Framing effects from misleading implicatures: An empirically-based case against some purported nudges

Shang Long Yeo

Abstract: Some bioethicists argue that a doctor may frame treatment options in terms of effects on survival rather than on mortality in order to influence patients to choose the better option. The debate over such framing typically assumes that the survival and mortality frames convey the same numerical information. But certain empirical findings contest this numerical equivalence assumption, demonstrating that framing effects may in fact be due to the two frames implying different information about the numerical bounds of survival and mortality rates. In this paper, I use these findings to argue that framing is presumptively wrong, because it violates a duty of proper disclosure. Along the way, I highlight morally relevant features affecting the permissibility of framing, tackle three objections, and draw some general lessons for the ethics of nudging.

Consider the case of:

> **Treatment Framing:** A doctor believes that some treatment best serves a patient's interests and wants to convince them to take it. They could describe the treatment in different ways (with more explored below):
>
> i)   "Out of 100 who took the treatment, 90 survive"
> ii)   "Out of 100 who took the treatment, 10 die"
> iii)   "Out of 100 who took the treatment, 90 survive and 10 die"

Suppose the doctor chooses i) because they know that patients are more likely consent to treatment framed in this way. This use of framing purports to be a nudge – that is, an intervention that makes it more likely that an agent will behave a certain way, primarily by triggering that agent's shallow cognitive processes, while preserving their choice set and being easily resistable.[1,2][a] A different example of a nudge involves changing how food is displayed, to influence people to choose healthier options.[2,3] Ethicists have extensively debated the ethics of nudging; similarly so with the ethics of framing treatment options.[4–12] Both proponents and opponents of treatment framing assume that options i)-iii) are interpreted by all patients as implying the same numerical information about the harms and benefits of

---

[a] The findings presented later may disqualify framing from being a nudge, because they show how framing works through rational behaviour informed by misrepresentations, rather than through shallow cognitive processes. Moreover, misrepresentation is not easily resistable.[1]

treatment[4–9] – that if a frame says that 90 out of 100 survive, patients draw the implication that 10 die.[b]

However, empirical findings contest this very assumption. In a thorough and innovative investigation of a similar framing effect, Mandel finds that the effect of frames on choice may instead be due to a lower-bound reading of outcome numbers – where subjects read "10 die" as "*at least* 10 die", while "90 survive" is read as "*at least* 90 survive".[13] That is, of the following explanations for why frames cause different choices, explanations 2 and 3 may be more plausible than previously thought:

1) The frames convey the same numerical information but emphasise different features or elicit different emotions.[7,10,11]
2) The frames convey *different* numerical information (because patients misunderstand framed disclosure as implying a lower-bound reading of outcome numbers).
3) The frames convey *different* numerical information *and* the frames emphasise different features or elicit different emotions.

If identical numerical information isn't conveyed by the different frames (ie. if 2 or 3 are true), then using framing is misleading, and it may undermine genuine informed consent.[c]

In this paper, I trace out the ethical implications of Mandel's findings for Treatment Framing. I argue that his findings show how the doctor's action in Treatment Framing is presumptively wrong because it violates a duty of proper disclosure. I start by outlining framing effects and Mandel's findings. I then examine the ethical implications of such findings, and tackle three objections to my arguments. Finally, I draw some general lessons for the ethics of nudging.

Framing effects, the numerical equivalence assumption, and Mandel's experiments

The effect of framing in terms of survival versus mortality was first popularised by the Asian Disease Problem, a hypothetical case with two policy options – an abridged version is presented below [mortality frame in brackets]:

> **Asian Disease Problem:** Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
>
> Program 1: 200 people will be saved[400 people will die].

---

[b] Director acknowledges debate about whether the frames "have identical semantic content" but does not pursue this.[8]

[c] Unjustified differential emphasis could also be misleading and hence undermine genuine informed consent.[12] For the purposes of this paper, however, I focus only on differential numerical information.

Program 2: 1/3 probability that 600 people will be saved [nobody will die], and 2/3 probability that no people will be saved [600 people will die].

