Spring 2022

Philosophy, Psychology, and the Ethics of Consent

Joanna Demaree-Cotton
Yale University Graduate School of Arts and Sciences, jodemaree@hotmail.com

Abstract

Philosophy, Psychology, and the Ethics of Consent

Joanna Demaree-Cotton

2022

Consent is morally transformative and suffuses our everyday moral and social lives. Valid consent makes the difference between permissible sex and rape; between a medical exam and assault; between entering a person's home and trespass; between an economic transaction and theft; between contact sports and physical attacks; between the sharing of information and invasions of privacy.

Moral philosophers writing on consent take themselves to capture and illuminate these ordinary practices, while claiming to do justice to commonsense claims about the circumstances under which consent is valid. Specifically, they purport to explain the way that ordinary consent can be morally transformative by offering theories of the kind of autonomous agency that is thought to underlie these ethical transformations. Moreover, these theories are then intended to capture what are taken to be commonsense distinctions between cases of valid consent, on the one hand, and invalid consent, on the other.

Yet this philosophizing is normally done with little or no empirical investigation of the nature of this practice or the nature of the autonomous agency that is claimed to be crucial in justifying it. How do real agents in fact come to give consent? In what sense can their

decisions be said to be autonomous? And to what extent do moral theories cohere with how consent is in fact conceived in ordinary reasoning?

In this dissertation, I bring moral philosophy together with careful examination of relevant experimental psychology in order to illuminate the ethics of consent.

In the first part of the dissertation, I show that we can reconcile the philosophical idea that consent must be autonomous with empirical findings that show that real decisions to consent are variable and influenced by trivial factors, without giving in to sweeping skepticism or revisionism about ordinary practices. Specifically, I examine challenges to the standard view of consent based on findings that decisions to consent are subject to framing effects. I argue that such challenges are best understood in terms of a claim about the extent to which frame-dependent consent decisions are autonomous. I propose a model of decision-making that captures how suboptimal decisions can be sufficiently autonomous for valid consent. Subsequently, I analyze arguments for the view that framing effects threaten the sufficient autonomy of consent. On philosophical grounds, I argue that being dependent on framing does not entail that consent is invalid. Furthermore, drawing on empirical work, I argue that frame-dependence does not make it likely that consent is invalid. Instead, variability in an agent's decision to consent due to the influence of framing is compatible with sufficiently autonomous decision-making based on a reasonable weighting of the agent's own values.

In the second part of the dissertation, I report three studies showing that the philosophical idea that autonomy is importantly linked to the validity of consent is also reflected in the folk concept, but that the kind of autonomy presupposed by the folk concept is much less demanding than many philosophical treatments take it to be. Specifically, while

philosophical accounts assume that consenters must exercise their autonomy in order to give valid consent (the "Exercised Capacity" view), the folk concept requires only that the consenter possesses the capacity to decide autonomously (the "Mere Capacity" view), even if they do not exercise this capacity.

Study 1 shows that when agents lack autonomous decision-making capacities, participants are less likely to view their consent as valid; however, failing to exercise this capacity and deciding in a nonautonomous way does not reduce consent judgments. Study 2 finds that specific and concrete incapacities reduce judgments of valid consent, but failing to exercise these specific capacities does not, even when the consenter makes an irrational and inauthentic decision. Finally, Study 3 shows that the effect of autonomy on judgments of valid consent carries important downstream consequences for moral reasoning about the rights and obligations of third parties. Overall, these findings suggest that laypeople embrace a normative, domain-general concept of valid consent that depends consistently on the possession of autonomous capacities, but not on the exercise of these capacities. Autonomous decisions and autonomous capacities thus play divergent roles in moral reasoning about consent interactions: while the former appears relevant for assessing the wrongfulness of consented-to acts, the latter plays a role in whether consent is regarded as authoritative and therefore as transforming moral rights.

Finally, I argue that the Mere Capacity view not only coheres better with the folk concept, but has independent philosophical promise for an account of the ethics of consent.

Philosophy, Psychology, and the Ethics of Consent

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

In Candidacy for the Degree of

Doctor of Philosophy

by

Joanna Demaree-Cotton

Dissertation Directors: Shelly Kagan and Joshua Knobe

May 2022

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support, advice, and encouragement of advisors, colleagues, friends, and family.

First and foremost, I would like to thank my dissertation advisors, Shelly Kagan and Joshua Knobe, for their continuous dedication, and for innumerable discussions of innumerable drafts. I have learned much from both of them; and their critical perspectives, without a doubt, improved my work and my thinking as a philosopher. I would also like to thank Daniel Greco and Gideon Yaffe as members of my dissertation committee for their support and for their feedback on my work.

I am deeply grateful for the collaboration of Roseanna Sommers. Her enthusiasm, encouragement and sharp intellect were invaluable, imparting in me the confidence and energy I needed to pursue my ideas.

In addition, I would like to thank friends and colleagues who created the enriching and supportive environment that was needed to learn and to research, especially: Chris Blake-Turner; Joanna Blake-Turner; Karamvir Chadha; Jennifer Daigle; Hugo Havranek; Reier Helle; Allison Piñeros-Glasscock; Juan Piñeros-Glasscock; and members of Yale's Philosophical Psychology lab. I must also single out Brian Earp whom I have to thank for enthusiastic support of my research.

I am grateful for the guidance of Regina Rini, who set an example for me, and whose continued and generous mentorship over many years has helped to guide me along my path in philosophy.

For helpful comments and discussion on aspects of this work, as well as those already mentioned, I would like to thank Mario Attie; April Bailey; Michael Dunn; Matthew Lindauer; Jonathan Pugh; Julia Rubin; Mark Sheehan; Kevin Tobia; Stella Villarmea; the Chicago/Michigan Psychology and Lab Studies Group; and audiences at the Yale-Oxford BioXphi Conference, the European Online Experimental Philosophy Conference, Michigan's Mind and Moral Psychology Research Group, the Group for Empirical Approaches to Morality and Society, and the Ethox Centre and Wellcome Centre for Ethics and Humanities Seminar Series.

This research received funding support from the Wachtell Program in Behavioral Law, Economics & Finance at the University of Chicago Law School, and from the Institute for Citizens & Scholars.

Finally, I want to thank Maxime Lepoutre, for his partnership, support, and unwavering belief in me.

INTRODUCTION

Consent is morally transformative and suffuses our everyday moral and social lives. Valid consent makes the difference between permissible sex and rape; between a medical exam and assault; between entering a person's home and trespass; between an economic transaction and theft; between contact sports and physical attacks; between the sharing of information and invasions of privacy. We need valid consent to borrow things, to exchange money, to perform medical procedures, to cut someone's hair, and to enter into legally binding contracts.

Moral philosophers writing on consent take themselves to capture and illuminate these ordinary practices, while claiming to do justice to commonsense claims about the circumstances under which consent is valid. Specifically, they purport to explain the way that ordinary consent can be morally transformative by offering theories of the kind of autonomous agency that is thought to underlie these ethical transformations. Moreover, these theories are then intended to capture what are taken to be commonsense distinctions between cases of valid consent, on the one hand, and invalid consent, on the other.

Yet, despite purporting to capture the ethical qualities of everyday consent transactions between ordinary decision-makers, this philosophizing is normally done with little or no empirical investigation of the nature of this practice or the nature of the autonomous agency that is claimed to be crucial in justifying it. How do real agents in fact come to give consent? In what sense can their decisions be said to be autonomous? And to what extent do moral theories cohere with how consent is in fact conceived in ordinary reasoning?

In this dissertation, I bring moral philosophy together with careful examination of relevant experimental psychology in order to illuminate the ethics of consent. In doing so, I aim to

illustrate how a neglect of empirical work can lead moral philosophy astray. One way it can do so is by presupposing an overly idealized picture of the autonomous agency of ordinary consenters. This can lead us to misunderstand the nature of consent itself, and to misdiagnose the source of its normative power. At the other extreme, a neglect or misunderstanding of empirical psychology has led some philosophers to highly skeptical and pessimistic conclusions about consent, underwritten by, as I will argue, the mistaken conviction that the normative requirements of consent cannot be reconciled with the flawed nature of ordinary decision-making.

Ultimately, I suggest that examining relevant empirical evidence, together with philosophical scrutiny, can help us better understand the normative power of ordinary consent. In the first part of the dissertation, I will show that we can reconcile the philosophical idea that consent must be autonomous with empirical findings that show that real decisions to consent are variable and influenced by trivial factors, without giving in to sweeping skepticism or revisionism about ordinary practices. In the second part of the dissertation, I will show that the philosophical idea that autonomy is importantly linked to the validity of consent is also reflected in the folk concept, but that the kind of autonomy presupposed by the folk concept is much less demanding than many philosophical treatments take it to be. I will argue that an alternative philosophical theory that coheres better with the folk concept has philosophical promise; as such, this view ought to be taken seriously. As philosophers, we need not, and ought not, simply defer to ordinary practice. But we must have an accurate grasp of the extent to which our theories imply a revision of ordinary practices, and provide adequate justification for any such revision. Moreover, we cannot successfully understand and illuminate the practice of consent unless we have an adequate understanding of the kind of decision-making and concepts that this practice ordinarily involves.

FRAMING EFFECTS, SUBOPTIMAL AGENTS, AND THE STANDARD VIEW OF VALID CONSENT

1. Introduction

Imagine the following situation:

> **Framing-Induced Surgery:** A patient has been diagnosed with lung cancer. Their doctor explains to them that surgery is one treatment option. He explains what this would involve and the likely impact it would have on the cancer. He also explains that all surgeries carry some risk, and states that 80% of patients survive surgery of this kind. The patient says he wants to have the surgery, signs the relevant consent forms, and, eventually, he is taken in for the surgery to be performed. However, had the doctor presented this patient with the *mortality rate* of the surgery—had the doctor stated that 20% of patients do *not* survive surgery of this kind—the patient would not have consented to surgery, and would have ended up pursuing a different treatment option, such as radiotherapy.[1]

Empirical evidence suggests that such a situation is not purely imaginary, and that, in fact, many real-life decisions might be affected by the way information is worded or "framed", without the person's awareness. This seems, on the face of it, rather troubling. We might have hoped that whether someone gives consent to an invasive and momentous procedure did not depend on something so trivial as a seemingly arbitrary choice between two equivalent ways of wording a piece of information. And that it does depend on something so trivial might make one worry about the quality of that consent.

This paper takes up the question of whether such worries are ultimately justified. More precisely, I take up the question of whether or not evidence that choices can be subject to framing effects undermines the standard view of consent, according to which (1) valid

[1] This vignette is similar to the one laid out in Hanna, 2011, p.520., and is loosely based on the vignette used in the empirical studies of McNeil et al., 1992.

consent is normally required for things like surgical interventions to be morally permissible, and according to which (2) the consent of competent adults normally succeeds in constituting valid consent in such contexts. If we are right to be troubled by cases like Framing-Induced Surgery, we may be faced with revisionary implications for our understanding of the ethics of consent.

2. Framing Effects

Framing-Induced Surgery describes a case in which someone's consent depends on wording in a way that seems troubling. Of course, dependence on wording is not troubling per se. After all, different wording can convey importantly different information! If a doctor chooses to describe a side-effect of some treatment as "not lethal but incredibly painful" rather than "minor", it should neither be worrying nor surprising if I am less likely to consent to that treatment as a result: these descriptions are importantly non-equivalent in their meaning, and the former description gives me information about negative consequences (incredible pain) that the second does not.

This feature appears to be missing in Framing-Induced Surgery, in which a choice depended on a wording in a more specific way: it depended on whether the risk of surgery was framed in terms of survival rates or mortality rates. But these seem to be equivalent ways of conveying the very same information, at least if we grant that the agent shares some basic background assumptions (for instance, that 'not dying' implies 'surviving', and

vice versa). So what is troubling about that case has to do with an apparent dependence on *mere* wording independently of the information conveyed by that wording.[2]

This is a known as a 'framing effect'. A piece of informational content can be presented in multiple ways, and a frame can be thought of as a way of presenting some information—the *form* in which some content it is presented. For instance, risk of death can be presented in the form of mortality or survival rates; information about side-effects can be presented orally or on paper. Framing effects occur when decisions are affected by putatively informationally equivalent variations in how the same choice is presented—when frames affect choice even though they do not affect meaning or informational content.[3] More precisely, we can define a framing effect as follows:

> **Framing effects:** a framing effect occurs when a choice depends on whether some piece of information is framed in terms of Frame A or Frame B, where Frame A and Frame B differ in form but do not differ in informational content.[4]

[2] We will see in Section 8 that some theorists have challenged the idea that changes in wording like this are, in fact, informationally neutral, at least in natural conversational contexts.

[3] Tversky and Kahneman (e.g., 1981). The extent to which such frames are in fact equivalent will be taken up later in this paper. Note that I am using the term "framing effects" more broadly here than is typically done in the behavioral economics and psychology literature, where it is used to refer to wording effects specifically. But for our philosophical purposes, the distinction between framing due to wording or some other seemingly choice-irrelevant aspect of the choice situation does not seem important.

[4] Other fields use the term "framing effects" differently. Political science research distinguishes between so-called "equivalency framing", which is the same kind of framing I am discussing, and "issue framing". Much of political science and media studies uses the generic term "framing effects" to mean issue framing. So-called issue framing involves emphasizing or highlighting different aspects of some option or situation—for instance, a politician or newspaper headline might "frame" a social welfare bill that cuts benefits in terms of "forcing people into jobs" or in terms of "creating poverty" (Slothuus, 2008). This kind of "framing" is not our current topic.

There are many types of framing effects, so defined. Much research has focused on so-called "attribute framing", and specifically on ways that some attribute or feature of an option can be framed either positively or negatively. Still, in the empirical literature, the terms "positive" and "negative" are used in a relatively broad, inclusive sense that may refer to a numerical contrast, an evaluative contrast, or some other type of contrast, such as the presence (positive) or absence (negative) of an event or action, or the satisfaction (positive) or not (negative) of a goal. For instance, numerical risk can be presented positively, in terms of the chance of some event obtaining, or negatively, in terms of the chance of some event being avoided. Similarly, some attributes can be framed positively or negatively depending on whether the description refers explicitly to an evaluatively positive or evaluatively negative attribute, as when mortality risk can be framed as 'x% chance of survival' or '100-x% chance of death', or when a side-effect of a medication on weight can be framed in terms of making you 'thinner' or 'less fat'.

Another example concerns the framing of actions and their consequences, which can be framed in terms of the positive consequences of doing an action ('if you have check-ups, we are able to detect cancer early') or the negative consequences of foregoing an action ('if you don't have check-ups, we are not able to detect cancer early'); this is sometimes known as "goal framing" (since it concerns the health-related goals that can be achieved by undertaking various actions) or "gain/loss framing" (referring to what is gained or equivalently lost as a consequence of performing such actions). An attentive reader may note that "goal framing" confounds two components which can also be manipulated independently: focus on the possible action or the possible inaction, and the positive/negative framing of the consequences. Moreover, compared to positive/negative attribute framing, positive/negative goal frames that switch between action and inaction require further assumptions of conversational pragmatics in order to render their content

informationally equivalent. To see this, notice that it is not coherent to affirm a positive attribute frame while disaffirming the corresponding negative attribute frame, as in, "the procedure has a 10% mortality rate, but does not have a 90% survival rate" (attribute framing). By contrast, it *is* logically coherent to say, "if you have check-ups, we are able to detect cancer early; but if you don't have check-ups, we are (also) able to detect cancer early" (goal framing). Still, under normal circumstances it would be conversationally infelicitous to believe this conjunct and yet to merely say "if you have check-ups, we are able to detect cancer early". In English, a speaker who says "if you have check-ups, we are able to detect cancer early" is normally taken to mean "*only* if you have check-ups are we able to detect cancer early".

In any case, when we say that a person's choice was influenced by a framing effect, we do not in general mean to say that they actually changed their choice—that they made one choice, were subjected to a new way of framing their options, and then made a different choice. Rather, we normally mean that someone's choice *would* have been different in some way *had* their options been differently framed. To be subject to a framing effect, then, is to be affected by framing such that one is *disposed* to have chosen differently than how one actually chooses given some difference in framing. One might be disposed to choose differently in the weak sense that one is disposed to choose the very same option but in a different way or with different attitudes—for instance, if framing affects the speed or confidence with which one chooses a particular option, but does not affect *whether* one chooses that option. This weaker sense of a framing effect will be set aside for the rest of this paper, for the most part. Instead, I'll focus on cases where one is disposed to choose *a different option altogether* in response to a different frame—for instance, a situation in which

one would not have consented to some procedure had the frame been different—since these cases are the ones that are most likely to pose problems for the quality of consent.[5]

There is robust evidence that different ways of framing the same attribute of some option affects judgment and choice in a range of contexts.[6] In medical decision-making, for instance, it has been shown that whether the risk involved in some medical procedure is framed in terms of its *x% survival rate* or the equivalent *100-x% mortality rate* affects how that procedure is evaluated, with mortality rates leading to more negative evaluations and lower likelihoods of the option being chosen.[7] For example, when faced with a choice between surgery and radiation therapy subjects were significantly less likely to express a preference for surgery when presented with the mortality rates of the procedures rather than the corresponding survival rates. The classic version of this study was done by McNeil and colleagues (1982), but a number of other studies have since found similar effects of survival/mortality framing on hypothetical preferences for surgery compared to other treatment options.[8] A meta-analysis by Moxey and colleagues in 2003 found that, across 7 studies on framing and surgical treatment, positive framing on average led participants to be 1.5 times as likely to choose surgery. The same meta-analysis also found that positive frames increase preferences for more invasive or toxic medication compared to negative frames. Framing effects have similarly been documented for other kinds of medical outcomes. For example, attitudes and choices with regards to cancer treatments are affected by the framing of their success rate (vs. failure rate; Levin, Schnittjer, & Thee

[5] I discuss the possible issue of non-decisive influence further in Section 4a.

[6] E.g., see Levin, Schneider & Gaeth, 1998.

[7] E.g., Haward, Murphy, & Lorenz, 2008; Levin, Schnittjer, & Thee., 1988; Marteau 1989; McNeil et al., 1982; Schneider et al. 2001; Wilson, Kaplan, & Schneiderman, 1987.

[8] See e.g., Blumenthal-Barby and Krieger (2015) for a systematic review of studies of framing effects and other biases or heuristics on medical choices. Their review includes 213 studies in all, and 72 studies concerning loss/gain framing effects.

1988), preferences for antibiotics are affected by the framing of their efficacy (vs. failure rate; Smith et al., 2020), and preferences for taking medication are affected by the framing of the chances of experiencing (vs. the chances of not experiencing) side effects (like nausea; Minton et al., 2020).

Another example concerns care decisions for premature newborns. Haward, Murphy, and Lorenz (2008) asked participants to imagine having gone into premature labor at 23 weeks. In such a situation, parents have to decide whether they want the medical team to attempt to provide the extremely premature newborn with intensive care (which involves hooking the newborn up to various machines designed to keep the newborn alive), or whether to provide 'comfort care' (which is palliative, aimed at minimizing suffering until the baby dies). The likelihood of positive or negative outcomes from resuscitation through intensive care were described either using positive framing or negative framing. Participants were asked which option they would choose for their newborn. What followed was a striking example of framing effects on parental consent: participants were significantly more likely to say they would choose intensive care for their premature newborn if they had been given the positive frame.[9] This evidence suggests that the following scenario can occur:

> **Framing-Induced Intensive Care:** A pregnant woman goes into premature labor at 23 weeks. Her doctor explains that, in these particular circumstances, it's up to her whether they give the premature baby comfort care, or intensive care. If they provide comfort care, the baby will not survive, but it will allow them to make the baby as comfortable and pain-free as possible until it dies, and the mother will be able to hold the baby. If they provide intensive care, there is some small chance that the baby will survive, and some smaller chance that it will do so without any severe developmental disabilities.
>
> Specifically, the doctor explains that 25 out of 100 infants will survive in these circumstances if given intensive care; and of those who survive, 15 out of 25 will not have severe developmental disabilities. The mother chooses intensive care for her newborn.

[9] This effect has been replicated elsewhere e.g., Gong et al., 2018.

> However, had the doctor instead explained that 75 out of 100 infants will not survive in these circumstances even if given intensive care, and that, of those who survive, 10 out of 25 will have severe developmental disabilities, the mother would not have chosen intensive care for her newborn, and would have asked for comfort care instead.

Most existing studies on framing effects and medical choices investigate the effect of frames on hypothetical, rather than actual, treatment choices (resource constraints and ethical issues make it difficult to experimentally study actual medical choices directly). Extrapolation to actual choice is merely inferred as likely on the basis of that evidence. However, some studies have examined framing effects on actual choices, or at least on patients facing choices close to that being studied.

For instance, Banks and colleagues (1995) presented participants' with gain-framed or loss-framed persuasive videos about mammograms (e.g., the gain framed video contained statements such as "detecting breast cancer early can save your life" and "if a cancer has not spread, it is less likely to be fatal", whereas the loss framed video contained alternative statements such as "failing to detect breast cancer early can cost you your life" and "if a cancer has spread, it is more likely to be fatal"). They found that which video participants had seen affected the likelihood that participants actually got a mammogram within six months of having watched the video. Similar framing effects on actual uptake of mammograms have been replicated in other studies (e.g., Bertoni, Corazzini, & Robone, 2020) as have loss/gain framing effects on actual uptake of post-stroke therapy (Yang, Wang & Chen).[10]

[10] Bertoni, Corazzini, and Robone (2020) found that loss-framed messages significantly increased uptake of mammograms compared to gain-framed messages – but only when the messages contained "enhanced" information about the implications of screening.

Gong and colleagues (2016) examined the effects of presenting the risks of thrombolysis (a treatment for stroke) using positive or negative frames on in-patients (or their proxies) one day after the patients were admitted to hospital after having had a stroke. They found that those patients given the positively-framed information were more likely to say they would consent to thrombolytic treatment. Although this study did not record what procedure was actually offered and used for these patients, it's compelling evidence that their actual consent could be affected by framing.[11]

Furthermore, positive/negative framing effects have been documented to occur similarly for written as for auditory messages (Kreiner & Gamliel, 2016), making it likely that vignette-based effects might extend to consent decisions following verbally-communicated information from doctors. Indeed, in a remarkable recent study, Fridman and colleagues (2021) used automated text analysis on transcripts of physician consultations with patients who had just been diagnosed with early-stage prostate cancer. With early-stage prostate cancer, it's typical for patients to make a choice between active treatment aimed to eradicate the cancer (such as surgery or radiation) and active surveillance (careful continued observation of the cancer to monitor possible progression). After a six-month follow-up, the researchers found that the use of "death" and related words (e.g., "demise", "lethal", "decease", etc.) on the part of the physician in the initial consultation predicted the likelihood that a patient would choose to pursue active treatment rather than active surveillance. While it's not possible to infer causality from a correlational study such as this, taken together with the body of experimental work, it adds to the evidence that frames can influence actual consent outside of the lab.

[11] There are many other examples of studies recording framing effects while using closely 'representative samples' of some kind e.g., Batchelder et al., 2020.

Altogether, this body of evidence supports the claim that, in all likelihood, many real-life decisions to consent depend on framing effects: that people give consent when they would not have given consent had the option been framed differently, just like in Framing-Induced Surgery. And, intuitively, there seems to be something defective with consent given in this way, even though, had we not known about framing effects, the consent would have seemed to be valid.

If this intuition is right, then we should expect ethical issues due to frame-dependent consent to be widespread. First, the potential threat of framing to our moral lives stretches beyond the domain of medical decisions. As already noted, decisions to consent are far more wide-ranging: we consent to sex, to economic transactions, to use of property. And there is no *prima facie* reason to expect framing effects to be limited to medical contexts. Much empirical research on framing effects has focused on economic decisions, for instance. So if framing effects pose a problem for medical consent, we have reason to be concerned about the ethics of these other interactions as well.

Second, in contexts such as medicine where patients choose between multiple options we face the problem of frame-dependent consent *even when the way an option is framed leads a patient to say "no" to that option*. Assume, for the sake of simplicity, that whether a patient is presented with a negative or positive frame is more or less random, and that their consent to a medical procedure is contingent on this framing. It might be thought, then, that we face only a 50-50 chance of ending up with defective, frame-dependent consent. This is because frame-dependent consent is restricted to the scenario where the patient is presented with a positive frame and thus consents to the procedure, which is then carried out without their valid consent. After all, if they are presented with the negative frame, and they dissent, then the procedure is not carried out; and although there might be ethical

problems with this scenario too (is the patient foregoing beneficial treatment?), at least we don't incur the violations involved in intervening on a patient without valid consent. Or so it might be thought.

