy objection based on ignorance. Our ignorance is often so subtle (we are not aware of areas of ignorance) and pervasive that there is no way that even the experts can know everything that is relevant to an intervention. *No one* can know with certainty and precision how our decisions will affect the future; we are not omniscient. Yet omniscience is not required for valid consent; all that is required for valid consent is the disclosure of what little information the intervener can reasonably be expected to know, of course with the proviso that even this little bit of information might need to be ranked by priority, because of the intervener's finite capacity to explain and the subject's finite capacity to absorb. The same is true for framing effects: just as we do not require omniscience about information, so also do we not require omniscience about meta-information. We cannot reasonably expect interveners to be aware of every framing effect, so we likewise cannot reasonably expect them to attempt to debias their subjects of all such effects.

7. Conclusion

In this article, I have argued, first, that framing effects need not invalidate consent and, second, that when they do invalidate consent we can make it valid again by debiasing. My thesis comes with several caveats: I do not discuss the relative merits of different debiasing techniques; nor do I discuss what to do when good faith debiasing is ineffective (the framing effects analogue to informed consent's therapeutic misconception). Indeed, I do not even discuss how extensively we must debias. Still, it is significant progress to note that the problem framing effects pose for consent is analogous to the problem ignorance poses for consent. Thus, we can learn a lot about how to resolve the problem of framing effects by looking to extant literature on informed consent, including in areas where there is as yet no consensus, such as (a) what to do when good faith disclosure is ineffective and (b) how extensively doctors must inform their patients. My point here has not been to give a practical primer that doctors can use at the bedside to eliminate framing effects, but rather to begin discussion on how best they can do that: by noting the parallels between framing effects and ignorance, including in areas of current controversy.

I hope it is also clear, then, that we need more research. We need empirical research on when framing effects occur and when they do not, so that we can tell when to be on the lookout for them, and we need empirical research on ways to debias, so that we can eliminate those effects efficiently when doing so is morally required. We also need further conceptual research. For example, we need further conceptual research on what we might call a demarcation problem, demarcating the boundary of physician responsibility from patient responsibility. The analogous demarcation problem for informed consent is already tricky: how do we determine precisely which bits of information a doctor must disclose, if her patient's consent is valid? Similarly, how should we demarcate exactly those framing effects from which medical professionals should attempt to debias their subjects? More generally, we need further conceptual research on other forms of bias — such as priming effects — that might contaminate consent. We need some conceptual guidance for when such biases invalidate consent, and we need some conceptual guidance for what we should do about those biases. I have here tried to make a start on those larger issues, by focusing on the interactions between framing effects and the moral force of consent.

Eric Chwang, Department of Philosophy, University of Colorado at Boulder, Box 232, Boulder, CO 80309-0232, USA. chwang@colorado.edu

Acknowledgments

Thanks to Brian Talbot and an audience at the University of Colorado at Boulder's Center for Values and Social Policy for helpful discussion. Special thanks to David Boonin, Jason Hanna and Adam Hosein for helpful comments on earlier drafts.

NOTES

1 Amos Tversky & Daniel Kahneman, 'The framing of decisions and the psychology of choice', *Science* 211, 4481 (1981): 453–458.