Tversky and Kahneman find that when the programs are framed in terms of saving lives, a majority chose Program 1; whereas when they are framed in terms of deaths, the majority chose Program 2. They contend that this is irrational because the frames are equivalent.[14] They explain the pattern of choice with loss aversion: the saving frame sets the reference point as the scenario where no one lives, so all options involve potential gains; the dying frame sets the reference point as all being alive, so the options involve potential losses. Subsequent research finds that such framing – which highlights one part of an outcome while leaving out its complement with a different valence – also influences choice of medical treatment for a single individual. When subjects chose between two cancer treatments – surgery versus radiation therapy – in a hypothetical scenario, their choices differed by more than 20 percentage points depending on whether a survival or mortality frame was used.[15]

Mandel observes that the assumption of equivalent frames is implicitly justified through a proof by arithmetic: if survival and death are the only possible outcomes, then for n patients, saying that x patients survive entails that n-x patients die.[13] But he notes that this argument only works if the outcome numbers are given an exact reading – that is, if "x patients survive" is read as "exactly x patients survive". He investigates whether people in fact give outcome numbers an exact reading, and how this affects the framing effects observed. First, he finds direct evidence for a lower-bound reading of outcome numbers instead (N=147, experiment 3). For outcomes that weren't fully specified with both complementary parts – e.g. when an outcome is only described as "400 die" without saying 200 survive – 64% of subjects said they gave outcome numbers a lower-bound reading, while 30% gave them with an exact reading.[13] (He also finds, however, that narrative context could modulate and discourage this lower-bound reading.[d])

Secondly, Mandel finds that the framing effect counterfactually depends on whether "at least" or "exactly" are explicitly used to modify the outcome numbers – these correspond to the lower-bound and exact readings respectively (N=228, experiment 2). When "at least" modifies outcome numbers of Program 1, a majority of subjects (67.7%) chose inconsistently across the two frames in the way observed in the standard framing effect. When "exactly" modifies the outcome numbers, however, a majority (73.3%) choose consistently between the two frames – the framing effect disappears.[13]

Mandel's findings indicate that a majority attribute lower-bound readings to outcome numbers, and that this reading is at least a partial cause of framing effects observed in the Asian Disease Problem. His findings are relevant to treatment framing too, given that it works

---

[d] In a different experiment (N=97) where the choice is described as arising from scarce medical resources, Mandel finds that most subjects adopted an upper-bound reading in the survival frame.[13]

in the same way as the Asian Disease Problem. Indeed, Mandel mentions treatment framing, saying that if mortality and survival rates "tend to be interpreted as lower bounds, then clearly the alternative frames would not convey the same information to patients."[13] I turn now to the ethical implications of this.

The ethical implications of the lower-bound reading

Suppose treatment framing works because of the lower-bound reading. Saying "x patients die" without specifying its complement encourages the lower-bound reading that *at least* x patients die, so this misleads many patients about the range of possible mortality rates justified by the evidence. This frame does not convey false information, however, since claiming that at least x patients die is consistent with exactly x patients dying. But it reliably implies to a majority (64%, according to Mandel's findings) that other values are possible or likely given the evidence, when in fact they are not.

This makes treatment framing presumptively wrong, because it violates a duty to disclose medically relevant truths to one's patient[7,16] – the truths being precise estimates of a treatment's survival and mortality rates, as contrasted with the misleading estimates implied by framing. Patients have a reasonable expectation that doctors communicate these rates precisely; these are one of the most medically relevant pieces of information being disclosed. This stands in contrast to other information whose relevance may be disputed – for instance, information about misattributed paternity that is inadvertently discovered in routine screening for an organ transplant.[16] The misleading implicature also stands in contrast to other effects of a frame which may or may not be morally justifiable – such as how much emotion is cultivated or recommended, or what reasons for action are provided.[7,10–12,17] Simkulet argues that for any treatment, there is an appropriate level of anxiety, fear, or hope a patient should have about it – and if using a frame reliably cultivates an inappropriate level of emotions, this is misleading, precludes understanding, and does not count as adequate disclosure.[7] Cohen objects that this relies on an overly strong notion of understanding that includes appraisal of an option's choiceworthiness – he argues instead that using a frame to elicit emotions does not violate adequate disclosure, because it still conveys accurate data about treatment options and their expected risks and benefits.[11] My argument parallels Simkulet's but uses Mandel's findings to show how frames convey misleading data about risks and benefits, separate from inappropriate levels of emotion. This prevents patients from fully understanding what they are consenting to, which could negate informed consent.[18]