But this thought would be mistaken for many contexts, and in the medical context we will often face this kind of violation *whichever* frame the patient is exposed to. This will be so whenever a patient faces multiple treatment options, where the default alternative to some procedure or treatment is to undergo some *other* procedure or treatment instead that also requires consent. In such cases, if a frame leads someone to dissent to one option, then this frame will likely be causally responsible for their *consenting* to some *other* option. This is the case in Framing-Induced Surgery: if the agent refuses surgery, he'll opt for radiotherapy instead. But then his consent to *radiotherapy* will be causally dependent on a frame. Similarly for the stroke patients who reject thrombolytic therapy as a result of framing, and so consent to non-thrombolytic therapy instead; or the parent who rejects intensive care for the premature newborn, and so consents to allowing the premature newborn to die after a process of comfort care. So if consent-dependent framing is problematic, then, in many cases, we will be faced with problematic, frame-dependent consent whichever frame occurs and whichever resultant framing-induced choice the patient makes.

3. The Threat of Framing Effects to the Standard View of Consent

So far, we have defined framing effects, and we have summarized evidence of framing effects on decisions pertinent to medical consent. This evidence supports the following empirical claim:

> **Generalization:** Many cases of consent (that otherwise seemed valid) depend on framing.

This, combined with intuitions that frame-dependent consent is defective and invalid, tempts us to accept the following conclusion as a result:

> **Conclusion:** Many cases of consent (that otherwise seemed valid) are not valid.

Of course, this conclusion does not validly follow from the empirical generalization on its own. We need to add a premise that framing effects undermine the validity of consent.

Before we investigate what form such a premise might take and what kind of argument might be given in support of it, I want to argue that if such an argument is ultimately successful, then we will be required to take some kind of revisionary position with regard to what I will refer to as the 'Standard View' of consent in moral philosophy.

Let's use the term "consent-relevant acts" as a rough, informal way of referring to those acts or transactions for which consent is at least ordinarily thought to be required—such as sex, medical treatment, sales, and so on. And let's use the description "seemingly-valid" to refer to consent that has no obvious signs of being defective—consent that ordinary, well-intentioned and reasonable observers would treat as valid. For instance, in cases where the consenter is a normal adult who indicates approval of the act in a context-appropriate way (e.g., saying "Sure, go ahead,", or signing a medical consent form), where she is not the subject of coercion or deceit, and where the circumstances make it clear what is being consented to, the consenter has given seemingly-valid consent.

Now, according to the Standard View of the ethics of consent:

> **Moral Transformation**: Consent-relevant acts are morally forbidden (all else being equal) without valid consent, but are morally permissible (all else being equal) with valid consent.
>
> **Ordinary Consent:** In many ordinary cases where seemingly-valid consent is given to a consent-relevant act, the consent is in fact valid.

Together, these imply:

> **Ordinary Permissibility:** In many ordinary cases where seemingly-valid consent is given to a consent-relevant act, the consent-relevant act is morally permissible (all else being equal).

Let me briefly unpack and expand on those claims. I begin with the claim of moral transformation. According to the Standard View, certain acts require valid consent to be morally permissible. This is because the presence or absence of valid consent determines whether or not certain wrong-making features obtain of certain acts. The relevant wrong-making features are often thought of in terms of rights violations, and the morally transformative function of consent in terms of the waiving of those rights. One could in principle analyze the wrong-making features in question in different terms, and speak, for instance, of whether persons have been *wronged* in virtue of an absence of valid consent, without speaking of rights violations. Either way, what's crucial to the Standard View is just that this wrong-making feature is present or absent in virtue of the absence or presence (respectively) of valid consent.

The violations in question can differ in gravity. For instance, having sex with someone without their consent constitutes a serious moral violation, but taking their pen without their consent is not serious, even though it does involve a moral violation due to lack of consent. Still, the presence or absence of these wrong-making features can, in many (though not all) real-life cases, make the difference between an action being all-things-

considered permissible and all-things-considered morally wrong. So, valid consent can transform a case of would-be rape into a case of permissible consensual sex, a case of would-be impermissible theft into a case of permissible borrowing, a case of would-be impermissible trespass into a case of a permissible visit, or a case of would-be aggravated battery into permissible, consensual participation in a sporting event (as when a boxer is punched in the face during a match).[12]

The clause "all else being equal" simply indicates that whether or not an interaction is consensual is, of course, not the only thing that determines permissibility. Certain acts will be wrong for reasons independent of consent. For instance, if adultery is wrong, then consensual adultery is wrong (even if it's not the wrong of rape). What's central to the standard view of the ethics of consent is just that valid consent can transform certain morally impermissible actions into morally permissible actions, supposing the action is question is not impermissible for other reasons.

This leads us to the second component of the Standard View, that regarding ordinary consent. The ability of the Standard View to capture the contours of ordinary moral lives and practices is a large part of its appeal. Importantly, therefore, according to the Standard View, it's not just the case that valid consent *would* be morally transformative *if* it were ever given; rather, valid consent *is* given in many ordinary cases ("Ordinary Consent"). And this explains the all-things-considered moral permissibility of many real-life, everyday interactions that would otherwise have been all-things-considered impermissible ("Ordinary Permissibility"). For instance, it explains the putative permissibility of engaging in sports that involve violence; it explains why most surgeries seemingly do not constitute

[12] Hurd, 1996, calls this the "moral magic" of consent.

battery and are free from moral wrong (even though rendering someone unconscious and performing surgery on them would, without consent, constitute a serious moral wrong); it explains the many cases of morally permissible sexual intercourse that occur between adults (even though having sex without someone without their consent is morally wrong and may constitute rape);[13] it explains the moral permissibility of accessing and sharing the private information of those who have given us permission to do so (even though doing so without permission can be a serious violation of privacy); and so on.

Now we have the pieces in place to see why the problem of framing effects threatens the Standard View of consent. If framing effects undermine the validity of consent, then many ordinary cases of consented-to acts in fact do *not* involve valid consent—thus Ordinary Consent is false. But if Ordinary Consent is false, then we are forced to give up either Moral Transformation or Ordinary Permissibility as well. For if Ordinary Consent is false, but Moral Transformation is true, this implies that all of these consented-to acts are in fact morally *impermissible*—thus Ordinary Permissibility is also false. Alternatively, if Ordinary Permissibility is true even though Ordinary Consent is false—if these cases involve permissible acts even though valid consent has not been given—then it must be the case that Moral Transformation is false—that valid consent is not generally required for the permissibility of consent-relevant acts. Thus, if framing effects show that Ordinary Consent is false, we are left with two possible revisionary positions with regards to the ethics of consent.

One revisionary position is that of the Consent Pessimist. The Consent Pessimist accepts the conclusion that many cases of consent are in fact invalid, and finds this conclusion

[13] For an overview of philosophical and legal debates about the definition of rape and its relation to consent, see Whisnant, 2021.

ethically worrisome, because they still accept Moral Transformation, the component of the Standard View that says that valid consent is normally required to transform morally impermissible acts into permissible ones. They are forced, then, to accept that many consented-to acts that we thought were morally permissible in fact are not, rejecting the component of the Standard View, Ordinary Permissibility, that says that many ordinary cases of consented-to acts are morally permissible.

It's not hard to see why the Consent Pessimist might be troubled by the discovery that many ordinary cases do not, in fact, involve valid consent. This might be especially so in domains like medicine, where many interventions are such that they would constitute serious moral violations without valid consent (if Moral Transformation is accepted): physical exams, medications, treatments and surgeries can have life-altering or lethal consequences and/or are intimate and invasive in a way that would seriously violate privacy and bodily integrity if done without consent. (We are not talking about consent to pen-borrowing here.) Imagine that the doctors and surgeons in some hospital explain procedures in English. Following a survey of past patients, it is discovered, to the horror of the hospital, that many of the patients—including ones who underwent serious non-emergency surgeries—don't speak a word of English. It is now apparent that serious procedures were conducted on these patients even though they had little or no idea of what the procedure was, what it was for, what it would involve, or what their other options were, and so their consent could not possibly have been valid. This is highly ethically troubling (even if the doctors in question are not to blame e.g., because these patients really seemed to understand).

The Consent Pessimist is troubled in just this way by empirical evidence that suggests that many cases of consent are not valid due to framing effects: for the Consent Pessimist, the

discovery that consent is subject to framing effects is akin to finding out that consent is subject to serious ignorance of what is being consented to due to language failures. Indeed, the problem of framing effects seems even worse than that faced by the imaginary hospital! For in the hospital case, we are limited to a specific hospital, the problem affects a limited subset of patients within that hospital, and it's fairly clear what the hospital can do now to address the lack of valid consent (implement language comprehension checks and provide translations, etc.) But framing, such as mortality/survival framing, affects a wide range of people across a wide range of contexts, not just a specific subgroup that can be easily identified ahead of time through something equivalent to a language check. What's more, even if at-risk cases could be identified, it's not obvious what one could do to solve the problem. After all, (i) it's necessary to communicate relevant information about procedures for informed consent to be given (otherwise they might not be able to understand the nature or consequences of the procedure to which they are consenting); (ii) to communicate this information, it's necessary that the information is *presented* somehow. So it doesn't seem possible to simply "do away" with frames as one can do away with language barriers.

To see why this problem is difficult to solve, take the communication of risk and the mortality/survival framing effect. A patient who is subject to mortality/survival framing will change their decision depending on whether the risk is framed negatively (20% mortality – they say no) or positively (80% survival – they say yes). Insofar as dependence on framing entails that consent is invalid, one cannot do away with the framing problem simply by trying to provide some more 'neutral' presentation. For instance, imagine you tell this patient that the procedure carries "a 20% risk of mortality and a corresponding 80% chance of survival", and the patient consents. Although this might seem like a more 'neutral' frame, for all that, this patient's consent is no less dependent on framing effects.

For, if "80% chance of survival" is informationally equivalent to "20% risk of mortality", then so too is "20% risk of mortality and a corresponding 80% chance of survival" informationally equivalent to "20% risk of mortality". And by hypothesis, this patient would *not* have consented if you had simply said that the procedure carries "a 20% risk of mortality".[14] Thus the patient's decision depends on mere wording, and their consent depends on a framing effect, despite efforts to choose avoid a frame. Indeed, some studies suggest that presenting 'both' frames introduces *another* kind of framing effect: participants can be influenced by whether the positively or negatively framed information is presented first (Kreiner & Gamliel., 2016).

Thus we arrive at a rather morally pessimistic view: the Consent Pessimist's position is that valid consent is morally required for many interactions to be all-things-considered morally permissible (Moral Transformation), but it turns out that valid consent is given much less often than we thought, so many more cases are in fact impermissible and involve wrong-doing than we thought.

A different kind of revisionary position one can adopt is that of the Consent Skeptic. The Consent Skeptic accepts the claim that many cases of consent are invalid, but ultimately rejects the intuition that we are justified in being ethically troubled by this. Instead, the Consent Skeptic prioritizes the claim that most interactions that *seem* morally permissible *are* morally permissible (Ordinary Permissibility), such as doctor-patient interactions that involve no deception, no coercion, and the supply of relevant information, and where the

[14] E.g., Smith et al (2020) report a study on preferences for antibiotic treatments where participants received either positively framed information about efficacy, negatively framed information (i.e. corresponding failure rates), or information framed both ways. They report that the condition that included both frames affected preferences in a similar way to the positive-only frame, leading to more favorable evaluations than the negative-only frame.

patients are adults with normal competencies, etc.; the Consent Skeptic squares this with the conclusion drawn from evidence of framing effects by rejecting Moral Transformation, the component of the Standard View that says that moral transformation occurs *only if* consent is valid. So—according to this position—if framing effects show that consent is hardly ever valid, this just shows that we were mistaken to think that valid consent is required for the elimination of rights violations, and thus morally permissible action, in the first place.[15] Something else (perhaps certain kinds of mere assent) suffices. This position is not pessimistic, since it is not committed to the claim that many acts that seem morally permissible to us in fact are wrongful; but it is skeptical about standard views about the ethical importance of valid consent, and in that way is highly revisionary.

Thus, given the plausibility of the Generalization premise that framing effects are widespread, anyone who accepts that framing effects undermine the validity of consent is forced to embrace some form of consent revisionism, whether in the form of Consent Skepticism or Consent Pessimism.

In what follows, I will argue that evidence of framing effects should not make Consent Skeptics or Consent Pessimists of us, even if we accept the empirical claim that many cases of consent are subject to framing effects.

[15] Hanna (2011) defends such a line of thought.

4. Generalization and the Extent of Framing Effects

According to the Generalization claim, many cases of consent (that otherwise seemed valid) depend on framing. As mentioned, since framing effects have been demonstrated in many contexts, and since frames are in some sense inescapable whenever information about choices is presented somehow, it's plausible that in some sense of "many", many decisions are frame-dependent. To the extent that *all* or *most* consent decisions depend on framing, the more plausible it is that the Standard View of consent must be abandoned in favor of either Consent Pessimism or Consent Skepticism (granting, for the moment, that framing effects do undermine the validity of consent). But presumably the motivation for adopting these revisionary positions becomes weaker to the extent that a smaller proportion of consent judgments are affected. So what exactly is the extent of the effect of framing—how many is "many"? And what implications does this have for the Standard View?

4.a. The Challenge of Counting Cases

Recall that frames can affect choices without exerting a decisive effect on them—decisive in the sense of altering whether or not a particular option is chosen. Non-decisive effects could include making an option more or less attractive, affecting the way an agent thinks about an option, making an agent more or less confident about their choice, etc., but not to the extent that the agent would have chosen a different option overall had the framing been different. I will focus specifically on the question of how many cases of consent are *decisively* affected by framing: that is, in how many cases an agent would have made a different choice had the options been framed differently, as in the case of Frame-Induced

Surgery. This choice of focus is based on the assumption that while decisive influence poses a potential ethical challenge to consent, non-decisive influence does not.

I won't offer a full defense of this assumption here, as I take it the burden would be on the Consent Revisionist to make it more plausible that non-decisive influence *does* pose a credible threat to consent. This seems a reasonable burden to place on the Consent Revisionist for two reasons. Firstly, unlike cases of decisive influence, cases of non-decisive influence are not intuitively troubling. Secondly, it seems unlikely (not logically impossible, but unlikely) that framing would affect an agent in a sufficiently dramatic way so as to undermine the validity of their consent while at the same time failing to make a decisive difference to the agent's decision. To conclude that frames affect decision-making in a non-decisive yet problematic way, therefore, we would have to posit a rather specific mechanism for how frames affect decision-makers, and argue that this undermines consent. Moreover, it's not obvious what this mechanism would be, or why we should believe that frames affect decisions according to such a mechanism. Accordingly, the burden would be on the Consent Revisionist to provide an argument for why non-decisive influence should, in fact, trouble us.

So we want to know the extent of the problem of decisive influence of framing on consent decisions—that is, the extent of frame-dependent consent. How many consent decisions we can expect to be frame-dependent depends on two factors: (1) the number of circumstances in which there are frames that exert effects on decisions; and (2) the chances of those frames exerting a decisive influence on decisions.

So the number of circumstances in which there are framing effects depends, in the first instance, on the number of circumstances in which there are frames *present* in the first place

of the sort that have been shown to have an effect on decisions. Estimating this is a rather murky business. Of course, almost anything can in principle qualify as a 'frame', since we have defined frames simply as features of the form or mode in which a decision is presented independently of the decision's content. Since all decisions are faced in some form or other, 'frames' as such are inescapable. However, we are only interested in those features of the way a decision is presented that have been shown to have effects on decisions. A revisionist about consent would not have a strong argument if it were based on the mere *possibility* that some irrelevant feature of the way a consenter is presented with their choice could affect their decision.

So we must restrict our focus to framing effects for which we have empirical evidence. For example, we have seen that whether a feature such as risk is framed positively or negatively can affect decisions. We can expect the presence of these frames to be ubiquitous in consent contexts that involve explicit descriptions of the options that may be consented to—such as medical contexts, legal contracts, and commercial transactions (e.g., in written information about products or services). In such contexts, information must be presented in *some* way, be that positively or negatively, before consent is given, and thus frames of this kind are typically present.

But it's worth noting that such frames may not be present at all in other consent contexts where the features of options are not explicitly described. And it trivially follows from the fact that they are not present that they cannot exert an effect on consent. For instance, it's possible for sexual encounters to involve no verbal descriptions whatsoever of what is being consented to or its consequences—shared understanding of what is being consented to may be entirely based on unvoiced background knowledge (of such situations and of the other person), on wholly implicit communicative cues, and on interpretations of

nonverbal behavior. In such cases, there may be no explicit descriptions to be framed positively or to be framed negatively, in which case there can be no effects of explicit positive/negative framing on the decision to consent to sex. There *have* been studies of the effect of framing on sex-related health behaviors. For instance, Richardson and colleagues (2004) found that safe-sex counselling with HIV positive patients using a loss frame ("We encourage you to make choices that do not put yourself or others at risk. Unsafe sex may expose you to other sexually transmitted diseases or other strains of HIV") led to lower rates of self-reported unprotected sex at follow-up compared to the counselling that used a gain frame ("We encourage you to make choices that protect yourself and others. Safer sex protects you from other sexually transmitted diseases and from other strains of HIV"). However, such interventions are less related to whether or not valid sexual consent has been given and more to do with whether or not people adopt safe sex practices when having sex.

Of course, even in cases that lack the explicit presentation of information about a decision, relevant information may be tacitly framed in various ways that could potentially be shown to affect decision-making independently of the content of the decision. But the point remains that the presence of frames that have been shown to affect decision-making is likely not uniform across all contexts, and this places some limits on the number of circumstances in which consent may be reasonably believed to be subject to framing effects.

Still, the more types of frames that there are that do affect decisions, the more ubiquitous that framing effects have the potential to be. So for the sake of argument, let's set aside possible circumstances in which there *are* no frames that significantly affect decision-making, and focus only on those cases in which such frames are, at least, present. Now we

want to know: for any given decision, what are the chances that the mere framing of the decision will exert a decisive effect?

When scientific studies find that some factor, like framing, affects decisions, it is tempting, I think, to interpret this as a finding that this factor has a very strong effect and that it on its own causally determines what people's decisions are.[16] So, given the earlier studies that show the existence of framing effects, it is tempting to conclude that they play a common and decisive role in our decisions.

But while this can in principle be the case for some factors that are found to affect decisions, it's often not the case. This is because even if an effect on decision-making is real, the number of decisions that the effect actually sways in a decisive way depends on the strength of this effect. The 'strength' of an effect, as I am using the term, is a function both of the way framing affects agents (e.g., how 'hard' it 'tugs' people towards or away from a particular choice) and on the agent's susceptibility to having their decision succumb to this effect in a decisive way. The stronger the effect of frames, the greater the likelihood that someone who would choose option A would instead choose option B had some piece of information been framed differently (for instance, negatively rather than positively), and therefore greater the number of decisions that will end up being swayed by framing. Correspondingly, however, if the effect of framing is weak, relatively few decisions will be swayed by framing, even if frames are ubiquitous.

[16] This may be due to the features of the way we ordinarily apply the concept of "cause" in everyday, nonscientific contexts. Scientific studies identify causal influences on behavior, where these causal influences may be very small, very large or anywhere in between. But I suspect that our colloquial, non-scientific usage of the term "cause" is more often restricted to stronger causal effects—unsurprisingly, since (i) stronger effects are more likely to be detectable without the aid of scientific study, and (ii) stronger effects have greater causal and therefore normative relevance to personal decision-making.

Importantly, the existence of framing effects would be compatible with the possibility that the strength of framing effects is quite weak. This is because even if the strength of an effect is very small, such that proportionally few decisions would actually be swayed by it, scientific studies are able to detect it as a real effect so long as there are enough participants taking part in the study. Here's an analogy to illustrate the point: if you give a small group of people a push each, but each push is only very gentle, it might be that nobody, or hardly anyone, actually falls over. But if you give hundreds of people tiny little pushes, some people are likely to fall as a result. Consequently, if we compare the group you're pushing to another group, and more people fall over in your group, we can detect this and conclude that your pushing is having a real effect and can cause people to fall. Nevertheless, it's still the case that your push is quite weak, so it's very unlikely to make any *particular* person fall over.

So the likelihood that any particular decision will actually be frame-dependent depends not only on the presence of frames, and not only on the reality of framing effects, but also on the strength of the effect that framing has on decision-making. Furthermore, even if framing can exert powerful effects, the strength of this effect may not be constant. Firstly, it may vary for different types of framing.[17] And, even for any given type of framing, the strength of the effect may vary (i.e. may be "moderated", in scientific jargon) according to context, and according to features of the individual agent.[18] Contextual factors could

[17] Levin, Schneider & Gaeth, 1998; Levin et al., 2002; Freling, Vincent, & Henard, 2014.

[18] For instance, Heilman & Miclea (2016) report finding differences in the 'risky choice' framing effect between different domains (e.g., financial vs. medical) and also for different scenarios within a domain; they speculate that levels of emotional involvement may be responsible for these differences. A review by Gallagher and Updegraff (2012) found that loss- vs. gain-framing had greater effects on illness prevention behaviors like skin cancer

include facts about the circumstances, the type of decision at issue, or ways choices are presented.[19] Maybe frames affect some kinds of decision more than others (as some results suggest e.g., Heilman & Miclea, 2016), or more so when particular types of option are on the table. For instance, it could be that survival/mortality framing has a stronger effect in the context of considering whether to choose a more serious and invasive procedure like a surgery than in the context of considering whether or not to receive an immunization. (A meta-analysis by Moxey and colleagues in 2003 suggests this very possibility, although they do not report any studies that attempt to directly compare these contexts.)

The strength of framing effects could also differ between individuals if certain traits make one more or less susceptible to being swayed by framing effects. Scientifically, this is referred to as moderation of an effect by "individual differences".[20] If there is little in the way of individual differences with respect to the strength of framing effects, then a

---

prevention, smoking cessation, physical activity, and safe sex, but weak or no effects for diet or vaccinations.

[19] An analysis of attribute framing literature by Freling, Vincent & Henard (2014) suggests that whether a decision focuses on abstract or concrete attributes, and whether the decision-maker is 'psychologically distant' from the choice, moderate the strength of framing effects.

[20] A systematic review (Covey, 2014) of framing effects in health contexts suggest that individual differences such as approach-avoidance motivation, regulatory focus, need for cognition and self-efficacy beliefs may moderate framing effects. For example, Levin and colleagues (2002) report findings that personality traits (specifically, Conscientiousness and Agreeableness) and 'faith in intuition' can moderate attribute framing effects. A number of studies report that framing effects are moderated by measures of cognitive style, such as 'Need for Cognition', but the findings on this have been very mixed overall—for instance, an early study found that only low NC participants were susceptible to framing (Smith & Levin, 1996), but other studies have found no effect of measures of cognitive style (e.g., Mandel & Kapler, 2018; LeBoeuf & Shafir, 2003; Levin, Schneider, & Gaeth, 2002), and more recent studies have reported effects in the opposite direction, with high NC participants showing equal or even greater susceptibility to framing effects (e.g., Dunegan, 2010). Evidence has been mixed as to whether framing effects differ for different age groups (e.g., Rönnlund et al., 2005; Kim et al., 2005). Cultural factors form another dimension that could play a moderating role in specific circumstances. For instance, a study by Ortiz, Martinez & Espino (2015) finds framing effects on end-of-life preferences for Latino, but not White, older adults.

moderate effect size indicates that framing exerts a moderately strong effect on all or most individuals. On the other hand, a moderate effect size could instead be the result of framing having stronger effects on some individuals, but weaker or no effects on others.

Because the likelihood that any particular decision is swayed by framing depends both on the likelihood that the decision is exposed to a particular frame (which is context-dependent, and difficult to provide an estimate of, especially when it comes to contexts that lack rigid, formal procedures); and because the strength of framing effects may vary according to the type of framing effect, the context and the individual (in many complex ways about which we do not yet have scientific consensus); and because many studies are unable to examine the effect of frames on actual consent decisions (relying instead on examining the effect of frames on judgments or reported behavioral intentions), ascertaining *how many* consent decisions can be expected to be swayed by framing on the basis of studies of framing effects with any precision is a very challenging task. Things get even more complicated very quickly when one considers the possibility that someone could in principle be affected by *multiple* frames at the same time—such as being affected both by whether survival rates are presented in terms of mortality *and* by the order in which options are presented to them. To assess the rates of decisive framing in such cases, we would need to know not only how strong the effects of survival/mortality and order effects are, but how they interact—whether they are additive, for instance, or cancel each other out, or have some other relationship.

4.b. Attempting an Estimate: What Does the Evidence Show?

So we can see that, in principle, it is rather difficult to come up with a general estimate. Nevertheless, let's see what we can glean from an examination of some of the studies in question, setting aside some of these complications for the sake of argument.