2 Jason Hanna, 'Consent and the problem of framing effects', *Ethical Theory and Moral Practice* 14,5 (2011): 517–531.
3 Shlomo Cohen, 'Nudging and informed consent', *The American Journal of Bioethics* 13,6 (2013): 3–11. Cohen's thesis is actually much broader than I have described it. He argues that doctors may nudge their patients towards the best options in a variety of ways in addition to the use of framing effects, in the sense of 'nudge' as described in Richard Thaler & Cass Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (New York: Penguin Books, 2009).
4 For more on cognitive biases, see for example Daniel Kahneman & Amos Tversky (eds), *Choices, Values, and Frames* (New York: Cambridge University Press, 2000) and Stuart Sutherland, *Irrationality*, (London: Pinter & Martin, 2007).
5 First, my concern in this article is with moral permission, not legal permission. Second, some writers instead define valid consent by reference to a particular list of criteria, for example competence and freedom from duress. It is then an open question whether valid consent, so defined, has sufficient power to make an intervention permissible. See, for example, Franklin G. Miller & Alan Wertheimer, 'Preface to a theory of consent transactions: Beyond valid consent' in F. G. Miller & A. Wertheimer (eds) *The Ethics of Consent: Theory and Practice* (New York: Oxford University Press, 2010), pp. 79–105. In this article I instead define valid consent as conferring sufficient moral power, leaving it open just what criteria (e.g. competence and freedom from duress) are needed for validity. I prefer the latter approach because it is analogous to how validity is defined in formal logic. There, a valid argument is defined as one that preserves truth, where this leaves open whether particular inference forms are valid. (For example, double negation elimination is valid in classical but not intuitionistic logic.) One could of course instead define validity by reference to a particular list of inference forms (e.g. including *modus ponens* and *modus tollens*), leaving open whether valid inferences in this sense always preserve truth. Though I prefer my way of defining validity, nothing in this article hangs on the difference.
6 The example is from Miller & Wertheimer op. cit., p. 89.
7 Steven Joffe & Robert Truog, 'Consent to medical care: The importance of fiduciary context' in F. G. Miller & A. Wertheimer (eds) *The Ethics of Consent: Theory and Practice* (New York: Oxford University Press, 2010), pp. 347–374.
8 Though see, for example, Judith Jarvis Thomson, *The Realm of Rights* (Cambridge, MA: Harvard University Press, 1990), pp. 187–188.
9 On criteria for understanding, see Ruth Faden & Tom Beauchamp, *A History and Theory of Informed Consent* (New York: Oxford University Press, 1986), chapter 9. On the therapeutic misconception, see Paul S. Appelbaum, Loren H. Roth & Charles Lidz, 'The therapeutic misconception: Informed consent in psychiatric research', *International Journal of Law and Psychiatry* 5,3–4 (1982): 219–329. For an argument against requiring understanding in research, see Gopal Sreenivasan, 'Does informed consent to research require comprehension?', *The Lancet* 362,9400 (2003): 2016–2018.
10 This line might be especially plausible if one is attracted to a proprietary gate model of informed consent, according to which a central function of informed consent is let people make autonomous decisions and prevent them from making non-autonomous ones. For a classic exposition of this sort of view, see Faden & Beauchamp op. cit., pp. 287–288.
11 This line might be especially plausible if one is attracted to a fair transaction model of informed consent, as described in Miller & Wertheimer op. cit., pp. 102–103.
12 There is a trivial sense in which we need not always be vigilant about framing effects, even in medical contexts, because framing effects might occur only in a very narrow range of cases. For example, if a procedure has a terrible prognosis (1% survival, 99% mortality) or a perfect prognosis (100% survival, 0% mortality), it probably will not matter much whether we frame that prognosis as survival or mortality.
13 Barbara J. McNeil et al., 'On the elicitation of preferences for alternative therapies', *New England Journal of Medicine* 306,21 (1982): 1259–1262.
14 Cohen op. cit., p. 6.
15 Onora O'Neill, *Autonomy and Trust in Bioethics* (Cambridge: Cambridge University Press, 2002).
16 Cohen op. cit., p. 6. This is one of the few instances where Cohen explicitly mentions framing effects: 'It is wholly appropriate that in bioethics we expect higher standards [than mere uninformed consent], but these can be nicely accommodated by, for example, a theory like Onora O'Neill's, where the function of [informed consent] is to prevent patients from being coerced or deceived. In the standard examples of nudging, patients are obviously not coerced but surely not deceived either, even if faulty reasoning is relied upon. (Recall again

the example of steering preference by framing outcome statistics in terms of success instead of failure, which — far from deceiving — conveys accurate, nonpartial information.)'

17 One might of course say that a person's consent while under the influence of alcohol is not genuinely hers and is thereby coerced, but (a) this seems very implausible, and (b) the same move can be made for a person's consent when that person is under the influence of a framing effect.

18 See J.S. Swindell Blumenthal-Barby, 'On nudging and informed consent — four key undefended premises', *The American Journal of Bioethics* 13,6 (2013): 31–33 for a similar criticism.

19 Cohen op. cit., pp. 6–7.

20 Cohen op. cit., p. 7.

21 See, for example, Hanna op. cit., e.g., at p. 530: 'Since framing effects are common and unavoidable, the plausibility of the [proprietary gate] model [according to which consent has independent moral force] may be in jeopardy unless it can be shown that those who are vulnerable to framing can give valid consent. It does not appear that such an argument is forthcoming.'

22 Faden & Beauchamp op. cit., p. 321. This is one of the few instances of bioethicists explicitly addressing the problem that framing effects pose for consent.

23 Hanna op. cit., p. 523. Franklin G. Miller & Luke Gelinas, 'Nudging, autonomy, and valid consent: Context matters', *The American Journal of Bioethics* 13,6 (2013): 12–13, at p. 13.

24 For some empirical data suggesting that visual aids, especially pie charts, can eliminate framing effects, see Rocio Garcia-Retamero & Mirta Galesic, 'How to reduce the effect of framing on messages about health', *Journal of General Internal Medicine* 25,12 (2010): 1323–1329.

25 Another plausible worry about neutrality is that even presenting information about prognosis presupposes, non-neutrally, that life is better than death. See D. Kirklin, 'Framing, truth telling and the problem with non-directive counselling', *Journal of Medical Ethics* 33,1 (2007): 58–62.

26 For a classic example, on the resilience of the 'knew-it-all-along' effect, see Baruch Fischhoff, 'Perceived informativeness of facts', *Journal of Experimental Psychology: Human Perception and Performance* 3,2 (1977): 349–358.

27 Ayanna K. Thomas & Peter R. Millar, 'Reducing the framing effect in older and younger adults by encouraging analytic processing', *The Journals of Gerontology Series B* 67B,2 (2012): 139–149.

28 Sammy Almashat, Briat Ayotte, Barry Edelstein & Jennifer Margrett, 'Framing effect debiasing in medical decision making', *Patient Education and Counseling* 71,1 (2008): 102–107.

29 See, for example, Thom Brooks, 'Should we nudge informed consent?', *The American Journal of Bioethics* 13,6 (2013): 22–23.