How misleading (and hence how wrong) treatment framing is will depend on empirical details about the lower-bound reading. To illustrate, consider the mortality frame "10 die out of 100" and suppose the lower-bound reading distributes probability equally to all outcomes where at least 10 die (Fig 1b), rather than the true 100% probability to exactly 10 dying (Fig

1a). The misleadingness of the reading depends on how much probability it attributes to outcomes other than the truth, and how far these are from the truth. This equal-probability lower-bound reading is thus quite misleading, since it attributes significant probability to outcomes far from exactly 10 dying.
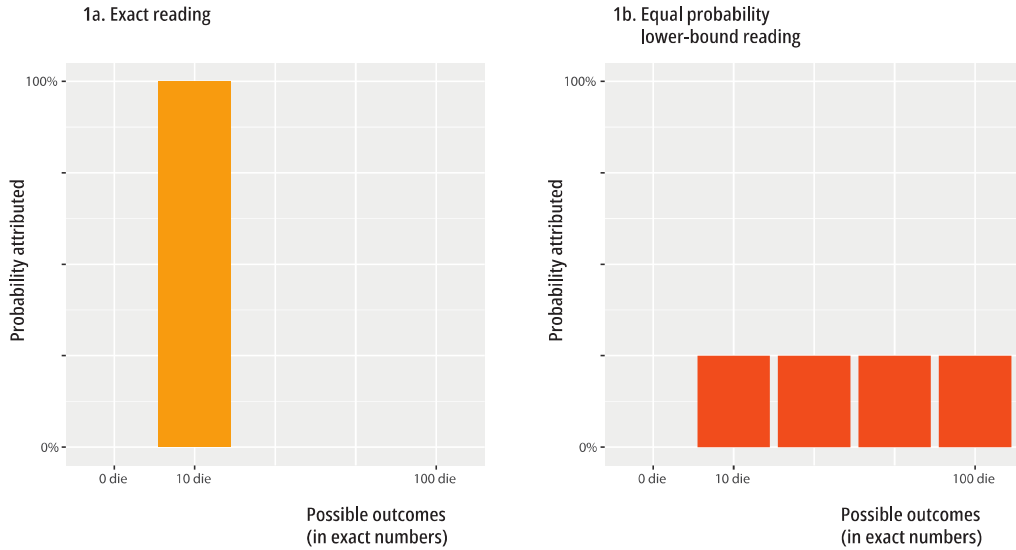