For a case study, let's focus specifically on the case of the framing of survival/mortality rates with regards to surgery—a relatively specific context where we would expect the explicit sharing of risk information to be the norm. Take the original experiment by McNeil and colleagues (1982)—the study that motivated the case of Frame-Induced Surgery described at the beginning of this chapter, and that inspired a lot of further investigation of framing effects in medical contexts. In this study, participants were given descriptions of two possible treatments for lung cancer: surgical treatment, and radiation treatment. They were then told the survival rates of both treatments over time, with radiation therapy having a better initial survival rate (since not all patients survive the surgery itself), but surgery having a better long-term survival rate (presumably since it is more effective). Half of the participants received these in terms of mortality (e.g., "Of 100 people having surgery, *10* will die during treatment, *32* will have died by one year and *66* will have died by 5 years") and half in terms of survival (e.g., "Of 100 people having surgery, 90 will survive treatment…" etc.).

While 75% of participants said they preferred surgery over radiation therapy when the information was presented using survival framing, only 58% of participants said they preferred surgery over radiation therapy when the information was presented with the mortality frame. This means that 17% fewer participants preferred surgery over radiation therapy when the information was presented with the mortality frame (correspondingly,

17% fewer participants preferred radiation therapy over surgery when presented with the survival frame compared to the mortality frame). Assuming that the only difference between the two groups is the frame they were exposed to, we can infer that the framing was responsible for the 17% difference in choices between the two framing conditions.

Now, let's compare this result to other studies that have tested this same case (or something very similar to it). Blumenthal-Barby and Krieger (2015) conducted a review of studies of biases and framing effects on medical decisions using systematic methods for searching potentially relevant literature. They identified 72 published studies that investigated positive/negative framing effects on medical decision-making.[21] Of these, 28 found a significant framing effect; another 28 found a significant framing effect but only in a subpopulation of participants studied; and 16 did not find any effect of framing. Out of the list of studies that found a significant effect for at least one subgroup studied, let's pick out those studies which (i) examined the case of the effect of framing of survival (mortality) rates on preferences for surgery, as in the McNeil study, or a similar case; and (ii) from which it is possible to extract information about the percentage of participants selecting surgery in each framing condition.

Doing this, it turns out that other studies have recorded similar differences in the proportion choosing surgery when given a mortality frame vs. a survival frame:[22]

[21] They call the category "loss/gain framing"; in it they include studies both of goal framing and attribute framing.

[22] There are a few studies that also report significant effects of mortality/survival framing on the scenario of choosing between surgery and radiation therapy to treat cancer, but that do not report the proportions of participants selecting surgery in the different conditions: this is the case for Almashat et al. (2008), and Marteau (1989); Woodhead, Lynch and Edelstein (2011) asked participants to use a 'think-aloud' procedure as they reason through the case, and they find a significant effect of survival/mortality framing for participants that they identify as using a 'data-driven' decisional strategy (i.e. those who base their decision on the information about survival rates provided to them) but not for those who

- LeBeouf and Shafir, 2003: 24% fewer participants preferred surgery for treating lung cancer given a mortality frame.
- Kim et al., 2005: for an older group of participants (aged 58+), 38% fewer participants preferred surgery for treating cancer given a mortality frame. For college-age participants, 17% fewer preferred surgery when given the mortality frame compared to the survival frame (but this effect did not reach statistical significance).[23]
- Wilson, Kaplan, & Schneiderman, 1987: 18% fewer participants preferred a high-risk surgery (one with 90% mortality rate) to treat a terminal case of liver disease when given a mortality frame.
- Smith and Levin, 1996: 21% fewer participants preferred surgery over radiation therapy for treating cancer given a mortality frame.[24]

---

use an 'experience-driven' decisional strategy (i.e. those that base their decision on their personal experience or background knowledge and beliefs). The effect for the 'data-driven' participants is large (approximately 30% difference between frames); however, the authors do not provide data that allows us to identify the proportion affected by framing collapsing across these two groups. Christensen and colleagues (1995) also examined the effect of framing survival risk on surgical vs. medical treatment choices for several specific medical cases, but in this study they asked participants with medical expertise to provide treatment choices for hypothetical patients, instead of for themselves—thus the study does not test the effect of framing on hypothetical decisions *to consent* to treatment, but rather to *recommend* treatment. In their study, 29% fewer participants overall recommended the surgical over the non-surgical option in a case involving treating a 30-year-old woman with bacterial endocarditis and aortic insufficiency; 12% fewer participants overall recommended the surgical over the non-surgical option in a case involving treating a 28-year-old man with a brain AVM. However, they found no significant of framing for 9 other medical problems. Furthermore, the effects found for these two vignettes summarized here were only significant for medical residents and experienced physicians, but not for medical students.

[23] In the study by Kim and colleagues (2005), these effects were found only in a "standard" condition; a different group of participants, both older and younger, who were asked to provide justifications for their choice did not exhibit framing effects.

[24] This is collapsing across different pools of participants. Smith and Levin found the framing effect to be significant only for participants who scored 'low' in what's known as "Need for Cognition" (NFC), but not those who were high in NFC. Amongst low NFC

The studies vary in their results. Variability between individual studies is to be expected, especially when the number of participants tested in the studies is relatively low. Still, many report approximately 20%-25% differences in the proportion of people preferring surgery depending on framing.

Of course, this little survey is only rough-and-ready, and has not been conducted in any kind of scientific manner. I have made no attempt to control for methodological quality, nor have I taken into account the number of participants involved in each study. In addition, many of these studies do have relatively small sample sizes (partly because of when they took place—smaller sample sizes were more common when participants had to be recruited in person instead of over the internet, which is cheaper and faster, and before more contemporary concerns about the replicability of psychological results, especially from smaller studies). In general, the fewer participants involved in a study, the more that random noise and chance can affect the results. Consequently, small studies that do find a significant result can overinflate the size of an effect, and the figures that come up in any one study cannot be assumed to be representative with any precision.

Using a scientific method called meta-analysis is the best way to address these problems and obtain a more reliable estimate of the size of an effect. Meta-analyses collate and analyze the results from many studies that examine some effect, while taking into account factors such as sample size. They also often assess studies for inclusion on the basis of methodological quality. Unfortunately, however, it is difficult to use existing meta-analyses in order to provide an estimate of the sort we are concerned with here: estimating the

participants only (N=50), the difference in the proportion of participants preferring surgery in the survival condition compared to the radiation condition is 33%.

proportion of participants who are liable to be decisively influenced by equivalent frames when giving consent. This is because scientific meta-analyses (i) often include studies on many decisions that are not relevant to consent; (ii) often include framing manipulations that are not plausibly equivalent, and thus are not good candidates for threatening consent; (iii) are not restricted to examining *decisive* influence on choice, instead examining the effect of framing more broadly, including non-decisive influence (e.g., how framing effects strength of preference or attitudes); and, finally and relatedly, (iv) provide estimates of 'effect size' that are not easily translatable into the question of the proportion of people that are, or would, be decisively influenced by framing.

For these reasons, I am relying here on a rough and informal approach. Still, it's illustrative. Across the studies just listed, approximately a fifth fewer participants prefer a surgical option when presented with mortality rather than survival framing. In the first instance, this tells us something about *how many people's hypothetical choices in the studies themselves* were frame-dependent: about 20%. This is based on two assumptions. Firstly, we assume that participants were allocated either to a mortality frame or to a survival frame at random, so there are no consistent differences between the two groups other than the framing. This allows us to reason as follows. Say we give 100 people a survival frame, and x out of 100 say they choose surgery over radiation. Our evidence suggests that 20 fewer people would have chosen surgery over radiation had we given them a mortality frame. So our evidence suggests that only x-20 would still have chosen surgery had they been given a mortality frame. This means that 20 of these people would not have chosen surgery over radiation had we given them a mortality frame; they would have chosen radiation therapy instead. The same reasoning applies in reverse when considering people given a mortality frame: if y out of 100 say they choose radiation in this condition, our evidence suggests that in fact

only y-20 would still have chosen radiation if they had been presented with the survival frame instead; 20 others would have switched to surgery.

So, across all frames, this seems to suggest that around 20% of these kinds of hypothetical decisions are frame-dependent: 20% pick surgery if presented with a survival frame, but radiation if presented with a mortality frame.

However, there are several reasons for thinking that this figure is not a reliable indication of the true number of frame-dependent consent decisions in such contexts.

One difficulty with generalizing from the rate of frame-dependence found in these studies to the rate of frame-dependent consent is that it's unclear in what way they generalize to different types of medical procedures than the ones studied. For instance, most of these studies focus on the effect of survival/mortality framing in the context of choosing to surgery to address life-threatening illness, replicating some of the classic early findings. But of course, many surgeries that take place are used to address illnesses that are not life-threatening. And there are many other types of illness, and many other types of treatment options for those illnesses for which people consider giving consent. To the extent the research has examined only a limited range of cases, the less confident we can be that the size of the effect will be similar in other cases as well.

Another difficulty with generalizing from these studies is that these studies examine hypothetical choices—judgments about what participants *would* choose *should* they face the choice in question. Even if about 20% of hypothetical choices are frame-dependent, it's not entirely clear what the proportion would be like for actual, real-life choices, and whether this proportion would be higher or lower. As already mentioned, it's more difficult

for both ethical and practical reasons to study the effect of framing on real-life decisions, especially for a specific choice such as whether or not to undergo surgery or radiation therapy, so direct evidence is more limited. (It's easier for other cases; for instance, there are more studies on the effects of framing information on real-life uptake of cancer screening, or immunization.)

It's not clear whether studies that use hypothetical decisions provide an under- or over-estimate of the true rate of frame-dependent decisions. On the one hand, it might be thought that facing the real prospect of mortality could *enhance* the effects of how such information is framed (empirically, the way that 'personal involvement' in a decision does or does not moderate framing effects remains controversial). A meta-analysis by Gallagher and Updegraff in 2012 of the effect of loss- vs. gain-framed messages in the medical domain did find stronger evidence for effects of framing on behavior (whether self-reported or objectively measured) than on intentions or attitudes. This renders some plausibility to the possibility that framing could have a stronger effect on actual decisions than it does on hypothetical ones.

However, the applicability of the studies reviewed by Gallagher and Updegraff to the current topic may be limited. Firstly, although some of the reviewed studies concern health behaviors that involve the giving of consent (e.g., cancer screening), most of them do not (e.g., smoking cessation, sunscreen use). Secondly, so-called 'goal frames' or 'gain- vs. loss-frames', which emphasize the positive consequences of performing behaviors or the *negative* consequences of *not* performing behaviors respectively, are not always closely 'equivalent' framings of the same information (as we noted in Section 2). For instance, in one study on smoking cessation, the gain-framed paragraph includes the claim, "In fact, not using tobacco is the best way to save lives"; the corresponding claim in the loss-framed

paragraph is "In fact, tobacco use is the leading preventable cause of death" (Steward et al., 2003).[25] Because of this, it's not clear to what extent the studies concern effects that are potentially consent-undermining in the way that Framing-Induced Surgery appeared to be.

Furthermore, real-life decisions involve many more complex factors and influences on decisions. For this reason, it may be thought that framing is likely to play a relatively *smaller* role for real-life decisions. It may also be harder to detect, especially without the availability of large samples of participants. Perhaps because of this, effects of framing have not always been detected for real-life decisions on the occasions that they have been studied, at least for the very specific cases and sample sizes that have been investigated (e.g., about cancer

[25] Failure of equivalence is not necessarily a problem for the stated aims of the studies, which are sometimes practically rather than theoretically oriented towards simply to identifying how we can use wording to best promote healthy behaviors or interventions. Nevertheless there are many other examples of failure of equivalence. Consider the contrast between the benefit of "more successful personal relationships" resulting from taking a hypothetical anti-alcoholism vaccine compared to "more troubled personal relationships" resulting from not taking the vaccine (Wirtz, Sar & Ghuge 2015); relationships can fail to become more successful without thereby becoming more troubled. Similarly, consider a study that contrasted a doctor's advice about continuing to eat bacon which would cause cholesterol to "significantly rise" and "greatly increase" chances of cardiovascular disease vs. advice about stopping eating bacon which would "significantly reduce" cholesterol and "greatly reduce" chances of cardiovascular disease (Peng, Jiang et al., 2013). Of course, being told that continuing unhealthy eating will significantly increase one's risks does not imply that stopping unhealthy eating will reduce one's risks; it only implies that stopping unhealthy eating means your risks will not significantly rise, which is compatible with one's risks staying the same—not necessarily decreasing. Other contrasts used in goal framing studies correspond more closely. For instance, Van 't Riet, Ruiter and De Vries (2011) contrasted the gain-framed message, 'when you check your skin for changes once a month, you can detect skin cancer in an early stage' with the loss-framed message, 'when you do not check your skin for changes once a month, you might detect skin cancer in a late state'. Another example is the contrast between phrases used in loss- or gain-framed information videos, such as the gain-framed "detecting breast cancer early can save a woman's life… When a woman gets regular mammograms, she is doing her best to detect breast cancer early. And, detecting breast cancer early can save her life" compared to the loss-framed "failing to detect breast cancer early can cost a woman her life…When a woman does not get regular mammograms, she is not doing her best to detect breast cancer early. And failing to detect breast cancer early can cost her life" (Gallagher et al., 2011). These sorts of contrasts are not logically equivalent, but norms of conversational implicature are such that the content of one frame is suggestive of the other.

treatment in a study by Siminoff and Fetting, 1989). Given that framing effects have been widely replicated across many studies, including studies using representative samples (i.e. participants who are actually facing the sorts of medical decisions that the study asks them to reason about hypothetically) and many actual decisions (although non-hypothetical surgery-vs.-radiation decisions, specifically, have not been studied), some null results here are unlikely to mean that framing effects aren't real for actual decisions. However, it's difficult to know their exact extent.

So it's unclear how hypothetical decisions relate to the rate of framing effects in real-life decisions. However, there are some reasons for thinking that the studies examined provide an *underestimate* of the true number of frame-dependent decisions.

The first is that the claim that around 20% of participants in the studies made frame-dependent choices is based on the assumption there aren't any people who are disposed to be affected by framing in the *opposite* direction to the majority of others—for instance, to prefer surgery when presented with a mortality frame, but to prefer radiation when presented with a survival frame. This is not an outlandish possibility. For instance, consider that these studies tend to include information suggesting that surgery has a better *long-term* survival rate than radiation. While immediate survival looms larger for most people given our tendency to disproportionately discount possibilities that are further in the future, it's conceivable that some people might focus more heavily on long-term results. If so, it could be that the long-term benefit of surgery relative to radiation is enhanced if considering this in terms of reduced mortality rather than increased survival. But such patterns, if there were any, would be not be revealed by most of the studies discussed here, given their design. Because of this, it's possible that the proportion of people in these studies whose responses are frame-dependent *in some way* is larger than the proportion of people in these

studies whose responses are frame-dependent *in the same way that the majority of people's decisions can be frame-dependent.*

However, it's unlikely that the number of participants who are decisively affected by framing in the opposite direction is large. For example, Levin and colleagues (2002) participants respond to differently framed versions of the same problem (with a week in between and "filler" tasks to make it hard for participants to directly compare their answers to the two versions). While they found a significant effect of framing, with a number of participants showing more favorable attitudes following positive framing, only a small minority showed a change in the non-standard direction (and due to the small number, it's not clear whether this is due to responding in an opposite way to framing, or simply due to random variation in responses).[26]

Secondly, occasionally the documented effect has been much larger. For instance, a bar chart published with the study by Wilson, Kaplan and Schneiderman suggests a much larger effect for a version in which the surgery for terminal liver disease was stated as having a 40% survival rate: here, almost 100% of participants favored surgery when presented in a survival frame, but only around 60% when presented in a mortality frame,

[26] Levin and colleagues (2002) similarly replicated risky-choice framing in a within-subjects design, but they found no significant effect for goal framing overall, with most subjects not affected by goal framing, and, of those who did alter responses to different frames, approximately the same number of participants responded more favorably to the positive frame as participants who responded more favorably to the negative frame. Indeed, in the context of goal framing, there is some evidence that individual differences can moderate whether gain frames (which emphasize the benefits of adopting a behavior or treatment) or loss frames (which emphasize the bad consequences of *not* adopting a behavior or treatment) are more likely to lead people to adopt a health behavior (such as applying sunscreen, attending mammography appointments, or flossing). See Covey (2014) for discussion. However, most findings regarding individual or contextual moderation of framing effects report finding that certain traits or conditions reduce or increase the extent of framing effects, rather than reverse their direction (e.g., the effect of age on risky choice framing, Best & Charness, 2015)

suggesting as many as 40% of people might be swayed by framing in such a case (however, the exact data for these conditions are not reported in the text of the paper).

On the other hand, there are other reasons for thinking that the figure of 'a fifth' is likely an *overestimate* of the number of consent-dependent decisions.

Firstly, some studies and conditions find smaller effects. This includes some conditions in the same Wilson, Kaplan, and Schneiderman paper: the effect they found was much smaller or non-existent in other conditions tested.[27]

Secondly, there may be a bias in the literature towards the publication of studies that found larger effects. This is because scientists are generally less able and therefore less likely to publish reports of experiments that did not find an effect. Consequently, even if there were more studies that tested the effect of framing on choices that found effects that were too small to reach statistical significance, or that found no effect at all, these are simply less likely to have been published. Now, it's important to note that meta-analyses of other types

[27] Specifically, the effect of framing was small at 40% and at 60% survival rates, and non-existent at 80% survival. Macchi and Zulato (2021) attempted a replication and extension of the study by McNeil and colleagues. While they find an unusually enormous effect in their straightforward replication attempt (from an 81.3% preference for surgery down to 20.0% in the mortality frame), they also find *no* effect of framing when using a different numerical format for presenting the survival rates over time; on this basis they argue that McNeil's original result was affected by participants misunderstanding numerical information presented in a cumulative format, as presented in the study by McNeil and colleagues. Both the apparent elimination of the effect in their modification and the very large effect size using the original wording should be interpreted with some caution, however, especially without further replication: many other studies have replicated the effect of survival/mortality framing on choices with the cumulative presentation format that Macchi and Zulato argue is confusing; on the other hand, significant effects of framing on healthcare decisions found in other studies have not been so dramatically large.

of framing effects have not found evidence of publication bias. But it remains a reasonable concern that it could affect cases of the sort in which we are interested in here.[28]

Thirdly—and this consideration is more decisive—in real life, the number of *decisions to consent* that are dependent on framing will be fewer than the number of *decisions in general* that are dependent on framing. In artificial experimental set-ups, participants are forced to choose between different medical procedures (such as choosing either surgery or choosing radiation therapy). In a set-up like this, failing to (hypothetically) consent to one procedure implies consenting to another. Thus all decisions are decisions to consent to *something*. However, in real life, people have the option of refusing to have any kind of treatment at all. In such cases, they do not provide consent. They dissent. Because of this, even if 20% (say) of real-life *medical decisions* were contingent on framing, not all of these would result in framing-dependent *consent*. Of course, frame-dependent *refusal of treatment* might pose

[28] A number of meta-analyses have tested for publication bias, but these tend to include attribute frames that are not equivalent, or focus on goal-framing. A 2014 meta-analysis of attribute framing by Freling, Vincent, and Henard did not find evidence of publication bias, but it should be noted that their analysis excluded studies that involved participants making a forced choice between two options, instead only including studies in which participants evaluated a single option/object, much of it from the consumer decision-making literature; it thus did not include studies like the ones discussed in the section involving choices between medical treatment options. They also employed a different definition of 'logically equivalent frames' than that employed in the present paper, as they included temporal frames (e.g., redemption intentions for gift certificates expiring soon vs. later) and what they call 'referent frames' (e.g., consumer attitudes evaluating ads for a charity that benefits the self vs. others) as well as valence framing (i.e. positive vs. negative frames) and numerical framing (e.g., cents off vs. percentage off). Their own valence framing study, reported in the same paper, also does not involve logical equivalence: participants were told either that "four out of six group members rated the target person positively" or "two out of six group members rated the target person negatively", but clearly these are only equivalent if you cannot report neutral attitudes about a person, which is not a plausible assumption and does not appear to have been controlled. A 2012 meta-analysis (Gallagher & Updegraff, 2012) did not find any evidence of publication bias for the effects of goal framing on prevention and detection behaviors in the health domain. Similarly, a meta-analysis of the effect of gain- vs. loss-framing on cancer prevention and detection found six studies between 2000 and 2020 which together show that loss framing is more likely to lead to cancer detection behaviors (e.g., mammography and cervical cancer screening); the authors did not find evidence of publication bias (Ainiwaer et al., 2021).

ethical problems of its own, insofar as this causes patients to forego treatment that is crucial for their wellbeing. But this, of course, wouldn't be an ethical challenge to the validity of their consent, since they do not give consent.

4.c. Risky Choice Framing and the Asian Disease Problem

At this point, it's worth taking a moment to discuss the relevance, or irrelevance, of 'risky choice' framing effects to the current discussion. Since these effects are well-known and are also dramatic, they might—mistakenly—be taken to mean that framing will have dramatic effects on consent, such that rates of frame-dependent consent will be very high indeed, and much higher than something like a fifth of decisions.

In the famous demonstration of framing effects using the so-called "Asian Disease Problem"—reported in Tversky and Kahneman's seminal 1981 paper—participants were asked to imagine that the U.S. was preparing for the outbreak of a disease that is expected to kill 600 people, and that they needed to choose between two possible programs to combat the disease. In the positively framed version of the problem, the options were described as follows:

> If Program A is adopted, 200 people will be saved.
>
> If Program B is adopted, there is 1/3 probability that 600 people will be saved and 2/3 probability that no people will be saved.

Other participants were given a negatively framed version of the same options instead:

> If Program C is adopted, 400 people will die.

> If Program D is adopted, there is 1/3 probability that nobody will die and 2/3 probability that 600 people will die.

In the positively framed condition, 72% of participants preferred Program A—the "sure" program which saves 200, but also, by implication, allows 400 to die—over Program B, the "risky" program where there is a chance of saving everyone but also a larger chance of saving no-one. However, when these programs were framed negatively in terms of the number of people who would die, only 22% preferred the "sure" program; instead, most people in the negatively framed condition preferred the "risky" option, Program D, over the "sure" option, Program C. So in this famous study, the proportion of frame-dependent decisions is very large: there was a full 50% difference in the proportion of participants who preferred the "sure" option, depending on framing.

However, there are several reasons to doubt that we can extrapolate from the size of this framing effect to the strength of framing effects on medical consent decisions. Firstly, studies suggest that different processes are responsible for each of risky choice framing, attribute framing, and goal framing (Levin, Schneider, & Gaeth, 1998; Levin et al., 2002).[29] Secondly, meta-analyses have indicated that the size of the framing effect in the Asian Disease Problem and structurally similar cases are unusually large compared to cases that use other domains or formats.[30] Several features of the problem may contribute to this.[31]

---

[29] Although this terminology is widely used in the literature and is standard, the literature is not entirely consistent in its application. For instance, occasionally the term "risky choice framing" is used to refer to positive or negative framing of risk as in the case of framing surgery in terms of survival or mortality rates (e.g., Peng, Jiang et al., 2013). This use of the terminology is not standard and is not in line with the typology of Levin and colleagues; it seems to be a mistaken interpretation of the terms due to the fact that the attribute in question is risk.

[30] Kühberger, 1998; see also Steiger & Kühberger, 2018.

[31] See Kühberger, 1998.

Consider the following respect in which the Asian Disease Problem is very different from typical consent decisions: the options in the Asian Disease Problem appear to have exactly the same expected utility,[32] and are identical in terms of most of their known attributes. Thus, this sort of framing effect is normally interpreted as showing something about preferences regarding something rather specific: the value people place on being sure of a certain gain (sure to avoid a certain amount of loss) compared to the value of having the chance of an even greater gain (a chance to avoid even more loss).[33]

This means that while framing has a very large effect on preferences about certain gains vs. risks for greater gains when all else is constant, this isn't strong evidence that framing will strongly sway real-life consent decisions. Firstly, it's not clear how many real-life consent decisions pit certainty vs. riskiness at all; often, we consider options that differ in their level of risk, or that risk different things. For this reason, we can't infer from the fact that framing has a very strong effect on preferences for risk vs. certainty that it will have a very strong effect on choices in cases of real-life consent. Secondly, even in cases that do differ with respect to certainty vs. risk, there are usually additional relevant differences between options that determine which one is preferred overall. For instance, say we altered the scenario so the choice was between saving 500/600 for sure or the standard option of 1/3 chance of saving everyone and 2/3 chance of saving no-one. Presumably if we framed this choice negatively—between the options letting 100/600 die for sure, or facing 1/3

[32] We might say that they have the same expected utility, but this would be an oversimplification. They might not. The options are identical in terms of the expected number of lives saved, calculated as the sum of the probabilities multiplied by the number of lives saved with those probabilities. But if further possible consequences of those policies are taken into account, they may not be equal in expected utility—for instance, unpopular policies cause more upset, anger, etc.