FIG 1: EXACT READING AND EQUAL-PROBABILITY LOWER-BOUND READING

But perhaps the lower-bound reading only shifts probabilities to outcomes near the framed quantity (Fig 2a), or it attributes probabilities which decay for outcomes further from the framed quantity (Fig 2b). These are less misleading than the equal-probability lower-bound reading, because they attribute probabilities concentrated nearer the truth. In general, the more the lower-bound reading disperses probability to other outcomes, and the more extreme these outcomes are, the more misleading it is. More empirical work is needed to determine how this reading works, which in turn influences the permissibility of framing.
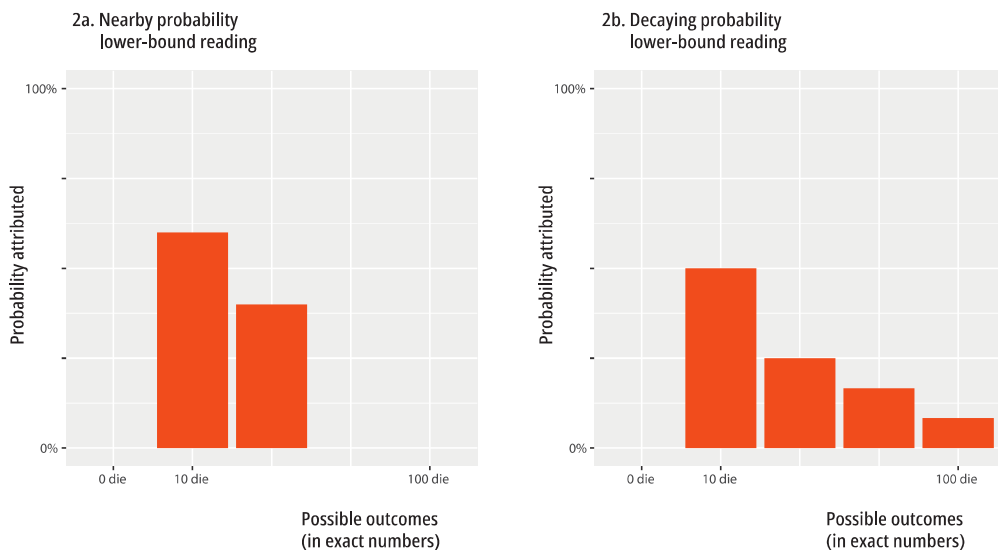

FIG 2: NEARBY PROBABILITY AND DECAYING PROBABILITY LOWER-BOUND READINGS

The mortality and survival frames could also be *differentially* misleading. Suppose the true outcome is that exactly 10 die and 90 survive, and suppose the lower-bound reading attributes a decaying probability beyond the framed quantity. Fig 3 compares the probability assignments, made to equivalent exact outcomes, by the two frames. As we can see, the mortality frame (left) is more misleading than the survival frame (right) – the mortality frame disperses more probability to more extreme outcomes, by virtue of there being more such outcomes to distribute to. Generally, the more extreme the framed quantities are, and the more the lower-bound reading disperses probability, the more the two frames will be differentially misleading. Here, the decaying-probability lower-bound reading imbues a previously irrelevant factor – the value of the framed quantity in relation to its maximum value – with moral relevance, given its impact on misleadingness.
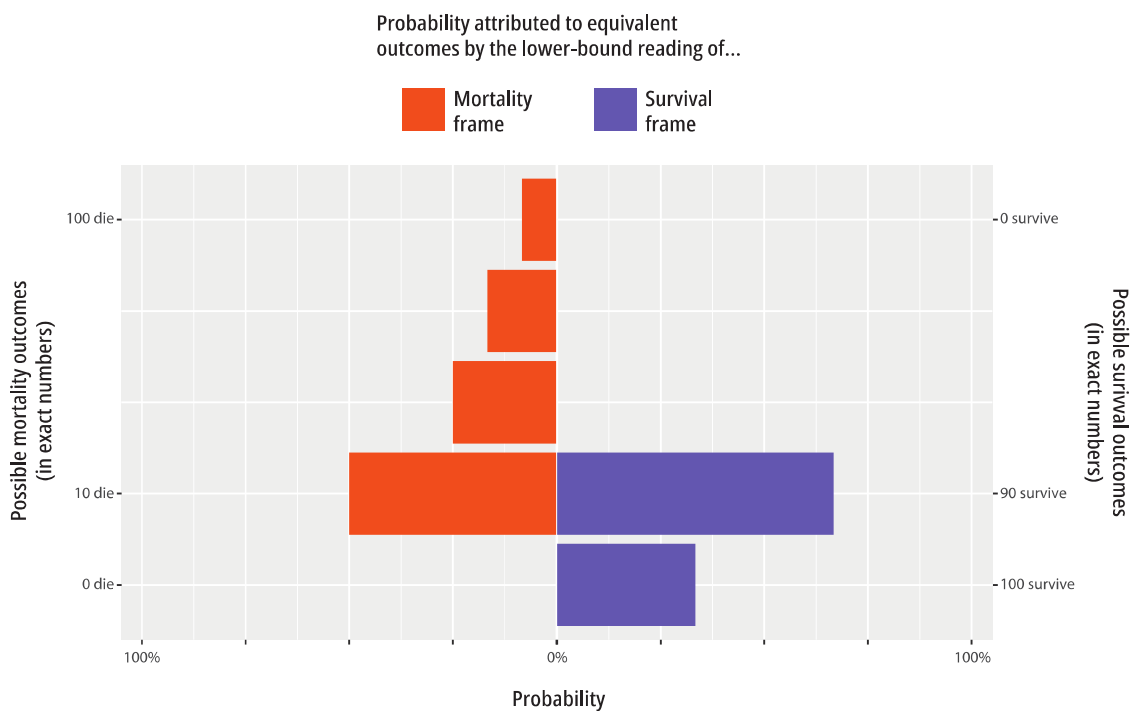

FIG 3: COMPARISON OF DECAYING PROBABILITY LOWER-BOUND READINGS OF MORTALITY AND SURVIVAL FRAMES

An asymmetry in the reasons justifying each frame could also arise from an asymmetry in the value (rather than probability) of potential outcomes. For instance, if treatment has equal chance of creating either mild benefits or great harms, while not treating is certain to create a neutral outcome, then it is more justifiable to use a frame to chill the enthusiasm of the patient. Still, using framing here is somewhat misleading, because it uses the wrong kind of reason – concerning probability of outcomes, rather than their value – to influence the patient.

Generally, treatment framing may be all things considered permissible if countervailing considerations prevail. I now draw on Sokol's useful flowchart for determining the permissibility of deception,[19] to argue that many possible justifications for deception do not apply to treatment framing, and that its permissibility faces two further challenges. (Note, however, that framing may not count as deception because its creation of false belief is not intentional, and it does not do so through convincing the patient. Framing may be manipulation instead.[20] Still, the justifications below similarly apply to manipulation, so I set aside the question of how best to classify framing.)

Start with possible justifications. First, it's clearly not the case that the patient is emotionally or cognitively incapable of deciding – since treatment framing is offered in contexts where the patient can choose between treatment options. Second, framing is not justified on grounds of preserving hope or reducing anxiety – which apply more to the disclosure of adverse diagnosis and prognosis. Here, framing aims primarily to get the patient to choose some treatment. Third, framing isn't typically offered as a temporary deception that is unmasked later, though this is an intriguing possibility worth considering. Fourth, perhaps the patient doesn't want the information they are being misled about (precise information about survival and mortality rates), but it's hard to see how proponents could reliably ascertain this before intervening. This leaves the prevention of great physical or psychological harm, and the enhancement of autonomy in the long run, which I concede may serve as potential justifications.

Even when such justifications are available, two more challenges arise: first, framing has a limited impact on probability of success: in the study of treatment framing, switching from a survival to mortality frame increased acceptance of a treatment option from 18% to 47%, an increase of 29 percentage points.[15] While this is considerable, recognising that framing isn't always efficacious reduces the justification for misleading patients about medically relevant information. Secondly, there may be other non-deceptive options that get the patient to choose the treatment just as effectively. For instance, if the doctor is permitted to make rational arguments and advocate for what they believe is the best course of action given the patient's values,[21] these may be just as effective as framing but not similarly misleading. The availability of such an option makes treatment framing impermissible. I conclude that when we consider the duty to disclose medically relevant information, treatment framing is permissible only in very specific circumstances: it must be required to prevent great harm or enhance autonomy greatly, it must be reasonably effective, and there must be no similarly effective non-deceptive option available.

Three objections

I now answer three potential objections. First, it might be objected that my arguments impose an overly demanding duty on doctors to avoid misleading implicatures from framing. In

response, I contend that avoiding this specific implicature imposes a minimal demand: to either use "exactly" when presenting outcome numbers in a single frame ("exactly 10 die") or to use a mixed frame specifying both parts of the outcome ("90 survive, 10 die"). Mandel's findings give us an independent, empirically-based reason for using these frames (contrary to ethicists who deny that there are such reasons[4,6]) – these frames convey the least misleading information, given the lower-bound reading. Even if the potential misleadingness is only probable and not certain, doctors are required to guard against it if the cost of doing so is adding a mere word or two. In general, doctors are required to avoid causing their patients to draw reasonable but misleading implicatures. They are not, however, required to disclose in a way that causes every particular patient to believe the truth – this is too demanding, since implicatures are not entirely within a speaker's control.