[33] Tversky and Kahneman's Prospect Theory offers the canonical theoretical explanation of this effect. However, Reyna's Fuzzy Trace Theory (Reyna & Brainerd 1991; Broniatowski & Renya 2018) offers a competing account.

chance of nobody dying and 2/3 chance of everybody dying—the relative preference for risk-seeking due to loss framing is unlikely to be enough to outweigh the fact that the sure option in this version is expected to save many more people. Correspondingly, real-life consent decisions normally involve choices between options that differ in very many attributes. In many cases, some of these differences will be more important to the decision than the certainty/risk contrast. It's therefore less likely that manipulating responsiveness to riskiness vs. certainty through framing will exert as large of an effect in most cases of real consent decisions.[34]

4.d. What Do the Numbers Mean? Normative Implications

Once again, when examining the evidence we ran into a number of difficulties when trying to estimate the number of consent decisions that would in fact be affected by framing. Nevertheless, it's useful to consider the normative implications of plausible outcomes.

Let's assume, on behalf of the Consent Revisionist, that framing would affect actual choices at roughly the same rates as they appear to affect choices within hypothetical vignette studies of the sort we have just been discussing.[35] And let's assume, for the sake

[34] We return to an analogous point in the case of attribute framing in Section 11.

[35] How appropriate is such an assumption? On the one hand, in a number of areas of psychology, real-life effects have been demonstrated to be weaker than effects found in the lab, so using effect sizes from lab studies to make projections about effect sizes in real life may seem to be unjustly friendly to consent revisionists (i.e. the Consent Skeptic or the Consent Pessimist). On the other hand, *some* lab studies suggest that factors like 'personal involvement' in the topic at hand—factors that one would expect to be heightened in real-life decision-making—can enhance the strength of framing effects in certain circumstances e.g., Wirtz, Sar & Ghuge 2015. But other studies have resulted in the opposite pattern, with personal involvement *reducing* susceptibility to framing effects – e.g., see Donovan & Jalleh, 2000; Gesser-Edelsburg et al., 2015. So it doesn't seem unreasonable to grant the Consent Skeptic such an assumption, if only for the sake of argument.

of argument, that this is about 20%. This would make it plausible, for instance, that about a fifth of actual decisions regarding this kind of surgical treatment are dependent on framing.

Assume, again for the sake of argument, that a decision being contingent on framing undermines the validity of consent. What should we take the normative implications of such a discovery to be? On the one hand, if the majority of all decisions were frame-dependent, we would have strong reason for suspecting that valid consent is not required for permissible medical intervention after all (i.e. strong reason for accepting Consent Skepticism). However, we are supposing that *a fifth* of medical decisions are frame-dependent. Is this enough to motivate the view that, after all, consent need not be valid in order to be morally transformative? This question does not admit of an obvious answer, and a Consent Skeptic would be tasked with generating a fuller argument for their favored normative interpretation of these results, and in particular that *this* many cases of framing effects is sufficient to warrant their favored philosophical overhaul of the Standard View of consent.

(The details of the Consent Skeptic's argument may have to depend, in turn, on how this proportion of faulty consent-decisions are distributed. On the one hand, to the extent that framing effects are not much moderated by individual differences, it may be that most people make frame-dependent decisions about a fifth of the time. On the other hand, to the extent that framing effects are heavily moderated by individual differences, it may be that some people make very many frame-dependent decisions to consent—much more than a fifth of their decisions—but others make very few—much less than a fifth. Whether or not one thinks that one or the other of these scenarios makes Consent Skepticism more

or less plausible depends on thorny issues to do with how to understand what is "ordinary" in the context of the coherence of ethical theory with "ordinary" practices.)

Even if we do not accept Consent Skepticism on the basis of such results, we might resign ourselves to a limited version of Consent Pessimism, and conclude that about a fifth of surgeries involve wrongful violations of consent (albeit perhaps blameless wrongs). Again, the flavor of Consent Pessimism seems different once again depending on how these wrongful violations are distributed: the case where there is some limited proportion of people whose consent decisions are frame-dependent much more than a fifth of the time seems intuitively more tragic than the case where the wrongs are evenly distributed between persons. Still, either way, the Consent Pessimist has a philosophically easier time compared to the Consent Skeptic of interpreting the significance of the finding that a fifth of consent decisions are invalid, since Consent Pessimism itself can be proportionally scaled: while we can be pessimistic about more decisions or fewer decisions, it seems we ought either to believe that valid consent is morally transformative or to be skeptical about this claim as a general matter of the nature of consent.

In any case, it's important to see that even this more limited, scaled-down version of Consent Pessimism would be a surprising and troubling ethical conclusion in its own right.[36] Therefore, ascertaining whether or not framing effects *do* undermine the validity of consent is of ethical interest, even if the Generalization claim is only true on an

[36] It would also seem to have methodological implications for using ordinary cases to build philosophical theories about consent. If many cases of ordinary and *valid-seeming* consent do not, in fact, yield permissible interactions (assuming, of course, that we can't tell of any particular case that someone would have been swayed by framing), this may make reflection on ordinary cases of consent a proportionally weaker source of evidence for theorizing about the relationship between consent and permissible interactions. However, such an argument depends on further details of the epistemology of using ordinary cases in moral theorizing.

interpretation of "many" that is relatively small—say, a fifth of consent decisions, or even fewer. Whatever the exact proportion turns out to be, we ought to care about whether that proportion of cases involves moral violations. Ascertaining whether or not framing effects do undermine the validity of consent is the topic to which we shall now turn.

5. Sufficient Autonomy

So, should we endorse the intuition that framing effects undermine the validity of consent? To answer this question, we must say a little bit more about what we mean by "valid consent".

Not all tokens of consent (e.g., the intentional utterance of "yes" or "okay" in response to a request) constitute valid, and thus morally transformative, consent. If an act or transaction between A and B is such that it requires B's consent, and it is carried out although the consent is invalid, then the act or transaction can be deemed non-consensual (in the moral sense of the term) and A wrongs B accordingly (assuming, for the time being, that Consent Skepticism is wrong). One could insist that only valid consent truly *is* consent, and that invalid consent is at most merely *apparent* consent.[37] Without taking a stand on this metaphysical question, for current purposes I adopt the terminological position of referring to cases of apparent consent as 'consent'. So, in this dissertation, use of the term 'consent' leaves open whether or not the consent in question is valid.

An intuitive thought, and one that is widely endorsed in bioethics and moral philosophy, is that there is a deep connection between valid consent and autonomous agency. So,

[37] Kleinig (2010, p.15), for instance, expresses the view that invalid 'consent' is never actual consent—perhaps just 'assent'.

although there is disagreement about how best to analyze the relevant notion of autonomy, supporters of the Standard View commonly take consent to be valid if and only if it is autonomous in some appropriate sense. The idea, roughly expressed, has intuitive appeal: If A gives B consent to x, but A does not consent of their own free will, or was not properly free to make a different choice, or if the consent is not properly their own decision, or if they are unable to decide whether or not to x on the basis of their own values or plans, then it seems reasonable to think that the consent does not make it morally permissible for B to x. The analysis of valid consent in terms of autonomy captures why paradigms of non-autonomous assent are also paradigms of invalid consent, such as coerced consent, or consent by incompetent agents (such as the extremely young or severely mentally ill): in these cases, the consent is not an expression of autonomous agency, and thus does not morally authorize.

Not only does autonomous consent require freedom and non-interference from third parties (such as freedom from coercive control), but—importantly for our purposes—it also imposes certain psychological decision-making requirements on the part of the consenter.[38] These requirements plausibly involve both cognitive and conative capacities. On the cognitive side, the consenter must be able to give consent with sufficient knowledge and understanding of their decision (for instance, of what they are consenting to, its important consequences, and what other options are available). On the conative side, the consenter must be able to give consent (or withhold it) in light of that understanding and whichever conative states are important for autonomous action—such as their own preferences, desires, values, or plans. (In this paper, I'll speak loosely of the

[38] The literature on autonomy sometimes puts something like this point in terms of a dual requirement of independence, on the one hand, and "self-rule," on the other; see e.g., Dworkin, 1988.

agent's 'desires' or 'values'; the reader can replace this with whichever conative states they take to be most relevant given their favorite theory of autonomous action.)[39]

These cognitive and conative requirements must be met to some sufficient degree for consent to be valid. However, no plausible theory of consent—certainly no theory that is supposed to capture contours of ordinary judgment and practice, as is any theory committed to the Standard View—could require optimality. This is because optimality in these capacities is most likely never achieved. On the cognitive side, it's unlikely that we are ever able to reason about a choice with perfect theoretical consistency and in full knowledge and understanding of all of the consequences involved. On the conative side, it's unlikely that we are ever able to make decisions in a way that is causally sensitive to all relevant desires and values, perfectly weighted according to their importance. Apart from anything else, we simply lack the resources: we make decisions in a limited amount of time, we are only able to process and attend to a limited amount of information at once, and introspectively accessing, weighing and applying relevant values takes time and attention. So we rarely, if ever, have the ability to pay attention to, evaluate, and respond to all relevant considerations when making a decision.

Even if we set aside resource constraints of time and attention, the task is difficult to accomplish perfectly. For one, there are epistemic difficulties in knowing what one most wants or values, even for highly competent and emotionally intelligent decision-makers. This may be especially true when there are competing considerations at work. For instance, do I value an increased 10% chance of being cured of my ailment enough to outweigh the

[39] Different theories of autonomy specify the cognitive and conative requirements in different ways, often in terms of rational competencies and authenticity; see Christman, 2020, for an overview.

disvalue of the expense and discomfort of a longer, more expensive treatment?[40] These kinds of epistemic difficulties are such that even sufficiently autonomous decision-makers can make mistakes within certain margins of error and yet still be considered to be making reasonable decisions and giving valid consent.

Because even highly skilled human decision-makers cannot be expected to decide in optimally autonomous ways—to make decisions with optimal understanding and optimal sensitivity to reasons grounded in their desires and values—valid consent can only plausibly require some sufficient level of understanding and sufficient level of sensitivity to reasons grounded in desires and values of the agent. Otherwise, we would be ruling out the very possibility of preserving the Standard View—according to which many ordinary, everyday consent-transactions are valid and permissible—from the outset.

For instance, it's generally thought that people can validly consent to receiving a vaccination without optimal understanding of all the mechanisms and consequences involved, and indeed with minor failures of understanding, so long as they have a sufficient grasp of important aspects of what they are consenting to (presumably the vast majority of vaccinations occur under such conditions). Similarly, people can validly consent to cosmetic surgery even if it's not the optimal decision in terms of what will satisfy their deepest desires, preferences and plans at that time, and even if cultural pressures and beauty standards play some role in forming the desires that drive such decisions (as surely is almost always the case, to some degree). People can also validly consent to sex if they do so freely and willingly even if in doing so, they are acting against their better judgment, exhibiting some weakness of will, or giving in to ordinary, understandable levels of wishful

[40] This may be an instance of what I describe and name 'psychological incommensurability' later on in this section.

thinking (surely an all-too-common occurrence, yet one that is untroubling—a subject, in cinema, of many romantic comedies).

Still, many ways of making decisions plausibly fail to be sufficiently autonomous because they involve *insufficient* understanding, rationality, or connection to autonomous desires and values. For instance, it's plausible that you cannot validly consent to a vaccination if you think you are being administered an anesthetic for an upcoming procedure instead; to cosmetic surgery if you're unaware that it involves significant medical risks, if you're a child, or if you're bullied into it by an abusive partner or boss; or to sex if consent is based on coercive pressure.

So, we can state the autonomy requirement on valid consent as follows:

> **Sufficient Autonomy:** Consent is valid only if it is sufficiently autonomous.

where this principle might be further analyzed into more specific requirements on valid consent related to autonomous decision-making, including:

> **Sufficient Cognitive Autonomy:** Consent is valid only if the agent's consent meets cognitive requirements of autonomous decision-making to a sufficient degree (e.g., consent is given with sufficient knowledge, understanding, or theoretical rationality).
>
> **Sufficient Conative Autonomy**: Consent is valid only if the agent's consent meets conative requirements of autonomous decision-making to a sufficient degree (e.g., consent is sufficiently dependent on, and sufficiently in accord with, the agent's reasons, desires, or values).

In sum, choices must be sufficiently autonomous for valid consent, where this allows for decisions to be made in a range of imperfect ways, but does not allow for serious imperfections in decision-making that too greatly interfere with the autonomy or

voluntariness of the choice. This makes sense if a theory of valid consent is to be compatible with the very possibility of consent being valid in many ordinary cases involving normal, competent adults making decisions based on desires with decent levels of understanding of their options and what they are consenting to.

The principle of sufficient autonomy gives us a natural hypothesis for why framing effects might undermine consent. Framing effects are a psychological phenomenon; they are a feature of the decision-making processes of would-be consenters. So a natural possibility is that they undermine consent because frame-dependent consent involves a violation of Sufficient Autonomy: frame-dependent consent fails, in some way, to meet adequate standards of rational processing based on the agent's reasons, either by falling short of cognitive requirements on autonomous decision-making or falling short of conative requirements on autonomous decision-making. For instance, perhaps framing effects are symptoms of a failure to grasp and understand important aspects of what it is that is being consented to or substantial failures of theoretical rationality; or perhaps they lead patients to make choices independently of their preferences, desires or values.

Indeed, many existing worries about the impact of framing effects on consent can be read as falling into these categories. For instance, the scientists responsible for the study of framing effects on parental decisions for premature newborns worry that framing effects "may compromise autonomy in decision-making and thereby compromise the integrity of the informed consent process"[41] because framing effects either compromise the patient's understanding of risk or constitute an external control on the parent's decision.[42] A number

[41] Haward, Murphy & Lorenz, 2008, p.114.

[42] Ibid., p.118.

of bioethicists and moral philosophers have also worried that framing effects compromise sufficient understanding: Beauchamp and Childress (2013, p.135) suggest that frame-dependent consent is invalid because the decision is driven by inadequate understanding (e.g., of risk of death), and Blumenthal-Barby (2016) also argues that frame-dependent consent is threatened by failures of understanding. Others have emphasized the impact of framing effects on the conative side: Schwab worries frame-dependent consent does not reflect autonomous desires (2006, p.575); Hanna argues that framing invalidates consent because framed choices do not "reflect the subject's settled values and preferences" (2011, p.525); Mills is concerned that frame-dependent consent is not "authentic" (2013, p.33). Still others have concerns that straddle both of these categories: Chwang suggests that framing effects are "incompatible with autonomous choice" because an agent who is subject to framing effects is thereby "incapable of reasoning consistently" and makes an "irrational" choice (2016, pp.274-275).

We can interpret these worries, in sum, in terms of the idea that framing effects threaten the validity of consent because frame-dependent consent violates sufficient autonomy. We haven't, of course, provided a precise definition of what counts as a sufficient level of autonomous decision-making. But framing effects are supposed to be showing us something new and surprising about the quality of ordinary decision-making, and the claim that we are evaluating in this paper is whether evidence of framing effects makes it such that many cases of consent that are ordinarily thought of as valid are in fact shown to be insufficiently autonomous in light of what we know about framing effects. Consequently, an appropriate rule of thumb for what counts as 'sufficiently autonomous' for our purposes is what we—or, at least, supporters of the Standard View of consent—regarded as the standard of decision-making achieved in many ordinary cases (at least, prior to

knowing about framing effects) by competent agents that we regarded as capable of giving valid consent.

Of course, a more stringent standard of autonomous decision-making could be argued for, one that is not consistent with the claim that these kinds of ordinary cases meet the requirements of sufficient autonomy. But if such a standard is accepted, then framing effects have correspondingly little work to do in leading us to reject the Standard View of consent; the higher standard would be enough to show that many cases of consent ordinarily thought of as valid are insufficiently autonomous, even if there were no such thing as a framing effect.

What kinds decision-making processes would frame-dependent consent, such as that described in Framing-Induced Surgery, have to involve to fail to meet standards of sufficient autonomy? On the cognitive side, as already discussed, patients clearly do not need to have the understanding of an expert physician or a statistician to have the ability to give valid consent, nor need they have perfectly consistent theoretical beliefs about their treatment, or valid consent would be exceedingly rare even in the case of routine procedures consented to by highly competent adults. Nevertheless, it would be worrying if patients consented to procedures on the basis of serious misperceptions of what its consequences were—for instance, if they didn't understand that an 80% chance of survival still meant that death was possible, or they explicitly believed that a 20% chance of mortality *was not* the same as an 80% chance of survival. If it turned out that framed decisions involve substantial failures of understanding or theoretical rationality such as this, especially where those failures of understanding concern an important feature of what is being consented to, and where this failure is critical to the agent's decision to consent, then the claim that frame-dependent consent is likely not valid would be compelling.

Alternatively, framing effects could undermine consent via the conative component of sufficient autonomy. Even if a failure to make the best decision in the best way available to the agent does not necessarily undermine consent,[43] it's plausible that a substantial departure from preferences, desires and values would be threatening to the validity of consent.

If framing effects are to threaten consent, therefore, they must show that agents' choices are divorced from an agent's preferences, desires and values in some more striking way than ordinary and commonplace errors that are not normally taken to be inconsistent with permissible, consent-based interactions. It wouldn't be enough, for instance, to show that a framed consenter only takes into account some limited proportion of all relevant information, acts on a desire that doesn't serve a deeper or more authentic life-plan, or arbitrarily weights a pre-existing and sincerely held value or desire more heavily than they ought to at the time of deciding. These are quotidian failures.

However, it would be plausible that framing effects undermine valid consent if, for instance: frames affect decisions through triggering overwhelming and fleeting emotion (such as fear), or through some other mechanism that mimics this in bypassing autonomous agency or in causing lack of control; if frames cause people to say "yes" to things they don't have any desire to do or have all-things-considered desire not to do; if framing causes the agent to temporarily weight his desires in such a way that is deeply irrational given his background desires and values (e.g., if the term "mortality" causes fear

[43] For this reason, we should not be convinced by arguments that framing effects threaten consent that rely on the claim that a framed decision may not reflect the agent's "true" or "most authentic" preference, insofar as that thought is based on the idea that only one possible choice is "the" autonomous choice for an agent at a given time.

of death to swamp the decision-making process at the expense of other relevant considerations of comparable importance to the agent in question in a way that is out of proportion to the importance he normally attaches to this).

On what basis might it be argued that frame-dependent consent lacks sufficient autonomy? In the coming sections, I outline different forms such an argument might take, and evaluate the success of such arguments.

6. The Argument From Variability and the Options Thesis

All cases of frame-dependent consent—that is, consent that is subject to framing effects—are defined by two features: firstly, the consenter's decision exhibits a kind of dispositional *variability*—a kind of counterfactual instability—in the following sense: they make one choice, but they might easily have made another. More specifically: they said yes, but they might easily have said no. Secondly, these variations in the consenter's decision are causally dependent on the presence of alternative ways of framing something about the agent's choice. Let's call this simply "variability" for simplicity (although, of course, only one choice is in fact made—the point is that the agent is disposed to have made other, alternative choices, contingent on framing).

Many commentators are tempted by the thought that the problem of framing effects lies in the variability itself—that being simultaneously disposed towards consenting and not consenting means that the decision to consent is unlikely to be sufficiently autonomous. According to this view, we don't need to know more about the causal explanation behind the variability; knowing that a consenter's decision is counterfactually unstable is sufficient to call the autonomy of the consenter's decision into doubt.

What might justify such a view? It seems to rely on the following thought: for an agent facing a given choice in a given situation, it's not the case that the decision to *p* and the decision to *not-p* could *both* be a sufficiently autonomous decision. On the assumption that frames do not relevantly alter the agent's choice or situation, this implies that an agent who is subject to framing effects is simultaneously disposed to decide that *p* and to decide that *not-p* with regards to a given choice in a given situation. It follows that an agent who is subject to framing effects is disposed to make an insufficiently autonomous decision. Furthermore—the argument goes—how the options are framed is independent of which of the choices is in fact the sufficiently autonomous one;[44] this means that there's no reason to think that how the options were in fact framed led to the sufficiently autonomous choice rather than the insufficiently autonomous choice. Therefore, if an agent's choice is dependent on framing, we have reason to doubt that the choice being made is the sufficiently autonomous choice for the agent.[45]

The crucial premise here is the assumption that, for an agent facing a given choice in a given situation, it's not the case that the decision to *p* and the decision to *not-p* would *both*

[44] This is analogous to the assumption of epistemological debunking arguments that some influence on the belief that *p* is independent of the truth of *p*. The assumption that frames are independent of the autonomy of medical choices will be questioned later in Section 8. However, we can grant it for current purposes.

[45] Again, this is similar in structure to epistemological arguments for the conclusion that an agent has formed an unreliable belief if the agent believes *p* but would have believed *not-p* just in case some factor, like a frame, that is not relevant to the truth of *p* had been different; see e.g., Sinnott-Armstrong (2008) for such an argument regarding the implications of framing effects for moral epistemology. The role played by the *truth* (of a belief) in an epistemological debunking argument is analogous to the role played by the autonomy (of a decision) in the argument that framing effects undermine consent. The epistemological debunker is correct to assume that truth is not subject to variability—that is, that either *p* is true, or not-*p* is true, but not both). However, as I will now argue, the analogous step in the consent debunker's argument—namely the assumption that either *p* is autonomous, or not-*p* is autonomous, but not both—is implausible.

be a sufficiently autonomous decision. We can call this assumption *No Options*. Say an agent is facing a decision about whether to choose surgery or to say 'no' to surgery and choose radiotherapy instead. According to No Options, only *one* of the possible choices would be sufficiently autonomous.[46] In such a case—absent additional insight into the agent's decision-making that could give us good reason for thinking that, out of the two possible decisions, the agent has made the sufficiently autonomous one—our chances seem at best to be fifty-fifty that the agent makes a decision with sufficient autonomy for valid consent.[47]

But No Options—the assumption on which this line of reasoning rests—is false. In fact, it's often the case that an agent in a given situation facing a choice between two (or more) consent-requiring options could choose either possible option with sufficient autonomy. The agent could thus validly consent to either option. We can call this the *Options* thesis.

The Options thesis thus laid out is 'substantive' in the sense that it concerns the options that the agent may choose: it says that not only could an agent validly consent to option A, it's also the case that they could validly consent to option B. But this also has implications for the agent's decision-making processes, since a process that leads to option

[46] No Options is consistent with the possibility that *neither* the decision to opt for option A *nor* the decision to opt for option B would be sufficiently autonomous. In this case, whatever the agent in fact decides, their decision cannot constitute valid consent. But No Options cannot establish this on its own – additional argument would be needed that neither decision is sufficiently autonomous.

[47] It is in fact a more complex matter than it might appear to get from the claim that only one of the possible decisions would be sufficiently autonomous to the conclusion that *we ought not believe that the actual decision was sufficiently autonomous*—a conclusion which I take it is crucial in establishing that framing effects threaten the Standard View of consent. In fact, elsewhere I have criticized the logic that variability implies at best a fifty-fifty chance of success, in the context of framing effects and moral intuitions (Demaree-Cotton, 2016). However, I won't pursue these issues here, and focus instead on the plausibility of the claim that at most one of the possible decisions could be sufficiently autonomous.

A being chosen must be different to a process leads to option B being chosen. The Options thesis thus implies that an agent in a given situation facing a choice could undergo different decision-making processes, where each of those processes could give rise to sufficiently autonomous, valid consent.[48] If the Options thesis is correct, then we cannot conclude, merely on the mere basis of the propensity to variable choices, that an agent is disposed to make a choice with insufficient autonomy.

Why think that the Options thesis is true?[49] One is the intuitively appealing idea that competent agents are often free to make autonomous decisions about what they want to do and thus to what they give valid consent. For example, many women are presented with multiple long-term contraceptive options—such as IUD's (intrauterine devices), subdermal contraceptive implants, and contraceptive injections—all of which must be fitted or administered by a doctor or nurse, and thus require consent. Intuitively, it's generally-speaking true that women face a genuine choice here: that is, many women who

[48] Of course, the difference between these processes may only be a fine-grained one; it's possible that the processes are not distinct at a coarse-grained level (e.g., "reflectively weighing pros and cons" vs. "going with your gut").

[49] I have already argued for the principle of sufficient autonomy, according to which valid consent does not require that decisions are made with perfect autonomy in order to be valid; only sufficient autonomy is required. It might be thought that this already shows that No Options is false. But this would be too quick, for two reasons. Firstly, it might be thought that even if we need not engage in an optimal decision-making process, valid consent still requires that we have selected the *option* that is optimal with respect to autonomy, even if we choose that option as a result of an imperfect process (e.g., a heuristic). Secondly, even if choosing optimally (in a procedural or substantive sense) is not required because choosing optimally is not a relevantly available option for normal human decision-makers, it might be thought that there is normally only one sufficiently autonomous way of choosing available—the most autonomous way of choosing that is available to the agent at the time. In addition, there are two further reasons why it's important to offer additional argument for Options. Firstly, doing so helps to shed light on the kind of decision-making that is required for valid consent, and the kind of justification that would be needed in order to argue that frames undermine valid consent. Secondly, it appears to figure as an assumption, albeit an implicit one, in much of the literature that does argue that frames undermine valid consent.

face such a choice at a given time could validly consent to either one of these options, and in doing so permissibly receive their chosen contraceptive. According to this intuitive picture, it's not determined ahead of time, so to speak, that they will validly consent if and only if they choose one particular option through one particular decision-making process. By contrast, No Options has the odd implication that, for many ordinary women facing this choice, there is at most one option to which they could validly consent, and their consent to another contraceptive choice could not be valid. This result is highly counterintuitive. This gives us a reason to reject No Options in favor of Options. (No Options is equally unappealing when considering consent in other, nonmedical domains. Take the sexual domain: imagine someone choosing between multiple potential sexual partners, or indeed between multiple possible sexual acts. The idea that only one option here can result in valid sexual consent seems obviously incorrect.)

Yet we need not rely solely on such intuitions to establish Options. In addition, Options is predicted and explained by a number of features of autonomous decision-making.

Firstly, the nature of agents' values is typically such that more than one choice can be chosen with *equal* levels of autonomy. More precisely: in many cases, for any choice that can be chosen with sufficient autonomy, there is some other choice that could be chosen with an equal, and therefore sufficient, level of autonomy. This can happen for two reasons. Occasionally, the available options may be exactly tied in the sense that both options satisfy the agent's balance of desires and values to equal degrees. More importantly—for precise ties might be rare—the *vagueness* and the *incommensurability* of competing reasons or values may often be such that the agent's values underdetermine the

exact autonomy-based merit of different available options, such that neither choice need be made less autonomously than the other, even in the absence of an exact tie.[50]

For instance, Vera is choosing between the hormonal IUD, which may cause acne but also reduced, lighter periods, and the copper IUD, which happily doesn't cause acne but also doesn't carry the benefit of reducing her periods (and she moderately values the convenience and comfort this would bring her). The two options are tied in all other respects she cares about. Vera moderately dislikes the prospect of a risk of hormonally-induced bad skin—a prospect that would be avoided by the copper IUD. But the disvalue she attaches to this is vague: there is no fact of the matter as to *how much precisely* she dislikes this prospect. Because the disvalue Vera attaches to the prospect of acne is vague, there *is no fact of the matter* as to whether the added comfort promised by the hormonal IUD outweighs the additional prospect of acne. Consequently, there is no fact of the matter as to whether the hormonal IUD or the copper IUD is the better choice overall for her given her values. (This is so even if her values do rule out other options, vagueness notwithstanding. For instance, from her point of view it's better to risk some acne than to use no contraceptive at all, given her very strong desire to control her risk of pregnancy, and it's also better than using an oral contraceptive that she has to remember to take every day, as this is something she strongly wishes to avoid).

Now let's assume that Vera could autonomously choose the copper IUD and thus consent to having it fitted—it's a sufficiently desirable and reasonable option for her, she is well informed about the choice, and is not subject to any autonomy-undermining influences. If there is no fact of the matter as to whether the hormonal or copper IUD is a better choice

[50] See Chang, 1997.

for her, it follows that Vera could also autonomously choose to have the hormonal IUD instead.

The same result comes about in a very similar manner if some of the values in question are of different types, resulting in limits in their comparability. For instance, say that Izzy slightly values the prospect of reduced periods: they don't normally bother her, but a reduction would be mildly better. However, her medical insurance means that she would have to pay $75 out of pocket for the hormonal IUD, but nothing for the copper IUD. (The options are tied in all other respects Izzy cares about—she is not concerned with a risk of acne, for instance.) So the hormonal IUD has the advantage of a small degree of extra comfort and convenience as a side-effect, and the copper IUD has the advantage of saving a bit of money. Even assuming that Izzy values these things to precise degrees, they might simply be too different to be meaningfully compared and weighed against one another with any precision.[51] The thought is that slightly reduced periods and $75 in cash are just not comparable types of things. This means that one prospect is not better than other, but nor are they equally good; they are just different advantages.[52] Again, since options tie in other respects, it follows that neither option can be said to be a better choice for Izzy than the other. But then we have reasons to think that Options is true. For assume, as is plausible, Izzy could autonomously consent to getting one of the IUD's—say, hormonal. In this version of events, the copper IUD would be no worse an option. It seems to follow from that this Izzy could consent to the copper IUD instead with *no lower* levels of autonomy than she has when consenting to the hormonal IUD. It then follows

[51] The possibility of valuing something to a precise degree and yet for that value to fail to be precisely comparable to another value depends on one's accounts of what it is to value something to some degree at all.

[52] See Chang, 1997.

that Izzy could autonomously consent either to having the hormonal IUD fitted *or* to having the copper IUD fitted. Thus the Options thesis holds for Izzy.

Some readers may reject the presupposition that values can be metaphysically vague or incomparable. They will thus reject this argument in favor of the Options thesis.[53] These readers might argue that while the values like comfort and money might *seem* to be metaphysically vague or incommensurable, that's only because they are only *psychologically* vague or incommensurable. That is, it's just really difficult to introspectively *tell* the precise amount one values or disvalues some things (psychological vagueness), and it's just really difficult to correctly weigh values of different types against one another (psychological incommensurability). But psychological difficulty doesn't show that there is no fact of the matter about how these values weigh against one another. Thus—the argument continues—insofar as the autonomy of a choice depends on the extent to which the choice coheres with one's values, whenever there isn't a precise tie between options, they could not both be chosen with equal levels of autonomy.

For the sake of argument, let's set aside metaphysical vagueness and incommensurability. Let's also grant, for the sake of argument, that therefore no two options could be chosen with *equal* levels of autonomy. Even so, there is still reason to believe in Options—that it's

[53] Supporters of value incommensurability would want to say of such cases that neither option is more valuable than the other and that the options are not of exactly the same value. Would they then want to insist that, strictly speaking, the options cannot be said to have *the same* autonomy-based merit, although it's also not true that either choice is *less* autonomous than the other? If so, they might object to the terminology of 'equal levels of autonomy', at least strictly speaking. I set aside this metaphysical question about levels of autonomy here, since it's not plausible that it affects the central normative issue of whether there is more than one option that can be chosen with sufficient autonomy for valid consent. For it to do so, we would have to endorse the implausible claim that whilst two options do not differ in autonomy-based merit, it's nevertheless the case that whilst one can be chosen with sufficient autonomy for valid consent, the other one cannot. This seems ad hoc, ill-motivated, and in tension with common examples of consent-giving.

often the case that an agent in a given situation facing a choice between two (or more) consent-requiring options could choose either possible option with sufficient autonomy. The reason is that there is a range of levels of autonomous decision-making all of which are above the level that is sufficient for valid consent, and the kinds of flaws of decision-making to which we are prone. Many of these flaws are sufficient to dispose us to choose different options, and to choose different options with lower levels of autonomy than we might have done, without thereby causing our decision to fall below levels of autonomy that are sufficient for valid consent.

For instance, go back to Izzy's choice between birth control options. In our example, Izzy was considering a choice between a hormonal IUD and a copper IUD. Imagine, not implausibly, that she's aware of a third option: a contraceptive implant. Further, imagine that, all-things-considered, the contraceptive implant, not one of the IUD's, would be the substantively optimal choice for her in terms of the overall satisfaction of her desires and values. But she doesn't stop to properly consider it because—well, because having something planted under your skin seems a bit icky, and she's more familiar with the idea of IUD's. Now, this rules out the possibility that she could pick an IUD with perfect autonomy. But we don't yet have reason to think that she couldn't choose one of these with sufficient autonomy, since we saw earlier that optimal decision-making could not plausibly be a requirement of valid consent.

So let's return to her consideration of these two options. We had previously stipulated that there was no precise way for her to weigh the advantage of increased comfort against the disadvantage of a higher monetary cost of an additional $75 associated with the hormonal IUD compared to the copper IUD. But let's get rid of this stipulation, and assume that there *is* a fact of the matter as to which she values more, and thus which option is the most

practically rational for her to choose, and that it's the hormonal IUD: in fact, the slight increase in long-term comfort is more important by her own lights. It follows from all we have assumed so far that this would be a sufficiently autonomous choice. But consider how flawed decision-making may lead her to choose the copper IUD instead:

> **Epistemic difficulty:** Despite sincere attempts at weighing all the pros and cons, it's difficult for her to introspect which of these features weighs more heavily with her, and she underestimates the value of a small increase in long-term comfort. So she plumps for the copper IUD for a short-term savings of $75.

Consider another flawed process of decision-making:

> **Neglecting reasons:** Although the information is available to her, she fails to give any serious consideration to the possible increase in comfort that the hormonal IUD could bring. So she simply does not dedicate any time to thinking about whether a chance of increasing her comfort in the long run, however slightly, is really a slightly weightier reason than the associated monetary cost. So she simply plumps for the copper IUD for a short-term savings of $75.

In both these cases, Izzy's decision-making is flawed: her decision-making process is procedurally suboptimal (she underweights or neglects a relevant reason, comfort, relative to another, cost), leading to a sub-optimal choice. Her decision is therefore less autonomous than it might have been had she more fully realized her desires and values through opting for the hormonal IUD instead. And even that decision would have been less autonomous relative to the optimal choice of the contraceptive implant, which she neglects to consider altogether. Nevertheless, in consenting to the copper IUD, she is still willingly consenting on the basis of a sufficient understanding of the consequences and on the basis of sufficiently good reasons—a desire to have a long-term but reversible solution to birth control. Albeit imperfect, these different processes and the resulting choices are sufficiently autonomous for consent. So Izzy has multiple birth control options that could be chosen with sufficient autonomy for valid consent. So the Options thesis is true of Izzy.

Furthermore, the kinds of considerations that make it the case that Options applies to the case of Izzy are mundane and ordinary features of values and decision-making that apply to many choices and many decision-makers. We thus have good reason to think that Options is true generally.

So Options means that, in general, there is a certain amount of wiggle room, within boundaries, for ways that decisions can be resolved by a given agent in a given situation, and for that decision to nevertheless yield valid consent. Of course, not everything goes, and there *are* boundaries on the types of decision-making and choices that can be considered adequate. The variations in the IUD example, despite suboptimalities in decision-making procedure and choice, all involved an agent making a voluntary choice to get an IUD on the basis of sufficient understanding and relevant desires and values that were truly the agent's own; we assumed that an important part of the explanation of her consent was that she understands that the IUD is a long-term but reversible contraceptive device, she wants to minimize chances of pregnancy for the time-being, and has no overwhelming reasons given her desires and values not to do so. Her decision may not have been sufficiently autonomous, say, if it was based on overwhelming irrational fear, coercive pressure, or significant misunderstanding of the medical implications of her options.

Because Options is true, and No Options is false, the mere fact of variability does not undermine consent in the case of framing effects. So, if a successful argument is to be made that framing effects undermine the validity of consent, a different strategy will be required.

7. The Entailment Argument and the Likelihood Argument

I have just argued that for many agents facing many choices, there is often more than one option that they could choose with sufficient autonomy in order to validly consent to that option. This means that it doesn't follow from the fact that an agent makes one decision, but could have made a different choice, that being subject to framing effects disposes agents to make choices with insufficient autonomy and thus to make choices that fail to constitute valid consent. This is because multiple decision-making processes and options can give rise to valid consent for a given agent facing a given choice.

But this defense of Options did not directly address the more specific situation where *frames* are causally operative in determining which option an agent chooses. So an argument that framing effects undermine consent may, then, appeal specifically to dispositional variability *caused by framing*, rather than only the variability of decision-making in itself.

One possibility here is that causal dependence on a framing effect *entails* that sufficient autonomy is undermined and that consent is invalid. Let's call this claim the "Entailment Principle". If we combine the Entailment Principle with an empirical premise that many choices are dependent on framing effects, we get the following argument:

The Entailment Argument

1. **Generalization:** Many cases of consent (that otherwise seemed valid) depend on framing.
2. **Entailment Principle:** If S's consent depends on framing, then S's consent is not valid.

3. **Conclusion:** Therefore, many cases of consent (that otherwise seemed valid) are not valid.

If you accept the Entailment Principle, then it seems the following reactions are available to this argument. First, you can accept the conclusion, leading to Consent Skepticism or Consent Pessimism, as discussed earlier. Alternatively, you can attempt to reject the Generalization premise: to deny that many cases of consent are dependent on framing effects, notwithstanding the empirical evidence of the type surveyed earlier. As discussed in Section 4, even if rejection of the Generalization premise were defensible, it would reduce the scale of the problem but not necessarily eliminate the problem entirely: the current state of the evidence makes it implausible that *no* cases of consent are affected by framing, and discovering that some decisions are being treated as valid consent when they should not be is worrying for the sake of those agents whom it affects, even if it does not have wider impact on our theorizing about consent.

The Entailment Principle gains prima facie intuitive support from cases like Framing-Induced Surgery—cases where dependence on framing effects seem, intuitively, to invalidate consent. However, despite the intuitive appeal of the Entailment Principle, it is false. We can see that it is false because it overgeneralizes.

Take, for instance, the following case:

> **Framed House Sale:** Matthew Mason's mother, Caroline Mason, recently died. Matthew inherited the old house in which Caroline had been living. It's the same house Matthew grew up in. In fact, as he well knows, Matthew's mother was born in that house. Although the house is very emotionally dear to him, and the house is beautiful and well-maintained, and he is financially stable, it's not a particularly convenient place for him and his family to live, and the money from the sale wouldn't hurt. So he puts plans in motion to sell the house.

> A contract is drawn up that includes a detailed description of the house and its history (including, near the beginning, the detail about his mother having been born there). At the bottom, Matthew must sign on a dotted line confirming that he "consents to the sale of [street address]". He does so.
>
> Suppose, however, that had the document instead required him to sign on a dotted line confirming he "consents to the sale of [street address], birthplace of Caroline Mason", he would not have signed, and would have discontinued the sale; this way of framing the personal significance of his decision would touch his emotions in such a way that he would not want to go through with it.

The actual and hypothetical versions of the contract are equally informative; they just contain the information about the mother's birth in different locations. Thus Matthew's decision to consent is contingent on mere framing in the same way as the patient whose consent is contingent on survival/mortality framing: Matthew is disposed to dissent if the very same information is presented to him in a different form. Whether the detail about the birth of his mother appears earlier or at the end of the contract does not affect how informative the contract is, Matthew's knowledge of the facts of the sale, or the terms of the contract. Consequently, according to the Entailment Principle, Matthew's consent to the sale of the house is invalid.

Intuitively, however, Matthew gives valid consent to the sale of the house, and the contract is morally (and, let's presume, legally) binding. Unlike Framing-Induced Surgery, the fact that Matthew's emotional resolve *might have* been weakened does not seem to invalidate the consent he actually gave. If Matthew later changed his mind and regretted the sale, it would rightly do no good for him to plead that the contract is void because he failed to be more gripped by the emotional salience of the loss at the moment of signing (although he might be justified in claiming the contract void in other situations, for instance if he had been directly misled into believing that it was a contract to lease the house rather than sell it, or that it concerned only the sale of the furniture).

Why don't we feel the same worry about Framed House Sale as we do about Framing-Induced Surgery? Plausibly, the reason is this: the extent to which Matthew's decision-making in House Sale is defective doesn't seem to dip below the level of sufficient autonomy required for valid consent. In part that is because the story involves explanations of the agent's behavior that are understandable in terms of the agent's reasons, so we need not resort to an explanation that involves a severe departure from understanding, his beliefs, or his desires.

To see this, let's explore the explanation of Matthew's decision-making. Why does he consent to the sale, even though he can be easily moved to refuse to do so if the emotional attachment he has to the house is made marginally more salient? It's not that he fails to understand that selling the house implies that he will be selling his mother's former home and place of birth, or forgets by the time he gets to the end of the contract. Rather, we think he consents to the sale because, although he cares deeply about this sentimental attachment, at that moment he decides to weight other considerations that he cares about (convenience and money) over that one. Moreover, although we cannot say with any certainty that this weighting is optimal, it seems sufficiently reasonable and in line with his stated values that we need not impute any severe defects of autonomy here.

Furthermore, the fact that Matthew's decision is subject to framing does not even entail that he is *disposed* to engage in decision-making that is insufficiently autonomous. Take the counterfactual in which the framing of the contract is such that he is gripped by his emotional attachment to the house and refuses to sign. Again, it might be that the refusal is not an optimally rational decision. It might involve a suboptimal weighting of some of his values (emotional attachment to the house) over others (convenience and money); it

might even be akratic (perhaps he knows he should really sell the house); it might even be based on some false beliefs (perhaps he irrationally overestimates the extent to which he will regret the sale). Still, the refusal is based on a good grasp of the meaning of his decision and on an authentically endorsed value of his own—the emotional attachment driving this decision is not some kind of alien force controlling his decision-making independently of his beliefs and values. So if he is non-autonomous, he is not severely so, not to the level that excludes one from being able to give authoritative consent.

That is not to say that his decision-making is free from defects, even if all the framing is doing is affecting which of his values he weighs more heavily at the moment of signing. For which consideration he weighs more heavily at that moment is determined by something independent of his reasons, namely the order of the wording on the contract. Many would argue that this entails that his decision-making process is not optimally rational. Indeed, many commentators believe that choice-dependent framing shows that we have inconsistent preferences.[54] Even if they are right, this does not imply that there is a sufficient level of irrationality to threaten consent.[55] So, Framed House Sale might be intuitively compelling as a case of valid consent precisely because it lacks any markers of deeply flawed agency, despite the dependence on framing.

The considerations that suggest that the Entailment Principle overgeneralizes are just like the considerations that led us to accept Options—the idea that a number of imperfect yet sufficiently autonomous ways of resolving choices are normally available—combined with

[54] E.g., Chwang, 2016.

[55] Many theorists think that framing effects entail *some* element of irrationality, such as the possession of an incomplete or inconsistent set of preferences. I take cases such as House Sale to show that that level of irrationality is not sufficient to invalidate consent.

the observation that we can be disposed towards one or another of these options due to a frame, without that further impugning the quality of one's decision-making. Although these sorts of instabilities in decision-making can be due to severe failures of rationality, they can also be generated by far less dramatic, more quotidian failures. Even before we had behavioral economics and contemporary cognitive psychology, we knew that our wills are not steadfast and constant; that our emotions are disposed to fluctuate; that judgments are disposed to be revised despite an absence of new evidence. These facts are signs of suboptimal agency, but of a quotidian sort, not the kind severe defects in rational and free choice that fall below the required bar for valid consent. The problem for the Entailment Principle, then, is that these kinds of human dispositional traits do not in general seem to be in tension with the validity of consent (plausibly, because the defects involved do not show that we lack a *sufficient* level of autonomy to give valid consent) even though they can be functions of equivalent variations in context or presentation—that is, of frames.

So cases like Framed House Sale suggest that we should reject the Entailment Principle as a way of capturing what we find troubling in cases like Framing-Induced Surgery.[56] Furthermore, Framed House Sale seems to be one such example amongst many. The Entailment Principle cannot adequately explain the difference between such cases and contingencies that potentially undermine consent. Therefore, the Entailment Argument for the impact of framing effects on the Standard View of consent fails.

[56] This counterexample also suggests that the Entailment Principle is false and cannot be the correct explanation of why consent is invalid in other particular cases. For instance, Hanna (2011) claims that consent is not valid in cases where agent consents *only* because they lack relevant information that is material to their decision (for instance, when a Jehovah's Witness consents to a transplant without realizing that this involves a blood transfusion). He goes on to diagnose these cases by invoking a principle which says that if the agent would dissent under an *equally* or *more* informative frame, then the consent is invalid. Even if Hanna is right that consent is not valid in these cases, this principle cannot be the correct diagnosis; it's a more expansive version of the Entailment Principle and suffers from the same counterexamples.

But a similar argument can be generated with a slightly more modest claim. Even if framing effects do not entail that consent is invalid, they might make it *likely* that consent is invalid. This more modest claim can be used to generate an argument that is very similar to the Entailment Argument, as follows:

The Likelihood Argument

1. **Generalization:** Many cases of consent (that otherwise seemed valid) depend on framing.
2. **Likelihood Principle:** If S's consent depends on framing, then it's likely that S's consent is not valid.
3. **Conclusion:** Therefore, many cases of consent (that otherwise seemed valid) are likely not valid.

The conclusion of the Likelihood Argument is only slightly more modest than the conclusion of the Entailment Argument, and accepting the conclusion that many cases of consent are *likely* not valid still seems sufficient to force us to give up some component of the Standard View of the ethics of consent. In particular, just like the Entailment Argument, accepting this conclusion seems to force us to choose between a modestly revised version of Consent Pessimism (many cases of apparently permissible consent transactions are in fact *likely* morally wrong) and Consent Skepticism (valid consent is not, in fact, normally required to make acts that appeared to require consent morally permissible).

Why might framing effects make it likely that consent is invalid, even if they do not entail it? This would be so if framing effects make it likely that there is a violation of Sufficient

Autonomy – that the agent is falling short of cognitive requirements of autonomous decision-making (sufficient understanding, knowledge, theoretical rationality) or conative requirements of autonomous decision-making (sufficiently practically rational decision-making based on the agent's reasons, desires or values). For instance, framing effects (such as mortality/survival framing effects on medical decision-making) could be *good evidence* that an agent is not meeting standards of Sufficient Autonomy if framing effects are *often* symptoms or causes of substantially non-autonomous processes.

If so, the Likelihood Argument has the resources to avoid the overgeneralization problem faced by the Entailment Principle, while capturing intuitive worries about cases like Framing-Induced Surgery. It thus shows more promise than the Entailment Argument. Furthermore, a proponent of the Likelihood Argument might suggest that although it's *possible* to give valid consent that is dependent on framing, as in Framed House Sale, this kind of autonomy-preserving story is not likely to be behind most framing effects of the sort that we have discovered in the lab, such as survival/mortality framing. For the autonomy-preserving explanation of the effect of framing on Matthew's decisions in Framed House Sale involves appealing to aspects of his choice situation that are highly peculiar to him and to his specific situation. But framing effects have been shown to affect randomized pools of unrelated participants assessing the very same framed choice-sets as one another, and to do so when applied to a number of different contexts. It's highly unlikely—the proponent of the Likelihood Argument might argue—that all of these participants have reasons and values that are balanced and lined up *just so,* so there are readily available reasons-based, autonomy-preserving explanations for *each* participant's propensity to be affected by framing in response to some particular choice-set that has been artificially dreamed up by the experimenter. Consequently, it might be argued, it's likely that many framing effects involve decision-making processes that are not of the

sufficiently autonomous kind. And, therefore, cases of consent that are dependent on framing effects are likely to be invalid.

I am going to argue that the Likelihood Principle is false. My strategy will involve considering the major existing explanations of attribute framing in the psychology and decision-making literature. The first explanation I'll consider is the 'information leakage' theory. This will be the subject of the section to follow. As we will see, this theory and supporting evidence suggests that framing effects do not themselves make it likely that agents are engaged in even sub-optimal decision-making, let alone insufficiently autonomous decision-making of the sort that could plausibly undermine consent. Afterwards I'll move to consider alternative approaches.

8. Information Leakage and the Likelihood Argument

8.a. The Information Leakage Theory of Framing Effects

According to the "information leakage" approach argued for by Sher, McKenzie and colleagues,[57] framing effects are explained by rational inferences listeners make based on information that is conveyed by which frame a speaker uses.

We can break down this account into two main claims. First, frames are not informationally equivalent: the way a speaker frames information licenses different inferences on behalf of the listener. Second, listeners in fact make accordingly different inferences as a result of frames, and this explains framing effects.

[57] E.g., Sher and Mckenzie, 2006, 2008.

Before expanding on these claims and evidence that supports them, let's clarify what these claims would show about framing effects, if correct. In line with standard definitions, recall that I defined framing effects in terms of a choice being affected by differences in frames that do not themselves differ in informational content. If one sticks strictly to such a definition, and if the information conveyed by a frame is to be understood as part of its informational content, then the information leakage account implies that the behavioral patterns in question do not constitute framing effects after all. It thus defeats much of the evidence in favor of the claim that framing effects occur. Alternatively, the information leakage account may be better understood as leading us to revise our understanding of what framing effects are: while the examples in which we were interested seemed to concern equivalent descriptions, it turns out that they are not informationally equivalent after all. I find it a more natural use of terminology to adopt the second approach and to take the information leakage theory as showing us something new about what we already dubbed 'framing effects. For this reason, and for reasons of simplicity, I will use this way of speaking when discussing the implications of information leakage theory. However, none of my arguments ultimately turn on this definitional issue (nor on the question of whether the information ought to be understood as having its source in the frame's content, on a proper analysis of "content"); the interested reader may substitute my talk of "framing effects" here to "putative framing effects" or "so-called framing effects" as they please.