But what about other framing effects? Some argue, for instance, that even the mixed frame may be problematic, because patients may be influenced by the order in which different parts of the outcome are presented (ie. whether survival or mortality numbers are specified first).[4,6,22] But there is empirical evidence against similar ordering effects, which defeats this worry.[23] In my view, we should treat frames as innocent until proven guilty, and only rectify them when we get positive evidence of reliably misleading implicatures (as in Mandel's findings) and when such rectification does not impose overly large costs.

Secondly, one could pose empirical objections: some failed to replicate Mandel's findings that framing effects depend on lower-bound readings,[24–26] so the lower-bound reading may not be necessary for such effects (though it is sufficient). Chick et al. contend that Mandel's direct evidence – from asking subjects which reading they attributed – is called into question by potential demand effects, where subjects are motivated to choose the reading that makes their previous choices most reasonable.[25] In response, I contend that the ethical case against treatment framing remains despite such empirical uncertainty. We should be wary of framing so long as we have a reasonable suspicion of misleading implicatures. And I believe Mandel's direct evidence establishes this reasonable suspicion, despite being attenuated by potential demand effects.

At a minimum, we should modify outcome numbers with "exactly" to refrain from misleading patients. This leaves open the possibility that framing effects from other causes – such as loss aversion or eliciting different emotions – remain and may be permissibly used. My arguments only tell against framing effects to the extent that they are caused by misleading numerical implicatures. Notice, however, that if the lower-bound reading is indeed necessary for the frame's beneficial effects, then these would be negated by the "exactly" modifier.

Thirdly, one could define frames in a way that avoids my criticism. For instance, Director argues that "by definition, framing effects cannot be deceptive" because the frames are defined as "logically equivalent ways of stating the same information, just with a different emphasis."[8] In response, I contend that this definition is unhelpful for normative guidance.

It remains an important question whether doctors may perform the communicative act of framing treatment by mentioning only one part of an outcome without its complement – the relevant communicative act being specified only in terms of its lexical properties, and not its downstream effects. If this act creates misleading downstream implicatures, they cannot be defined out of existence – they must be acknowledged, for as I have argued, they have important normative implications for the permissibility of the communicative act.

Conclusion

Mandel's findings reveal that some framing effects may be caused by numerical implicature. This vindicates Simkulet's cautionary claim that if two frames lead to different behaviours, this is defeasible evidence that they convey different information (though the findings work through a different route to what Simkulet imagined).[7] I've argued that the findings show how treatment framing is presumptively wrong because it violates duties of proper disclosure.

I draw three general lessons for the ethics of nudging. First, ethicists shouldn't simply accept dominant assumptions and explanations about purported nudges – there may be alternative, empirically-grounded explanations with significant implications. If Mandel's findings and explanation are accurate, treatment framing may not even be a nudge, because it works through misrepresentation rather than loss aversion. Furthermore, his explanation implies significant moral downsides of treatment framing. A reasonable credence in these alternative explanations may be sufficient grounds for forbidding framing, or at least for shifting to different forms of it (for instance, using an "exactly" modifier when presenting only one part of an outcome without its complement). Secondly, we shouldn't speculate from the armchair about these alternative explanations. Mandel's findings illustrate a method for how we can empirically ascertain the implicatures created by treatment framing, which can in turn inform our investigation into its permissibility. Thirdly, my arguments exemplify a highly piecemeal approach to evaluating (purported) nudges. The problems with misleading numerical implicatures only arise for numerical framing of the kind outlined, not for other types of framing – like loss/gain framing that does not use numbers (e.g. saying that a test for disease "can save your life" versus that foregoing it "can cost you your life"),[27] or framing that uses "cuddly language designed not to upset"[7]. With regards to numerical treatment framing, at least, I've argued that doctors are often forbidden from using these – for such framing reliably misleads about medically relevant information.