Information leakage theory has the potential implication that framing effects do not, in fact, undermine the autonomy of consent. What does this theory claim, and what evidence is there to support its claims? Let's begin with the first major claim, that concerning the denial of informational equivalence. According to information leakage theory, although

phrases like "20% chance of death" and "80% chance of survival" seem informationally equivalent, this is often not the case when such phrases are uttered by a speaker in communicative contexts. This is because in communicative contexts—such as the context of a doctor speaking with a patient about her options—which frame the speaker uses licenses further inferences about the background beliefs and assumptions of the speaker (Sher and Mckenzie, 2006). One way of putting this is that although "20% chance of death" and "80% chance of survival" are logically equivalent in internal content, "The doctor said that there is a 20% chance of death" and "The doctor said that there is an 80% chance of survival" are not logically or informationally equivalent—there is pragmatically conveyed informational content as well.

Why would this be? The reason is that speakers don't select which frame to use at random. Instead, which frame they use depends on a number of factors related to their background beliefs and attitudes, the conversational context, and what they want to convey. When speakers take a property to be greater than a relevant reference point (or particularly representative of the object, more intrinsically notable, or more pragmatically consequential given the context), they will adjust their choice of language accordingly.[58] For instance, speakers are more likely to describe a glass that was previously empty as now being "half full", whereas they are more likely to describe a glass that was previously full as "half empty". Frame selection might also convey evaluative information. For instance, speakers are more likely to describe the success rates of a research team made up of talented, valiant, impressive, and highly-qualified individuals in terms of their rates of success, but a team made up of uninspired and not highly qualified individuals in terms of

[58] E.g., Sher & McKenzie, 2008, pp.87-88. Holleman and Pander Maat (2009) especially emphasize the speaker's intended implications.

their failure rates (even if the success/failure rate of each team is numerically equivalent).[59] In each case, the speakers' selection of a frame is intended to communicate something about the object being described: in the case of the water, whether the level is going up or down; in the case of the team, whether the speaker thinks well or badly of their capabilities.

In this way, frames are selected according to whether or not the speaker intends to express favorable or unfavorable attitudes towards the object. The same is true in medical contexts. For instance, people are more likely to describe a treatment with a 50% mortality rate in terms of mortality when it is being compared with an alternative treatment with a better survival rate, and in terms of survival when it is being compared with an alternative treatment with a worse survival rate (McKenzie & Nelson, 2003). Similarly, if a new cancer treatment is evaluated positively, speakers are more inclined to frame it in terms of its survival rate; if the same treatment is evaluated negatively, speakers are more inclined to frame it in terms of its mortality rate (Holleman & Pander Maat, 2009). That is, a "mortality" frame is used when the treatment's riskiness is comparatively worse and the treatment is otherwise assessed negatively, and a "survival" frame when the treatment's riskiness is comparatively better and when the treatment is otherwise assessed positively. So we can expect doctors to vary their choice of mortality/survival frame depending on their background beliefs and attitudes regarding the treatment in question, such as whether they think the procedure is to be recommended overall, and how it compares to other options open to the patient.[60]

[59] Sher & McKenzie, 2006, Exp. 5.

[60] Evidence suggests that there is no difference between patients and doctors in how they relate different framings of mortality/survival rates to correspondingly different conclusions regarding a treatment's effectiveness e.g., Perneger & Agoritsas 2011.

The fact that speakers alter their language in these ways means that which frame is used "leaks" information (about what the speaker recommends or their background information) to the decision-maker, making it rational for the decision-maker to respond differently to different frames.

This brings us to the second main claim of the information leakage account: that listeners do in fact make accordingly different inferences as a result of frames, and this causes and explains framing effects. For instance, participants infer that a glass described as "half empty" used to be full, whereas a glass described as "half full" used to be empty.[61] When asked to estimate what proportion of shots typical high school basketball players get in, most participants who are told about an "unusual" player who "makes 40% of his shots" provide estimates under 40% for the typical player, whereas most participants who are told about an "unusual" basketball player who "misses 60% of his shots" provide estimates higher than 40% for typical players.[62]

Similar inferences occur about the background values of the speaker in the medical context. For instance, in a recent set of studies by Altay and Mercier (2020), participants were asked whether a statement about vaccines was more likely to have been uttered by a pro- or anti-vaccination individual. Whereas 99% of participants attributed the positively framed statement "999 out of 1,000 don't have any severe side effects" to a pro-vaccination speaker, only 46% of participants did so for the negatively framed "1 individual out of 1,000 has some severe side effects".[63]

[61] Sher & McKenzie, 2006, Exp.1, Exp.2; McKenzie & Nelson, 2003.

[62] Leong et al., 2017.

[63] Altay and Mercier, 2020, Experiment 1. The same study found a similar effect for "90% of medical scientists think that vaccines are safe" vs. "10% of medical scientists don't think that vaccines are safe".

Furthermore, there is some evidence that framing effects are greatly attenuated if the relevant inferences are blocked, suggesting that these inferences play a causal role in framing effects. For instance, the inferences can be blocked by pre-existing background knowledge: framing effects on the evaluation of NBA basketball players who "make x%" or "miss 100-x%" of free throws are greatly attenuated for participants who are very knowledgeable about NBA basketball. This is presumably because they possess background knowledge related to typical performance that blocks the relevant inferences based on the frame, even though those same participants show typical framing effects when evaluating the effectiveness of a fictitious medical treatment described in terms of survival or mortality rates (presumably because they have no background knowledge that would block frame-based inferences here).[64] There is even some preliminary evidence that frame-based inferences can be blocked based on whether the listener has reason to expect the speaker to be a cooperative communicator in the context at hand.[65]

8.b. Applying Information Leakage to Medical Consent

If the information leakage theory of attribute framing effects is right, we should expect patients to make choice-relevant inferences from frames in cases where they are considering whether to undergo a treatment or procedure. For instance, patients might infer that a doctor using a mortality frame believes that a 10% mortality rate is undesirable or unusually risky under these circumstances and that they do not recommend for that reason. On the other hand, patients might infer that a doctor using a survival frame believes

[64] Leong et al., 2017.

[65] Leong, unpublished manuscript, 2020, Chapter 2.

that an 90% chance of survival are good odds that are worth taking under the circumstances.

Indeed, there are special features of the relationship between doctors and patients which should make the medical context a particularly fertile ground for facilitating and justifying the use of frame-based inferences such as these in decision-making. This is because patients often seek guidance and recommendations from doctors to help inform medical decisions, especially in light of the doctor's greater knowledge and experience of the treatments or procedures in question, and the norms of beneficence governing the doctors' relationship to the patient. (It is thus not surprising that, in some studies of actual patient decision-making, the vast majority of patients have been shown to choose their doctor's primary treatment recommendation.[66]) As patients we seek information and advice from a physician, a medical expert who has significantly greater knowledge and experience of medical issues than we do, and we often trust them to use this expert knowledge in order to help us make choices which are best for us.[67] This is so even in cultural contexts, like the U.S., where there is an expectation that patients retain autonomy in medical decisions and engage in shared decision-making with physicians.[68] We should expect this feature of the medical context to affect the conversational norms which govern how speech is used and interpreted. For instance, we tend to assume that the speech of the physician will be governed by norms of truthfulness, by a goal to disclose choice-relevant information, and a norm of beneficence, a goal to advise and guide the patient so that they are able to make choices which promote their welfare.

[66] e.g., Siminoff & Fetting, 1991.

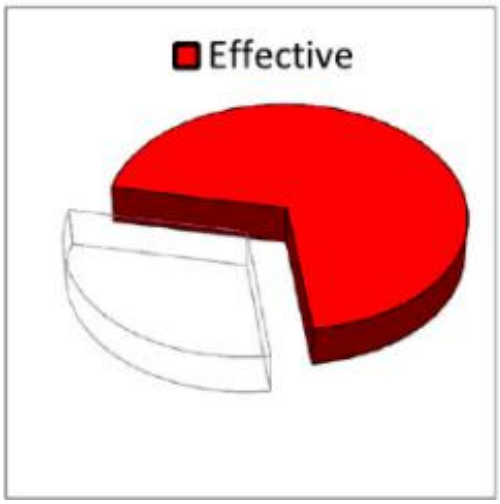
[67] E.g., Joffe & Truog, 2010; Douglas & Proudfoot, 2013.

[68] In some other countries, such as Iran, cultural norms involve a greater expectation that the doctor-patient relationship is paternalistic and attributes greater authority to doctors (Tabesh, Tabesh, & Moghaddam, 2019).

So in giving recommendations we should expect the speech of physicians to be shaped by their background knowledge and their medical recommendations. It would not then be surprising if patients interpreted the speech of doctors in this light, nor if mortality/survival frames could influence decisions via providing information about the physician's assumptions about and evaluations of the treatment in question.

The hypothesis is easily extended to other types of framing. Imagine that there is a local HPV vaccination drive. You tell your doctor you don't know much about the HPV vaccine, and you're wondering whether you should get it. They say they've compiled an information leaflet about it. They give one to you to take away and read. Included in the information leaflet is the following image:[69]

The vaccine is effective against HPV types that cause a certain proportion of cervical cancers, as described in the figure:

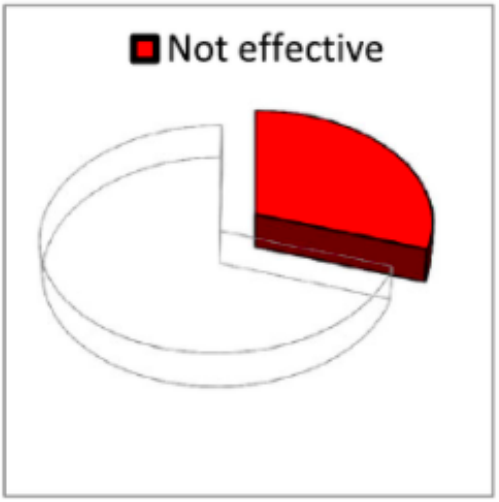


What would you think your doctor's opinion is about whether or not getting the vaccine is worthwhile, if you had to guess?

[69] Images taken from Kreiner and Gamliel, 2017.

Now imagine that the leaflet had instead contained the following image:

The vaccine is not effective against HPV types that cause a certain proportion of cervical cancers, as described in the figure:



Now, having seen *this* image, what would you think your doctor's opinion is about whether or not getting the vaccine is worthwhile?

I would bet that your credence that your doctor believes the HPV vaccine is worth having would be higher in the second case than in the first. Although these images are nearly identical, and although they communicate the exact same numerical fact—they both illustrate the proportion of HPV that is prevented by the vaccine, and the smaller proportion that is not—they are clearly not informationally equivalent. As the idiom would have it, a picture speaks a thousand words, and more information is communicated here than a bare numerical proportion, given the way that color and phrasing are used to communicate emphasis, and given that the act of emphasis is a familiar communicative device—one that can be achieved visually as well as with the tools of language. And, although the communicative inferences I am predicting have not been tested, Kreiner and

Gamliel (2017) used these images to test for framing effects, and indeed their evidence suggests that the negatively framed image led participants to rate the vaccine as less effective and to be less likely to recommend the vaccine to a friend.[70] Moreover, this framing effect was found irrespective of participants' numeracy levels.[71] This is consistent with the hypothesis that the effect of framing on consent and related decisions can be driven by more domain-general processes by which we make inferences about background information based on how the information is framed.

8.c. Information Leakage: Implications for the Likelihood Argument

To the extent that framing effects are explained by information leakage, the fact that a particular decision is frame-dependent does not make it likely that the decision is insufficiently autonomous for valid consent, any more than responding to a piece of information or a recommendation makes it likely that a decision is insufficiently autonomous. In fact, to the extent that the framing effect is explained in terms of accurate inferences about relevant information, frame-dependence does not even show that the decision is being made in a sub-optimal way, any more than other ways of inferring and using relevant information in forming a decision does. If we take the information-leakage hypothesis seriously, framing effects are simply evidence that patients are disposed to alter their decisions based on pragmatically inferred information. Patients' decisions, then, might be determined by endorsed values they have with regards to this inferred information—such as a reflectively endorsed desire to trust their physician's

[70] I use the weaker "suggests" instead of "shows" only because the HPV vaccine was one of three scenarios used in the study, and the authors averaged across the scenarios when reporting results "to avoid redundancy" (Kreiner & Gamliel, 2017, p.778).

[71] i.e. no significant interaction between numeracy and framing was found.

recommendations. Additionally, the decisions might be determined by more general endorsed values, such as the desire to make choices which balance reasonable risk with quality of life, and the information inferred from the framing (e.g., about the reasonableness of the risk and the quality of the procedure) tells the patient something about which outcome is more likely to satisfy those desires and values.

9. A Model of Sufficient Autonomy for Variable Consent

To the extent that framing effects are explained by information leakage—and there is some convincing evidence supporting that theory—the fact that a particular decision is frame-dependent does not make it likely that the decision is insufficiently autonomous for consent.

Nevertheless, there are some scientific explanations of framing effects that do not involve the rational inference of relevant information, and whether all or even many framing effects are to be explained in terms of information leakage remains a matter of scientific controversy. It might be objected, therefore, that my criticism of the Likelihood Argument so far hinges on a contested scientific claim. In particular, while we have good evidence that information leakage is responsible for some framing effects in some cases, it's plausible that it is not responsible for all framing effects. To the extent that frame-dependent decisions are more frequently best explained by mechanisms other than information-leakage, and to the extent that these other mechanisms bypass or significantly obstruct autonomous decision-making, it remains possible that frame-dependent decisions in general are likely to be non-autonomous.

So we will turn, momentarily, to consider alternative accounts of framing effects—ones that do not assume that frames implicitly communicate informational content—and evaluate the Likelihood Argument in light of those accounts. But before that, we will need to revisit the question of what standards of decision-making need to be met in order to qualify as sufficiently autonomous, even if the decision-making in question is suboptimal and prone to vary. Developing a more detailed framework for evaluating whether variable decision-making qualifies as sufficiently autonomous will provide us with more precise standards with which we can evaluate the picture of decision-making offered by alternative models of frame-dependent decision-making.

Recall that, in Section 6, I argued for the Options thesis: for an agent facing a given choice in a given situation, there can be multiple ways of resolving the choice that would result in a sufficiently autonomous decision. An example of this was the case of Izzy, who was choosing between different kinds of long-term contraceptive, using imperfect, yet sufficiently autonomous, ways of resolving this choice. Moreover, in Section 7, we saw, using the case of Matt the house-seller as an example, that ordinary people engaging in ordinary modes of resolving decisions can be prompted towards one or another of these sufficiently autonomous options by normatively irrelevant features of frames; the possibility that decisions could be variable in this way without leading to an insufficiently autonomous choice led to the rejection of the Entailment Principle, the idea that making a frame-dependent choice entails that the choice is insufficiently autonomous.

These claims apply generally, for many decision-makers facing many choices. Just like Izzy making a choice about contraceptives, or Matt making a choice about selling his house, there are often many ways of resolving a choice in a suboptimal manner—involving epistemically suboptimal reasoning about the choice or about one's own values, or

practically suboptimal decision-making in terms of one's responsiveness to one's desires and values—that affect which option is chosen, but are nevertheless sufficiently autonomous for ordinary, valid consent. For instance, competent decision-makers commonly fail to consider all relevant reasons when making a choice, or fail to adequately take account of the full weight of all reasons; they prioritize certain desires or values just because they seem more salient or pressing at the time, even if a bit of soul-searching would reveal that this is short-sighted or neglects an important value; or sometimes there are too many competing considerations to consider and they plump for a reasonable choice just to make a decision so long as it seems to be good enough given the deliberation they've engaged in so far. Being prone to these flaws, and to suffering from these flaws in different ways (e.g., either neglecting *that* reason or *this* one), leads to multiple possibilities for how people resolve their decisions. Yet, so long as they still make their decision in a sufficiently rational way based on sufficient understanding and relevant desires and values, and the choice is not far away from what their desires and values determine they ought to do, such suboptimalities are not normally enough to vitiate consent.

We now are now in a position to extract conditions of sufficiently autonomous decision-making in a more explicit and precisely stated manner. This will allow us, in the sections that follow, to evaluate whether framing effects *tend* to cause agents to decide in insufficiently autonomous ways, even if they do not entail this. Because this is our ultimate goal, I won't argue that these conditions are all necessary for a decision to be sufficiently autonomous. It is better that we treat them as necessary at least for the sake of argument, making us more confident that *if* a decision meets the conditions, then it is in fact sufficiently autonomous.

Before we begin, there's an important point to make regarding the scope of the conditions and the model of decision-making that I'll be using. The model focuses exclusively on conditions that apply to the agent's decision-making as it affects the autonomy of their decision. My intention, then, is that we can treat the conditions as jointly sufficient for sufficient autonomy, *all else being equal,* where things might not be equal due to facts about the agent's circumstances that are independent of his decision-making (i.e. independent of his psychological states such as beliefs, desires, and values, his decision-making processes, and his decision-making capacities). For instance, some might think that being coerced undermines autonomy and invalidates consent even if the coercion does not affect the agent's decision-making process or abilities; if so, it would be possible for an agent to meet the conditions specified by the following model and yet still fail to make a sufficiently autonomous decision. Even if this is correct, such issues are not plausibly relevant to the question of whether *framing effects* undermine consent, which specifically concerns whether the agent's decision-making processes and capacities meets certain standards. The model, then, focuses exclusively on the latter, and sets independent issues, such as coercion, aside.

So let's start with a simplified model of how agents make decisions between two or more options. Options have various attributes that realize values to different degrees. For instance, possible medical treatments have attributes like mortality risk, various types of comfort and discomfort, various types of side-effects, cost, convenience, etc. According to this simplified model, agents make decisions by applying weights to different attributes that are possessed by the different options. The agent assigns a single weight to an attribute based on a product of two factors: the extent to which options are perceived to possess the attribute, and the amount that the agent takes themselves to value that attribute. For instance, an agent might place great weight on the fact that some treatment will rid them of most of their pain, a little weight on the inconvenience of the monthly trip to the

hospital required by this treatment (they disvalue the inconvenience but the amount of inconvenience involved is not high), etc.[72] The balance of weights then determines the choice: the option with the highest total is selected. (This model captures what we might alternatively think of as the weighing up of different reasons, or the weighing up of different values that are instantiated by these attributes.)

For any given decision an agent faces, there are a range of different decision-making processes that could be used to resolve the decision, distinguished by the set of weights accorded to different attributes. In this model, an attribute accorded no weight is one that does not effectively feature in the agent's decision-making process at all. On the other hand, attributes that do feature in the agent's decision-making could be given greater or lesser weight. For instance, instead of placing great weight on the pain-reducing qualities of a treatment, an agent could instead place great weight on the disvalue of inconvenient monthly trips, and little weight on the reduction in pain; this weighting might lead them to choose a less effective but more convenient option. So agents resolve decisions by, in effect, 'selecting' one possible decision-making process—that is, assigning one possible combination of weights for different attributes, and basing their decision on these weightings, thus determining which option is chosen.[73]

[72] In some cases, as we will see momentarily, it's clear when these two factors diverge – for example, when an agent mistakenly thinks that something they value greatly is not realized at all by some option they are considering. In many cases, however, it will not be clear how a weight breaks down into these two factors, and so it's more useful to simply consider the overall weight an agent applies to an attribute at hand. For example, in a case where an agent places great weight on the significant pain-reducing qualities of a treatment, we might say that they greatly value large reductions in pain, where this is realized to a moderate extent for the treatment in question; or we might say that the agent somewhat values reductions in pain, but this is realized to a great extent by the treatment in question. Questions of how such cases ought best to be interpreted will not be pertinent to the discussion that follows.

[73] For the purposes of the model, we can talk simply in terms of values, although of course cognitive states like beliefs must feature in decision-making as well. Basing a decision on a

This simplified model abstracts away from questions about the extent to which these steps are implicit or explicit, unconscious or conscious, and so on; instead, it captures the functional contours of how the choice is resolved. Assigning a weight to an attribute may be realized consciously or unconsciously, explicitly or implicitly. So to say that an agent weighs some attribute heavily does not necessarily mean that they consciously attend to it and decide to assign it great importance (although of course they *could* do so)—the weight given to this attribute could be realized entirely implicitly. For instance, imagine I offer you either some Aspirin or some Tylenol for your headache, and you're choosing which to take. It never consciously occurs to you that one benefit possessed by both options is that neither will cause instantaneous death. Still, you give the fact that neither Aspirin nor Tylenol will cause instantaneous death a lot of weight—only implicitly. We would have seen that your decision-making process involves assigning a lot of weight to this factor if I had also presented you with you a bottle labelled 'POISON': the fact that you so heavily weigh avoiding instantaneous death means there is no chance of you picking that option over Aspirin or Tylenol.

Thus the relevant sense of assigning weights to different attributes is a functional one. For instance, say an agent is choosing between a palliative and a non-palliative treatment option for some serious illness, where the palliative option is superior in relieving pain, and the non-palliative option extends life for longer. Let's say that, in resolving this choice, the agent assigns heavier weight to the value of pain-reduction than to the value of life-extension. To say that the agent places greater weight on the pain-reducing qualities of a treatment *is* for that attribute to count more strongly in deciding between treatments, and

value as I use that concept here can be read as shorthand for 'basing a decision on a value and the belief that the decision realizes that value', or something to that effect.

thus it is to say that the agent will select a pain-reducing option unless comparably weighty considerations count against it. So it means that, in this case, the agent will pick the palliative option (unless there is some other set of attributes possessed by the non-palliative option that have not been mentioned here on which they place equal or greater weight). Moreover, to say that the agent places greater weight on the value of pain-reduction in resolving this choice means that if the palliative option had in fact been *inferior* in relieving pain than some *third* option, the agent would have picked the third option instead (again, all else being equal).

We can now distinguish between better and worse decision-making processes. In the ideal case, the weights assigned to different attributes perfectly reflect (1) the agent's background values—that is, how much the agent in fact cares about the different attributes in question, and (2) the presence of the attributes in the world—that is, the extent to which the attributes are in fact possessed by the options in question. We can say this ideal case is one in which "all attributes are perfectly weighted according to the agent's values", where a perfect weighting reflects both the agent's background values and the application of the attributes to the options at hand. For instance, ideally, the weight that the agent assigns to pain-reduction relative to life extension when they make their decision perfectly reflects how much more (or less) the agent in fact values the amount of pain-reduction at issue in their choice compared to how much they value life extension. (We say 'perfectly', but not 'precisely', since, as discussed earlier, values may or may not admit of precise comparisons. So weighing pain-reduction much more heavily than extension to life may be said to perfectly correspond to the much greater value an agent places on pain-reduction relative to life extension, even if this comparison is not a precise one. It is perfect in the sense that it could not *better* reflect the relative importance of the agent's values.)

Of course, we may never satisfy this ideal of autonomous decision-making. But we need only approximate this ideal to a sufficient degree in order for a decision to be sufficiently autonomous. I propose two conditions that, if met, suggest that an agent is making a sufficiently autonomous decision. The first condition is this:

> **Reasonable weighing of values:** the agent accords sufficiently reasonable weights to different attributes in making their decision (where the reasonableness of the weight is determined by how well it corresponds to how much the agent values the attribute in question).

This condition requires that the agent at least sufficiently approximates the ideal of perfectly weighing all of the attributes that are relevant to her decision; it requires that she does not weigh her reasons in a way that deviates excessively from their appropriate weights, in light of her values. This has the result that the choice she makes is what we might call 'within reason': it is sufficiently reasonable in light of her values.

One way that an agent can fail to use an ideal decision-making process and yet satisfy this condition is if they neglect or give insufficient weight to some reasons, but those reasons are not very important ones. Recall the 'Neglecting Reasons' case from Section 6. In this case, a woman chooses to have a copper IUD on the basis that it will satisfy her desire to have a long-term, safe solution to birth control while also offering her a savings of $75 compared to a hormonal IUD; yet she fails to consider the value of the increased comfort that the hormonal IUD might provide for her. Her decision is based on a reasonable weighting of her values, even though her decision underweights some values. Although that agent neglects the value of the slight increased comfort of the hormonal IUD—thus underweighting it—this error is not excessive, for the value is only slight, and consequently she still chooses a very reasonable option in the sense that it satisfies her most important values to a significant degree. An example of a failure to satisfy this condition would be if

her underweighting of comfort-related considerations was so severe that she would choose a slightly cheaper IUD even if past experience with it told her this particular method causes her terrible pain, whereas the slightly more expensive one is likely to cause no discomfort and in fact to increase comfort somewhat in the long term. This would arguably be a case of severe underweighting of a value that leads to a choice that is not within reason. But if an agent accords reasonable weight to their values, this makes it unlikely that their decision privileges minor values over very important ones.

Here's another example. Say an agent is deciding between two treatment options, where one has unpleasant side effects including nausea (let's call it Nausea Treatment), but the other is associated with a slightly lower survival rate (let's call it Risky Treatment) although it has no such side effects. If the amount he in fact disvalues suffering nausea is not too different from the amount he disvalues a slight mortality risk, it may be simultaneously true that it would be sufficiently reasonable to weigh the disvalue of an extra 10% risk more heavily than the disvalue of nausea—thus choosing Nausea Treatment—and also true that it would be sufficiently reasonable to weigh the disvalue of an extra 10% risk less heavily than the disvalue of nausea—thus choosing Risky Treatment. Thus either way of resolving the choice would be sufficiently reasonable for this agent (even if one would be better than the other). By contrast, it might have been insufficiently reasonable to weigh nausea more heavily than a very large increase in mortality, thus choosing a third option—Very Risky Treatment—instead. The relative weight placed on nausea compared to mortality risk would be too great here, too out of line with how much he in fact disvalues nausea compared to mortality risks.

Many things can affect the weights an agent assigns to different attributes, and thus how far they deviate from the ideal decision-making process. For example, agents can fail to

attend to some of their values; some attributes can loom larger than others because they are more striking, more salient, or more immediate, making them seem more important than they are; it can be epistemically difficult for to determine how much different attributes are valued relative to one another. Yet such influences can be compatible with choices that remain *sufficiently* reasonable in light of the agent's values. In Section 7, for instance, we saw the House Sale example of how an agent might be caused, due to an arbitrary feature of how a contract is presented, to weight an emotional attachment more heavily than practical considerations at the moment of decision, and thus refuse to sell. Yet, he still, intuitively, makes a choice that is sufficiently autonomous; for the emotional attachment is something he truly values to an extent that approaches the importance of the practical considerations that count in favor of selling, and thus the weights he gives to these considerations in making his decision remain broadly consistent with his values.

Importantly, however, this condition also excludes a wide range of potentially troublesome cases that might be thought to involve insufficiently autonomous decision-making.

Firstly, meeting this condition suggests that the agent is making a value-based choice. This is because this condition excludes those cases where the agent gives no values any weight in their decision-making, where their decision is best explained in a way that is independent of their values, such as reflexes—cases that are widely thought to be distinguished from reasons-based autonomous choice. It excludes such cases because placing little or no weight on any of the attributes one values is not sufficiently reasonable. On the other hand, it allows cases where an agent bases their decision on a mere subset of their values (for instance, an agent who values living a long life may, on that basis, choose surgery to treat an otherwise lethal cancer).

Secondly, it requires that although the agent may not weigh her reasons exactly as she should, she does not give overwhelmingly important reasons little or no weight in her decision, nor does she accord excessive weight to very minor reasons. This allows us to rule out cases where, although the decision is value-based, the agent neglects or seriously underweights their most important values when making a choice. This includes cases, for instance, where one value 'swamps' decision-making at the expense of the consideration of other important values. Take the example of overwhelming fear. Say that you are facing two medical options, one surgical and one non-surgical. The thought of surgery induces overwhelming fear; on that basis you choose the non-surgical option. Acting on the basis of this fear may well be to act on the basis of *a* value. But, insofar as this fear is overwhelming, there may be many values that bear on the decision which are nevertheless accorded no weight whatsoever—too many, and of too much importance, for them to be acceptably neglected. Such a case is ruled out by the requirement that the agent accords sufficiently reasonable weights to different values in making their decision.

Thirdly, because the condition requires that the agent bases their decision on a reasonable weighting of values given the actual strength of their own values (and disvalues), it's going to rule out cases where the agent bases their decision on values that are, in an important sense, alien, or not truly their own—such as choosing an option because it's what someone else wants, even though the agent doesn't want that option (and doesn't strongly value the fact that it's what this other person wants); or choosing an option based on a sudden urge or fleeting emotion that is significantly out of line with the agent's values.

Fourthly, it's going to rule out cases where serious misunderstanding is material to the agent's decision. There is much debate over what kinds and levels of misunderstanding are so serious so as to invalidate consent, if it does. Of plausible relevance is whether it involves

misunderstanding of the very type of act being consented to in a coarse-grained way (e.g., believing one is consenting to a medical examination rather than sex; to surgery rather than chemotherapy) and whether the misunderstanding concerns something very important to the agent so that the misunderstanding is material to the agent's decision, especially if the person asking for consent knows about the misunderstanding.[74] I will not enter this debate here, as my later argument won't rely on a precise analysis of this phenomenon, but at least for the sake of argument we can treat cases of consent based on serious misunderstanding as insufficiently autonomous for valid consent. Such cases are ruled out by the requirement that the agent places sufficiently reasonable weights on different values: even attributes that are very important are given no weight in the agent's decision if they are not known about, thus if it is something that the agent cares about very strongly, then they end up

[74] In fact, I do not believe that misunderstanding that is material to an agent's decision can, in fact, be sufficient to invalidate consent in and of itself. For instance, it's surely the case that many medical procedures of, say, the 18th Century were based on serious misunderstanding of biology and illness. Moreover, these misunderstandings were material to people's decisions: if they had known what is now known by modern medicine, many would, I presume, not have consented. Yet it seems rather ludicrous to say that none of these medical decisions were consensual—at least, not for this reason. A case like this—as well as the case of consenting to an examination under false pretenses—is ruled out by the condition under discussion if we define the 'reasonableness' of a weight in part objectively, in terms of whether a value is in fact realized by an option, as I have done to rule out sex-by-deception cases. If instead we define 'reasonableness' subjectively in terms of what it is subjectively reasonable for an agent to believe about whether various values are realized by an option, then the case of 18th Century medicine is no longer ruled out; but neither is the sex-by-deception case, which is a more plausible candidate for invalid consent. It might be that the category of 'serious' misunderstanding is in fact better resolved by conditions on consent that are not to do with the quality of the consenter's decision-making at all—in fact, not to do, in and of itself, with the quality of the consenter's understanding—but rather to do with the *consent-seeker* and their relationship to, or ethical obligations towards, the consenter. This bears the promise of distinguishing between the case of misunderstanding in 18th Century medicine on the one hand—where we presume the doctor him- or herself can't know any better either—and intuitively troubling cases on the other: both the case of sex-by-deception (which involves ill intent and control by the deceiver), and potentially cases of blameless yet serious medical misunderstanding in a contemporary context. In any case, for the purposes of arguing that frame-dependent decisions are sufficiently autonomous, it is better to grant stronger requirements for the sake of argument that rule out too many cases, and nevertheless show that frame-dependent decisions meet the stronger requirement.

seriously underweighting this value relative to others. (Correspondingly, they overweight values that they believe apply to the option consented to but that do not in fact in apply.)

This allows us to distinguish between acceptable and unacceptable types of misunderstanding. Say you consent to a vaccine, but you don't really understand how the vaccine works, biologically speaking. The vaccine mechanism is given almost no weight. But this is not unreasonable, because it's not something you really care about anyway. By contrast, say you consent to an examination thinking that it's a professionally conducted medical procedure, but in fact it's a sexual encounter. Given your ignorance, you place no weight on the disvalue of receiving unwanted sexual attention; but it should in fact weigh very heavily in your decision, given how much you care about this.

Finally, the requirement that the agent bases their choice on a reasonable weighting of their values rules out *many,* but not all, cases of incapacity (e.g., caused by overwhelming emotion, external interference and control, intoxication, serious mental illness or disability, etc.). These very often lead to unreasonable weighting in light of the agent's own values. Take the example of overwhelming fear. Say that you are facing two medical options, one surgical and one non-surgical. The thought of surgery induces overwhelming fear; you choose the non-surgical option. What, if anything, might be wrong with that choice? In many instances, such a choice will already be ruled out as insufficiently autonomous by a failure to meet the requirement that decisions be based on a reasonable weighting of the agent's relevant values, insofar as overwhelming fear "swamps" other relevant and potentially weightier considerations.

However, it's still *possible* for a choice influenced by overwhelming fear—or by some other kind of incapacity—to be based on a reasonable weighting of the agent's values. Imagine

that the non-surgical option is in fact very reasonable in light of the agent's values (comfort, efficacy, cost, etc.). And imagine that the agent chooses this option on the basis of those values. Still, the mere thought of the surgical option induces overwhelming fear, so that the agent is unable to properly assess its merits, and wouldn't have been able to give it reasonable consideration even *if* the non-surgical option had not been suitable, and even if the surgical option had, all-things-considered, presented a tremendously better choice. The overwhelming fear puts 'blinders' on the agent: it leaves them unable to consider certain values when making their decision. It's plausible that such cases are not sufficiently autonomous for consent. To rule out such cases, we need to add a second condition:

> **Absence of incapacity:** in making their decision, the agent was not significantly hampered in their ability to consider and weigh values in a reasonable way.

This condition rules out remaining forms of interference, control, or illness, immaturity and so on that undermine the autonomy of a consent decision. Such conditions hamper agents in their ability to consider and weigh values in a reasonable way, even if they in fact happen to get it right, and act as 'blinders' in the sense that they leave the agent unable to consider reasonable alternatives to the values that in fact form the basis of their decision.

So, in summary, I have proposed a simplified model of value-based decision-making. Within this model, agents select amongst different possible sets of value weightings and resulting options. Some of these selections constitute sufficiently autonomous decision-making processes, and others do not. They constitute sufficiently autonomous decision-making processes if:

1. **Reasonable weighing of values:** the agent accords sufficiently reasonable weights to different attributes in making their decision (where the reasonableness of the weight is determined by how well it corresponds to how much the agent values the attribute in question).

2. **Absence of incapacity:** in making their decision, the agent was not significantly hampered in their ability to consider and weigh values in a reasonable way.

The proponent of the Likelihood Argument needs to show that it is likely that at least one of these conditions is violated in cases of frame-dependent consent. There are various ways that frames could affect decisions that would make this true. For instance, a number of writers who are troubled by the potential impact of framing effects have stated that framing effects involve serious misunderstanding; bypass or seriously interfere with the agent's ability to reason about their choice; play on overwhelming emotion; cause people to make choices that are largely independent of their own personal values and preferences; cause people to make decisions based almost entirely on the one framed attribute to the neglect of others (a case of what I called 'swamping' earlier); or betray deep-seated irrationality. If any of these claims about the effects of framing are true, this would, indeed, make it likely that frame-dependent consent violates at least one of the conditions on sufficiently autonomous decision-making.

10. The Likelihood Argument Revisited, Part One: Other Theories of Framing

If the proponent of the Likelihood Argument is correct, decisions that are frame-dependent are likely to violate at least one of the two conditions outlined in the previous section. For this claim to be plausible, it must be the case that frames affect decisions in ways that mean that agents weigh attributes in an unreasonable way, or at least are impaired in their capacity to weigh attributes in a reasonable way, in light of the agent's values. By contrast, if I am correct, then decisions that are frame-dependent are not likely to violate either of these two conditions.

As discussed in Section 8, frame-dependent decisions are not likely to violate conditions on sufficient autonomy to the extent that their frame-dependence is explained by the fact that the presence of a frame "leaks" choice-relevant information. In such cases, even holding the agent's values constant, it becomes reasonable for them to alter the weight they assign to different values in response to different frames, because it becomes reasonable for them to believe that different attributes are realized (or are realized to different degrees) in response to different frames (e.g., to infer that "is the option my doctor recommends" is realized to a greater degree in the case that they use a positive frame).

But let's set aside this explanation of framing effects—cases where agents end up weighting an attribute more heavily because they have evidence that it applies to a greater degree. What are the implications of alternative explanations of framing effects for the likelihood that they violate conditions on sufficient autonomy?

Here are two competing models of what an alternative explanation could look like. The one friendly to the would-be proponents of the Likelihood Argument is as follows:

> **Insufficiently autonomous framing effects:** (a) frames affect decision-making by substantially altering the weight that agents assign to different attributes, and/or (b) framing effects are associated with a significant inability to consider and weigh attributes in a sufficiently reasonable way.

By contrast, I am now in a position to propose a rival model:

> **Sufficiently autonomous framing effects:** (a) frames affect decision-making by slightly altering the weight that agents assign to different attributes, and (b) being subject to framing effects is not associated with a significant inability to consider and weigh attributes in a sufficiently reasonable way.

The latter suggestion offers a picture for suboptimal and variable yet sufficiently autonomous consent. If this prediction is correct, then it is not the case that framing effects give us good reason to believe that decisions are likely to be insufficiently autonomous, because they do not make it likely that either of the two conditions on sufficiently autonomous decision-making has been violated.

Of course, this account is compatible with the possibility that a *given* framed decision violates one of the conditions of sufficiently autonomous decision-making. For instance, there could be an agent who suffers from serious misunderstanding or incapacities for independent reasons, and who is *also* subject to framing effects. It could even be that in a small number of cases framing itself does lead to insufficiently autonomous decision. But, if my suggestion is correct, it means that a decision being frame-dependent as such does not make it *likely* that the decision is insufficiently autonomous. This suggestion is not compatible with the hypotheses of those troubled by framing effects—such as the claim that framing effects involve serious misunderstanding of risks, lead to the swamping of decision-making by one attribute, induce overwhelming emotion, or lead to decision-making that is independent of the agent's preferences, etc. All of those hypotheses imply a large alteration of the weights assigned to different values.

Support for my preferred model—the sufficiently autonomous framing effects model—will be built as follows. In this section, I will consider recent empirical theories of framing effects other than the information leakage theory, and I will argue that they are broadly compatible with the picture of sufficiently autonomous framed decision-making that is posited by my model, and that they do not comport well with the rival model of framing effects according to which framed decisions are likely to be insufficiently autonomous. In

the following section, I will consider whether existing empirical evidence provides more direct support for my model over the rival model.

Recall that the term "framing effects" is used to refer to a number of different phenomena in the empirical literature.[75] For this survey, we are especially focusing on theories of what are called "attribute" framing effects. This is the category of framing effect used in the empirical literature that best covers the sorts of examples we have been concerned with and that have sparked concerns about the validity of consent, such as framing outcomes in terms of chance of success vs. chance of failure, survival rates vs. mortality rates, the proportion of people who experience the side-effect vs. the proportion of people who don't experience the side-effect, etc.[76]

[75] Levin, Schneider, & Gaeth, 1998.

[76] This means I am mostly setting aside the well-known and widely discussed case of 'risky-choice framing' and the prospect-theoretic models designed to explain that phenomenon (Tversky and Kahneman, 1981), discussed earlier in Section 4.c. Although attribute and risky-choice framing are often conflated in philosophical discussion of framing effects, and attribute framing (e.g., mortality/survival framing) can concern ways of framing a risk, so-called 'risky-choice' framing effects actually refer to a more specific phenomenon, where positively or negatively framed information affects whether agents prefer options that involve risk where the outcomes are only known with some less-than-certain probability compared to options where the outcome is certain. Attribute framing is a more general and, in some sense, more basic phenomenon (e.g., see Teigen, 2015) than risky choice framing. In addition, the latter only applies to a more specific set of choices, and so poses less of a potential threat to consent. Prospect Theory makes some specific predictions for how framing will affect attitudes towards a procedure with x% risk *compared to* procedures (i) with no risk; (ii) with greater risk; (iii) with lesser risk, but it does not make predictions for how positive or negative framing will affect evaluations of the procedure considered on its own. We will also be setting aside "goal framing"—firstly, because many goal framing studies concern voluntary health-related behaviors that are not pertinent to questions of consent (e.g., examining the likelihood that people will stop smoking after either receiving messages that focus on the negative consequences of failing to stop smoking, or messages that focus on the positive consequences of stopping smoking); secondly, as discussed in Section 2, Section 4.b and note 25, because the frames used in most goal framing studies are often not relevantly equivalent. We are also setting aside "framing" as used to mean non-equivalent "issue framing" in media and political studies where this does not informational equivalence, as explained earlier in note 4.

I will first briefly run through different accounts of framing effects, before turning to evaluate their implications for autonomy. I will argue that, despite differences between the accounts, they bear on autonomy in similar ways.

Though they posit different mechanisms, all of the theories can be connected to the model of decision-making laid out in the previous section in a similar way. In particular, all of the theories suggest that frames lead us to place more weight on some attributes and less on others, either by altering which attributes we attend to in the first place (and correspondingly which attributes we neglect), or by altering how heavily we weigh an attribute that is being attended to relative to others.[77] A crucial question for us to consider, after briefly outlining the theories, will be whether frames do this in a way, or to a degree, that leads to insufficiently autonomous decisions.

The Associative Theory

The classic account of attribute framing effects is the Associative Theory developed by Levin and colleagues in the 80s and 90s and discussed in an influential review of by Levin, Schneider and Gaeth (1998). Levin and colleagues suggest that attribute framing is primarily to be explained in terms of associations: positively valenced frames (such as survival) tend to evoke positive associations, whilst negatively valenced frames (such as mortality) tend to evoke negative associations.

[77] The Associative Theory, Query Theory, and Fuzzy Trace Theory all posit mechanisms that fall in the former category of mechanisms that affect which attributes receive attention. The Associative Theory also includes a mechanism in the second category, one that affects how much weight an attribute is given that is already being attended to; Prospect Theory falls into this category as well, as it concerns the difference in how much weight we give to a given amount of an attribute relative to a reference point depending on whether it is framed positively or negatively relative to that reference point.

These associations can affect decisions for at least two reasons: one cognitive, and one non-cognitive. The cognitive reason is that associations affect which reasons you are more likely to consider and attend to. A positive frame tends to have associations with other positive concepts and attributes, making those more accessible to thought; it is thus more likely that you will attend to more positive attributes of the option in question. Correspondingly, a negative frame with negative associations will tend to prompt you to attend to more negative reasons against the option.[78] The second reason associations can affect decisions, according to this theory, is that words are associated with positive (or negative) emotional valence, and the evoked emotional valence of the frame directly 'colors' how favorably you feel about the option in question, independently of which attributes one attends to and considers. So in the same way that you might simply feel a little better about a treatment proposal because you're in a very beautiful and comfortable room (even if you recognize it's a bad proposal), you treat the positive feelings evoked by a positive frame as if they transfer onto the option itself. This effectively leads you to treat the positive frame (rather than just the attribute that is referred to by the frame) as if it is an attribute that *in itself* weighs positively in favor of the option in question (or, in the case of a negative frame, as an attribute that in itself weighs negatively against it).

[78] There are different mechanisms that can drive such an associative effect. According to traditional accounts of associative processes, this happens because associative connections 'activate' connected concepts, and this makes it quicker and easier to access those concepts. More recently, proponents of the Associative Theory have begun to describe the theory in terms of the way that associations affect which reasons one is more likely to pay attention to (Kreiner & Gamliel, 2018). The associative theory also predicts that frames will activate concepts that are associated with the frame but that do not necessarily have the same valence—for instance, because there is a conceptual association between them although they differ in valence.

Query Theory

Like the Associative Theory, Query Theory—developed by Eric Johnson, Elke Weber and colleagues—also posits that frames affect which subset of attributes we are more likely to attend to and thus weigh in our decision, but the theory offers a different picture of why this is, situated within a broader account of decision-making.[79] According to this theory, we make decisions (at least implicitly) by retrieving evidence in answer to different 'queries' we might make about a decision (e.g., What are reasons in favor of doing it? What are reasons against doing it? How will it affect my quality of life? How will it affect my family? Etc.). Furthermore, we only process a limited amount of evidence and number of reasons when making a decision. Consequently, order matters: we tend to gather more evidence in response to whichever 'query' we happen to process first, and so more reasons related to earlier queries actually factor into our decision, whilst it tends to be the case that fewer reasons related to later queries are attended to and factored into the decision.

So, according to Query Theory, attribute framing effects arise because frames can affect which 'query' is asked and answered first; and, since we only generate a limited number of reasons when making a decision, this affects which reasons are factored into the decision. Although the theory hasn't been directly tested in the context of medical decisions, we can imagine how this might go: positive frames like survival rates might lead participants to first think about the positives in favor of the treatment or procedure, and only subsequently weigh negatives. By contrast, negative frames like mortality rates might lead participants to first think about all of the negatives that count against that option. If so, patients making decisions in response to a survival frame might be more likely to think of

[79] E.g., Johnson, Häubl, & Keinen, 2007; Weber et al., 2007. Hardisty, Johnson and Weber (2010) apply Query Theory to attribute framing effects in the political domain. Wall and colleagues (2020) apply Query Theory to risky choice framing.

more positives and so be more likely to say 'yes' compared to patients making decisions in response to a mortality frame, who will tend to bring more negatives to bear on the decision, and thus are more likely to say 'no'.

Fuzzy Trace Theory

Fuzzy Trace Theory—developed by Valerie Reyna and colleagues—also posits a mechanism by which frames affect which subset of attributes and values we attend to when resolving a choice. But Fuzzy Trace Theory offers a different explanation for this, situated within its own broader account of decision-making.[80] Rather than focusing on positively or negatively valenced associations (Associative Theory), or the order in which we consider different attributes (Query Theory), Fuzzy Trace Theory focuses on the way frames lead us to *represent* the option in a different light and thus leads us to consider a different selection of our own values.

Specifically, according to Fuzzy Trace theory, we process options along a continuum from more "fuzzy", abstract representations that get at the overall meaning, import or "gist" of the option (e.g., "a high-risk procedure") to what the authors term "verbatim" representations—representations that capture all of the surface-level details in a thorough way (e.g., "a surgery with a 40.2% survival rate"). When trying to decide between options, we start by paying attention to more gist-like representations. Based on these representations, we retrieve relevant values from memory to apply to the decision at hand (e.g., considering "a high-risk procedure" might prompt me to think about the disvalue I place on taking high levels of risks). If this fails to give us an adequate reason to make a

[80] Reyna & Brainerd, 1991; Broniatowski & Reyna, 2018. Reyna and colleagues (2015) discuss the implications of Fuzzy Trace Theory for medical decision making. Gamliel and Kreiner (2020) apply Fuzzy Trace Theory to attribute framing.

choice (for instance, because one option seems sufficiently good and/or sufficiently better than the other), we move to increasingly verbatim representations, until we are able to resolve the choice.

According to the theory, frames can affect the "gist" of an option—the fuzzy representation of the option— and thus it affects the likelihood that certain values are brought to bear on the decision. For example, "a surgery with a 95% chance of survival" might be represented in terms of "a very safe surgery", whereas "a surgery with a 5% chance of death" might be represented in terms of "a surgery in which I may die". Thinking of the option in terms of these different "gists" affects the values that are brought to bear on the decision, and thus results in it being viewed more positively or more negatively; and, depending on what alternative option this is being compared to, this may be considered a sufficiently good reason on which to base the choice. (For example, the mortality frame just mentioned may lead me to think about the strong value I place on avoiding risking my life, and this might, under the circumstances, be enough to lead me to reject the surgery in favor of an option with no mortality risk. On the other hand, thinking about the surgery as a safe procedure might do little to help me distinguish the option from a safe non-surgical intervention; moving on to consider more of the details, I think about how I value the fact that the surgery, unlike the non-surgical option, will free me of long-term reliance on medication, and this is enough to sway me in favor of the surgery.)

So, according to the Associative Theory, Query Theory, and Fuzzy Trace Theory, frames can affect decisions, not because they provide new information as such (as Information Leakage theory would have it), but because they lead us to place more weight on some attributes and less on others, either by altering which attributes or values we attend to in the first place (according to all of the theories), or in addition (according only to Associative

Theory) by directly affecting the degree to which we weigh an attribute (positively or negatively, depending on the frame).

The crucial question is whether the way this occurs is likely to lead to insufficiently autonomous decisions. I suggest that *none* of these theories predict that susceptibility to framing effects depends on, or is associated with, features that undermine the sufficient autonomy of a decision.

Notice, first, that none of the theories predict that framing effects involve features thought to be typical culprits of undermined autonomy, such as: significant misunderstanding; decisions that are reflexive or bypass the agent's values; decisions that are based on alien values, such as inappropriate social influence or sudden impulses; or temporary or unusual impairments to reasoning or decision-making. Instead, they all assume that decisions are mostly based on the agent's own values—they just offer different stories about how some values end up playing a more prominent role in an agent's decision depending on the frame. They also all assume that the influence of frames is a side-effect of normal decision-making processes. They don't, for instance, posit that framing effects are driven by a particularly impaired subpopulation, or that they're explained by temporary and overwhelming angst inspired by words like "mortality". Indeed, to Fuzzy Trace Theory, the kind of processing that gives rise to framing effects— "gist-based" processing—tends to give rise to *better* quality reasoning and decision-making, and is associated with greater decision-making expertise within a domain (Reyna, 2018) because it is related to the ability to extract the overall meaning and significance of options. We might say that, according to Fuzzy Trace Theory, gist-based processing is what allows us to see the forest for the trees.