References

1   Saghai Y. Salvaging the concept of nudge. *Journal of Medical Ethics*. 2013;39:487–93.

2   Thaler RH, Sunstein CR. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press 2008.

3   Rozin P, Scott SE, Dingley M, *et al.* Nudge to nobesity I: Minor changes in accessibility decrease food intake. 2011.

4   Hanna J. Consent and the Problem of Framing Effects. *Ethical Theory and Moral Practice*. 2011;14:517–31.

5   Cohen S. Nudging and informed consent. *Am J Bioeth*. 2013;13:3–11.

6   Chwang E. Consent's Been Framed: When Framing Effects Invalidate Consent and How to Validate It Again. *Journal of Applied Philosophy*. 2016;33:270–85.

7   Simkulet W. Nudging, informed consent and bullshit. *J Med Ethics*. 2018;44:536–42.

8   Director S. Framing Effects Do Not Undermine Consent. *Ethic Theory Moral Prac*. Published Online First: 22 August 2023. doi: 10.1007/s10677-023-10412-1

9   Blumenthal-Barby J, Ubel PA. Truth be told: not all nudging is bullshit. *Journal of Medical Ethics*. 2018;44:547–547.

10   Ploug T. Physicians' framing and recommendations. Are they nudging? And do they violate the requirements of informed consent? *Journal of Medical Ethics*. 2018;44:543–4.

11   Cohen S. Truthful nudging. *Journal of Medical Ethics*. 2018;44:545–6.

12   Simkulet W. At odds with the truth. *Journal of Medical Ethics*. 2018;44:548–50.

13   Mandel DR. Do framing effects reveal irrational choice? *J Exp Psychol Gen*. 2014;143:1185–98.

14   Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science*. 1981;211:453–8.

15   McNeil BJ, Pauker SG, Tversky A. On the framing of medical decisions. *Decision making: Descriptive, normative, and prescriptive interactions*. New York, NY, US: Cambridge University Press 1988:562–8. https://doi.org/10.1017/CBO9780511598951.028

16   Sokol DK. Truth-telling in the doctor-patient relationship: a case analysis. *Clinical Ethics*. 2006;1:130–4.

17 Blumenthal-Barby JS, Burroughs H. Seeking better health care outcomes: the ethics of using the "nudge." *Am J Bioeth*. 2012;12:1–10.

18 Millum J, Bromwich D. Informed Consent: What Must Be Disclosed and What Must Be Understood? *Am J Bioeth*. 2021;21:46–58.

19 Sokol DK. Can deceiving patients be morally acceptable? *BMJ*. 2007;334:984–6.

20 Cohen S. Manipulation and Deception. *Australasian Journal of Philosophy*. 2018;96:483–97.

21 Savulescu J. Rational non-interventional paternalism: why doctors ought to make judgments of what is best for their patients. *J Med Ethics*. 1995;21:327–31.

22 Miller FG, Gelinas L. Nudging, Autonomy, and Valid Consent: Context Matters. *The American Journal of Bioethics*. 2013;13:12–3.

23 Peters E, Levin IP. Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision Making*. 2008;3:435–48.

24 Fisher SA. Meaning and framing: the semantic implications of psychological framing effects. *Inquiry*. 2022;65:967–90.

25 Chick CF, Reyna VF, Corbin JC. Framing Effects are Robust to Linguistic Disambiguation: A Critical Test of Contemporary Theory. *J Exp Psychol Learn Mem Cogn*. 2016;42:238–56.

26 Simmons J, Nelson L. "Exactly": The Most Famous Framing Effect Is Robust To Precise Wording. Data Colada. 2013. https://datacolada.org/11 (accessed 16 March 2024)

27 Banks SM, Salovey P, Greener S, *et al.* The effects of message framing on mammography utilization. *Health Psychol*. 1995;14:178–84.