So then the question is whether the kinds of decision-making processes that these theories *do* posit—in particular, processes that prioritize certain values at the expense of others—do so in a way that fails to meet standards of sufficient autonomy, at least when frames end up exerting a decisive influence on decisions.

Such a failure does not follow *merely* from the fact that *at least some* attributes are neglected and underweighted when agents make decisions, according to the decision-making processes posited by these theories. As discussed in the previous section, decisions can be sufficiently autonomous so long as the agent accords *sufficiently* reasonable weights to different attributes, and this is consistent with some attributes being neglected.

Indeed, having decision-making processes that involve some way of prioritizing and limiting our attention to some attributes over others is likely to be necessary for us to make sufficiently autonomous decisions, even though in doing so we necessarily depart from the decision-making of an ideally autonomous agent. This is because real human decision-makers have limited time and processing capacity. Because of these limitations, if we attempted to meet the ideal of attending to all relevant factors before making a decision, we wouldn't be able to make one at all—ironically leaving us even further away from the ideal of autonomous decision-making. If suboptimal agents are to make sufficiently autonomous decisions, therefore, it's important that their decision-making processes have a system for determining which limited subset of potentially relevant considerations are brought to bear.

Of course, conditions on sufficient autonomy would be violated if these processes of prioritization resulted in decisions that are *overly* focused on one or two values *at the expense of important others.* But—I'll now argue—none of these theories posit or predict this either.

This is because these theories say that different frames can lead to different decisions because they serve as a prompt for one to consider and bring to bear certain values (ones that are associated with the frame, linked to a frame-associated query, or that are relevant for evaluating a frame-inspired gist). This does not, however, presuppose that the considerations that *have* been highlighted by the frame will be treated as far more important than they, in fact, are, when they have simply been brought to attention; the agent may then evaluate them according to their relevance and importance to the decision in question, and adjust their evaluation of the options accordingly. Of course, the non-cognitive component of the Associative Theory, according to which the emotional valence of the frame has a direct influence on evaluations, does imply that the option will thereby be perceived as more valuable independently of whether it is so; but this effect, like priming effects more generally, are only thought to be subtle, giving things an extra evaluative "glow", but not distorting perceptions to a significant extent. So these theories do not predict that the importance of attributes and values highlighted by frames will be greatly over-weighted.

Furthermore—and this is important—these theories do not assume that frames are the *only* cues that can and do act as prompts that bring certain attributes or values to the agent's attention. Crucially, another such cue is whether an attribute or value is important and potentially decisive with respect to the decision at hand! If so, when the value is important to the decision, we are likely to include consideration of it in our limited pool of considered reasons, irrespective of framing. So these processes are not likely to lead to the neglect or significant underweighting of important values.

For example, according to Levin and colleagues, framing results from an associative effect that is supposed to only act like a small nudge in one direction or another, leaving your

decision to be primarily based on whichever values are most important to the decision. For this reason, they argue that attribute frames should be most likely to have an effect when agents have little else to go on to make a choice or when agents don't feel strongly about the choice either way (Levin, Schneider, & Gaeth, 1998, p.160, p.164). Consequently, a decision that depends on a frame exerting an associative influence of this sort is still largely to be explained in terms of facts about the compatibility of the option with the agent's desires and values; if the option was not largely recommendable on the basis of the agent's desires and values, then the agent wouldn't choose it, associative glow notwithstanding. The same considerations would seem to apply to Query Theory: while a frame might help some values get to the front of the queue, as it were, it doesn't prevent you from considering your most important values, which suggests that framing should only have a decisive effect when the options you're considering are relatively comparable.

Supporting the thought that frames aren't especially likely to lead you to seriously underweight important values, Fuzzy Trace Theory predicts that one will only neglect details about one's choice to the extent that "fuzzier", gist-based representations are sufficient to help you decisively resolve the choice you are facing. It thus does not predict that you are likely to ignore *important* details. So although the theory does imply that we fail to consider all of the details of options and correspondingly fail to bring to bear all of our relevant values, it hypothesizes that we only stop processing—stop progressing to more detailed representations of options—once our representation and resultant reasoning about the valuableness of an option is deemed sufficient for resolving the problem at hand.

So none of these theories gives us a reason to think that dependence on framing makes it likely that the first condition on sufficiently autonomous decision-making has been violated—namely, the condition that requires the agent to base their decision on a

sufficiently reasonable weighing of attributes in light of the agent's values. What about the second condition, requiring that agents retain the *capacity* to weigh different attributes in a reasonable way? We again find that the mechanisms posited by these theories do not make a violation of this condition particularly likely just because a decision is frame-dependent. This is for much the same reason that they do not make it likely that the first condition has been violated: these theories do not posit incapacities or controlling influences on decision-making that block agents from being able to bring multiple values to bear, including whichever are the most important in the current situation. According to all of these theories, frames set you up to view a choice in a particular light and to bring related values to bear in the first instance; but they don't predict that you are unable, then, to change the way you're thinking about it of your own accord, to bring other values to bear if they seem sufficiently pressing, or to bring more considerations to bear if the first way of thinking about it doesn't resolve your decision. Because of this, it's difficult to make plausible the claim that a decision affected by frames in the way described by these theories is thereby unlikely to be sufficiently autonomous.

In sum, none of the major theories of attribute framing predicts that framed consent is *likely* to be insufficiently autonomous. Instead, all provide explanations of framed consent that imply that it is likely based on the agent's desires and values, and none invokes a significant failure of cognitive requirements (such as misunderstanding) or conative requirements (such as the introduction of alien, unendorsed values or controlling influences) to explain framing effects. They thus suggest that framing effects do *not* make it particularly likely that an agent has been caused to make a decision in a way that gives different attributes insufficiently reasonable weights.

11. The Likelihood Argument Revisited, Part Two: Empirical Evidence

Although I argued that none of the theories surveyed in the previous section invokes or predicts significant failures of autonomous decision-making in order to explain framing effects, we did not survey any evidence that was strictly speaking *inconsistent* with this possibility. We will now attempt to assess what empirical evidence there is that bears more directly on this question.

If the Likelihood Principle is correct, then susceptibility to framing effects should make it likely that one or both of the two conditions of sufficiently autonomous decision-making has been violated. Violation of the first condition occurs if the agent assigns unreasonable weights to different attributes. This can happen in two ways: either because the agent has inaccurate beliefs and misunderstanding regarding whether important attributes apply to the options they are considering; or because the agent does not give those attributes appropriate weights in light of how much they value them. Violation of the second condition occurs if the agent is unable to assign reasonable weights to different attributes.

By contrast, recall that according to my account of sufficiently autonomous framed decisions, which rejects the Likelihood Principle, framing affects decision-making by only slightly altering the weight that agents assign to different attributes, and being subject to framing effects is not associated with a significant inability to consider and weigh attributes in a sufficiently reasonable way.

A number of predictions fall out of this account. Firstly, framing effects should not be associated with significant misunderstanding of one's options. Secondly, framing effects

should be moderated by the extent to which one's values are decisive with respect to one's options. Thirdly, susceptibility to framing should not be associated with significant reasoning or decision-making incapacities. Finally, frame-dependent decisions should be likely to be sufficiently autonomous whichever frame they happen to have been affected by. In what follows, I will consider these predictions in turn. Although available evidence is limited, I will argue that what evidence there is supports these predictions.

11.a. Prediction 1: Understanding

If the sufficiently autonomous account of framing effects is correct, we should not find that framing effects are particularly associated with misunderstanding of important attributes. But some theorists have claimed that frames *do* lead to poor reasoning or poor understanding, even for those who otherwise are able to reason and understand their options well. Along these lines, Ploug and Holm argue that:

> "Nudging by framing information in terms of survival rather than mortality rate or vice versa poses a problem because it intentionally thwarts the patient's reasoning and understanding of information about an intervention." (Ploug and Holm, 2015, p.34).

Similarly, Beauchamp and Childress's (2013) pessimism about consent in the face of framing effects centers on a claim that patients under the influence of framing misunderstand risks and therefore decide non-autonomously. Discussing mortality/survival framing effects, they write:

> "These framing effects reduce understanding, with direct implications for autonomous choice. If a misperception prevents a person from adequately understanding the risk of death and this risk is material to the person's decision, then the person's choice of a procedure does not reflect a substantial understanding and his or her consent does not qualify as an autonomous authorization." (Beauchamp & Childress, 2013, p.135)

Unfortunately, few studies have directly examined associations between susceptibility to framing effects and levels of understanding (although associations between susceptibility to framing and general numeracy and reasoning ability have been studied; we will discuss this momentarily). However, I know of no evidence that supports such conjectures about the impact of framing, nor—as we saw in the previous section—do leading theories of the mechanisms that explain framing effects predict that framing interferes with reasoning or understanding. And there is some evidence that counts against the prediction that framing effects turn on misunderstanding. Armstrong and colleagues (2002) presented participants with information about the benefits of a preventative surgery in the form of graphs that showed mortality rates over time, with one line for people that have the surgery, and one line for people that do not. Participants either viewed this information in terms of mortality rates over time, or graphs showing equivalent survival rates over time. Armstrong and colleagues also tested participants' understanding of the mortality information that had been presented. While the framing manipulation affected participants attitudes towards having the surgery, susceptibility to this framing effect was not related to participants' level of understanding. (Indeed, there was a small trend—though this was not statistically significant—towards greater susceptibility to framing amongst those who answered all of the understanding questions correctly.) This is not consistent with the hypothesis that frames interfere with one's understanding of statistical information like mortality rates.

11.b. Prediction 2: Decisive Values

If the sufficient autonomy account of framing effects is correct, and framing effects do not make it likely that either condition of sufficient autonomy is violated, then framing should not cause agents to assign unreasonable weights to different attributes. To the extent that framing only affects the weights an agent assigns to an attribute to a reasonable

degree, framing should not lead people to select options that are all-things-considered very far out of line with their values, but they could lead people to select different options that are sufficiently in line with what their values recommend.

An important prediction of this account, then, is that framing effects should be moderated by the degree to which an agent's values count strongly in favor of one option over another. Specifically, framing effects should be less likely to the extent that an agent's values count strongly in favor of one option over the other. We might call such cases ones in which the agent's values are more 'decisive' with respect to the different options. Correspondingly, framing effects should be *more* likely to the extent that an agent's values do *not* strongly favor one option over the other—cases where each option is all-things-considered comparably valuable. By contrast, if framing effects are likely to involve assigning *unreasonable* weights to different attributes, then we should expect framing to sway people's decisions even if the agent has values that count strongly in favor of one option over the other, because it is in such cases that framing has the potential to lead to an unreasonable choice in light of the agent's values.

My prediction here is related to the widely cited claim made by Levin and colleagues that framing effects are mitigated in cases of strongly held attitudes (Levin et al., 1998); they argued that when participants have strongly held pre-existing attitudes about the topic in question, they are less susceptible to framing effects, with the framing itself having little influence on the decision. My prediction, however, must be more precisely stated: framing effects should be mitigated in cases where attitudes *strongly favor* one option *over* another. (If one has comparably important and strongly held values that favor opposing options, one might in some sense be said to have strongly held attitudes about the choice without it being the case that, on balance, one's attitudes strongly favor one option over the other.)

Levin and colleagues' claim is widely cited, though, surprisingly, not very much research has directly tested and explored it. However, there are a number of studies that are suggestive in this regard, which, put together, give us good reason to believe that framing effects are not likely to occur when the agent's values are decisive.

The main evidence that Levin and colleagues cited in support of their claim is the failure to find framing effects with controversial or hot button topics. In particular, they refer to a study that found significant framing effects for medical decisions on non-controversial issues but failed to find framing effects on decisions to have an abortion in light of information about risk of severe disability in the fetus. The authors of that study suggest that this might be because many people have desires and values that overwhelmingly determine their choices in this case. For example, many people might have strong anti-abortion values that decisively lead them to reject abortion, irrespective of framing. Unfortunately, however, the values of the participants were not measured or tested by the study in question, so interpretations of the study along these lines—while intuitively highly plausible—has remained speculative.

More convincing, however, is a study mentioned earlier, by Haward, Murphy and Lorenz (2008) concerning the effect of framing on perinatal care. Recall that participants were less likely to choose intensive care over palliative comfort care for a premature newborn if the risks of intensive care were framed negatively (in terms of mortality and disability rates) rather than positively (in terms of survival and rates of being disability-free). Interestingly, religiousness significantly modified this impact of framing on decisions: participants who reported being highly religious were more likely to choose intensive care over comfort care, and, unlike nonreligious or mildly religious participants, this choice was not affected

by framing. That is, highly religious participants were not made any more likely to choose comfort care instead when the chances of surviving intensive care without disability were framed negatively. A natural explanation of this finding (and along the lines of one explanation suggested by the authors of the study) is that it has to do with differences in the background values of the highly religious compared to the mildly or non-religious. In particular, this finding would be explained by the assumptions that (i) frames affect the nonreligious, somewhat religious, and highly religious in the same way, where (ii) this effect does not block or overpower the role that the agent's background values play in their decision-making, but rather (iii) works by somewhat amplifying the role that one or another value plays in the decision, through inferences about its application to the situation, through drawing attention to it, or through weighting it somewhat more or less heavily.

According to this explanation, both the religious and nonreligious participants make sufficiently autonomous decisions based on desires and values. However, the highly religious participants were not affected by framing because they were more likely to have desires and values that overwhelmingly counted against comfort care (allowing the infant to die), perhaps because they are more likely to see the moral value placed on preserving life as requiring absolute prohibitions on such options (as in absolute prohibitions on abortions), or simply because it is regarded as an overwhelmingly weightier value than other values that might be highlighted by certain frames (such as the painfulness of intensive care, or any disvalue patients might attach to disability). This kind of valuing might mean that they are liable to refuse comfort care and choose intensive care irrespective of possible frame-induced variations to the value attached to the consequences (for instance, pragmatic inferences about what the doctor recommends, or the weight attached to the possibility of the surviving infant having a disability).

By contrast, according to this proposed explanation of the results, those who were mildly religious or non-religious were less likely to have values that decisively favored comfort care over intensive care irrespective of framing. For instance, they might have been less likely to strongly believe in absolute prohibitions on allowing newborns to die, and more likely to have values that are consistent with trading off the value of attempting to save the newborn's life against the disvalue of living with disability or the possibility of subjecting the newborn to needless suffering. Consequently, they were more likely to have non-decisive values—values that mean that either option could be chosen with sufficient autonomy. Because of this, frames that highlighted certain considerations (e.g., led them to think of more of the positive or negative consequences of intensive care, respectively, or led them to weigh the disvalue of living with disability a little more) could lead those considerations to be weighted slightly more heavily, which in turn could be sufficient to sway their decision; we thus have cases of frame-dependent decisions that still involve sufficiently reasonable weightings of different attributes.

There are other examples that we can imagine admit of similar explanations: more generally, people are less susceptible to framing when their attitudes strongly favor one option over the other, because in such cases small alterations of weights due to framing are not sufficient to decisively alter one's choice. A recent, suggestive example concerns a *failure* to replicate framing effects when attitudes towards an option are already strongly positive or negative. Finkelstein and colleagues (2021) recruited advanced cancer patients with a prognosis of one year or less left to live as participants in their study. Participants were asked whether they would try a hypothetical new treatment that aimed to slightly increase their three-year survival rate by 5 percentage points (alternatively framed as aiming to decrease the three-year mortality rate by 5 percentage points). The study failed to find a significant effect of framing. Notably, when we look at how participants responded to the

question, responses were more likely to be at extremes: in both frames, the most common response was that they were "very likely" to accept the new treatment; and of the minority that rejected the treatment, more selected "very unlikely" than merely "unlikely". It seems highly plausible that, given their situation in the late stages of cancer, participants already had strong values regarding whether or not they would try new treatments to extend their life, so small adjustments to the weightings of these due to framing would not be sufficient to exert a decisive effect – for most patients, this meant they would try a new treatment irrespective of framing.

Of course, there could be many reasons why no framing effect was found in this particular case. But other studies that have directly compared groups of participants further support the hypothesis that framing effects are less likely to be found for groups that have stronger, settled views about the options. This supports the possibility that the null result in the study by Finkelstein and colleagues was due to a feature of the participants' attitudes, and not merely an incidental artifact of the particular study. For instance, positive/negative framing of the chance of flu as a side-effect of infant immunization affects attitudes towards infant immunization—but not for women who have an infant, are pregnant, or are intending to get pregnant within the next 12 months (Donovan & Jalleh, 2000). Presumably, women who fall into one of the latter categories are more likely to have actively considered the issue of infant immunization, and are more likely to have strong views about it, whether for or against. On the other hand, women who do not belong to these categories—that is, women who are not currently parenting infants or preparing to do so imminently—are relatively more likely to be ambivalent, uncertain, or to not have

settled views on the matter, as it's not an issue that currently has practical consequences or high stakes for them.[81]

The extent to which agents feel certain about a decision could also serve as a proxy for the likelihood that they have values that decisively favor one option over another: all else being equal, having decisive values makes decisions easier to resolve, while having less decisive values makes decisions more difficult to resolve, resulting in greater uncertainty as to which option is to be preferred, all-things-considered. While this also remains understudied, evidence suggestively supports the hypothesis that uncertainty affects the impact of framing on decisions. For example, Jacoby and colleagues (1993) tested framing effects on real decisions about anti-epileptic medication; although the sample size in the study is too small to be reliable, the trends in their results suggest that framing may only be effective for those who are initially uncertain about whether or not they want to continue with anti-epileptic medication.

Other supporting evidence comes from studies of attribute framing outside of the medical domain. In consumer research, for instance, framing has less of an effect on decision-making the more relevant information participants have to go on. In a famous study, participants reported more favorable attitudes to beef described as 75% lean than 25% fat; but the effect of wording is reduced if they are actually allowed to taste the beef first (Levin & Gaeth, 1988). Similarly, framing effects are less likely to be found when participants are comparing options that differ markedly in their value, or are evaluating a product that is obviously extremely bad rather than a product with only a minor flaw. For instance, Beach et al. (1996) show no attribute framing effects on toaster evaluations when toasters were

[81] See also Gesser-Edelberg et al., 2015.

missing most key attributes, but show attribute framing effects on toaster evaluations when toasters were missing only minor attributes. A natural way of interpreting such cases is again that agents are less likely to be affected by framing when the information they have is already decisive with respect to their options in light of their values.

A final illustrative example comes from the domain of political psychology. Druckman (2001) gave participants versions of the Asian Disease problem—in which, as you will recall from earlier discussion, a "risky" program that might save all six-hundred people (1/3 chance) but might save no-one (2/3 chance) is pitted against a "sure" program that has the same expected utility but will certainly save some (200 people) but correspondingly will certainly mean some die (400 people). Participants shown the traditional version of the problem, in which the two options were labelled Program A and Program B, exhibited the classic framing effect—they tended to prefer the "sure" option when the options were framed in terms of the numbers saved, but the "risky" options when framed in terms of the numbers who would die. But the framing effect disappeared for a version of the problem in which the exact same two options were referred to as the Democrat Program or the Republican Program; in this case, the primary predictor of preferences was simply party loyalty. Again, the most natural explanation of this finding is that framing only slightly alters the weights one gives to different attributes (in this case, to risk), and that this is enough to decisively affect choices between options that are otherwise very similar and broadly compatible with people's values (the classic problem), but not enough to decisively affect options when one of the options is superior on an attribute, such as political affiliation, that many people have strong views about (the political problem).

In sum, while direct empirical investigation of this remains limited, evidence suggests that the decisive impact of framing is moderated by the extent to which one's values

differentiate strongly between the options. Religion moderates the impact of framing on choices about allowing sick infants to die; being close to death moderates the impact of framing on willingness to undergo treatment to try to extend one's life; certainty moderates the impact of framing on medication choices; politics moderates the impact of framing on choices between otherwise similar public health interventions. This is line with the hypothesis that framing only slightly alters weights agents assign to different attributes, and thus with my claim that we should expect framing effects to only decisively affect decisions when both options are broadly consistent with the agent's values.

Based on findings that the strength of an agent's values moderates the impact of framing, I have suggested that framing only slightly affects the weights people assign to different attributes, and that this mechanism applies both to people who have values that bear strongly on the choice and those whose values are less decisive with respect to what they should choose. But the evidence amassed here does not conclusively rule out an alternative way of interpreting the findings that is consistent with the possibility that framing substantially alters the weights agents assign to different attributes: this could be the case if having values that bear strongly on a decision (e.g., religiosity in the context of infant mortality) protect one against framing effects, but framing substantially alters the weights assigned to different attributes for those who do not have values that bear strongly on the decision. However, such an explanation is more complex and seems ad hoc. This is especially so since the explanation that preserves sufficient autonomy, but not the autonomy-undermining explanation, is predicted by the extant theories of framing effects discussed in the previous section. Furthermore, evidence considered in the next subsection provides further support for the sufficiently-autonomous model of framed decisions.

11.c. Prediction 3: Reasoning Capacity and Numeracy

If framing effects are modified by having decisive values because frames only tend to slightly alter the weights an agent assigns to different attributes (and thus are unlikely to lead agents to assign weights to attributes that are out of line with their values), then we should expect framing effects to be independent of significant incapacities or lower quality of decision-making. If this prediction is borne out, this would support the sufficiently autonomous model of framed decisions in two ways. Firstly, it would further bolster support for the claim that susceptibility to framing does not make it likely that the agent has made a decision in a way that involves an unreasonable weighing of values (thus it is not made likely that the first condition of sufficient autonomy has been violated). Secondly, it would support the claim that susceptibility to framing does not make it likely that the agent is impaired in their capacity to weigh values in a reasonable way (thus it is not made likely that that the second condition of sufficient autonomy has been violated).

On the other hand, to the extent that framing effects are associated with markers of poor reasoning or decision-making ability, this would make it likely that frame-dependent decisions are insufficiently autonomous. A common thought along these lines is that framing effects result from, or cause, deep irrationality, or that they involve an inability to properly reason about or understand the framed attribute—for instance, because of an inability to adequately understand mortality risks.

Existing evidence, however, does not support the hypothesis that framing effects are associated with significant failures of reasoning or decision-making capacities.

Firstly, susceptibility to framing effects is not associated with general markers of cognitive ability. Indeed, there is evidence that intelligence and reasoning abilities can be associated with *increased* susceptibility to framing effects.[82] For example, Dunegan (2010) found that students with higher GPAs were more susceptible to attribute framing effects because they were more likely to make inferences about the intended meaning of a message in light of its framing. In addition, having greater working memory capacity has been associated with increased susceptibility to risky choice framing effects (Urs, Goodmon, & Martin, 2019). Other studies have simply found that measures of cognitive ability are not associated with susceptibility to framing (Stanovich & West, 2008), and neither are correlates of cognitive ability, such as education level (Armstrong et al., 2002). All in all, these findings suggest that susceptibility to framing effects does not make it likely that an agent has reduced general capacity for reasoning well about their decision.

Furthermore, experience or expertise within a domain—which we would expect to be associated with increased capacity for good decision-making within that domain—can increase, or at least fail to reduce, framing effects within that domain (e.g., see Corbin et al., 2015). Consider medical expertise. Even medical experts are subject to survival/mortality framing. For instance, Bui, Krieger, and Blumenthal-Barby (2015) presented physicians with a hypothetical vignette about a woman with breast cancer who is considering whether to have a second mastectomy. The vignette detailed the estimated impact that choosing not to have the additional mastectomy would have on this particular

[82] There may be multiple, mutually compatible reasons for this. Firstly, cognitive ability is likely to be related to the propensity to make inferences about background information on the basis of frames, of the kind discussed by information leakage theory. Secondly, Fuzzy Trace Theory suggests that the ability to process an option in terms of its "fuzzy", overall meaning—zooming out from all of the precise, "verbatim" details, which can obscure the proverbial forest for the trees—is a mark of better, more intelligent decision-making (Reyna, 2018), though it is also more susceptible to framing effects.

patient's chance of survival; this was framed either in terms of the estimated decrease in survival rate or the equivalent estimated increase in mortality rate that would result from not having the procedure. The sample of 159 physicians taking part in the study had, on average, been practicing for 19 years. Still, this framing manipulation affected whether physicians said they would recommend the second mastectomy to the patient.[83]

Even if framing effects are not associated with failures of general reasoning or decision-making capacities, a skeptic might hypothesize that framing effects are associated with a more specific type of incapacity—namely, an inability to understand mathematically expressed risks or probabilities. Although we would not expect physicians to be especially poor in this regard, it is well-documented that physicians are prone to making errors in statistical reasoning.[84] So medical expertise does not guarantee excellence in the grasp of risk and statistics. Furthermore, it's at least conceivable that some subgroup of medical experts could be poor in this regard, and that this subgroup is responsible for the pattern of framing effects within this population.

One thing to say about this hypothesis is that its applicability is limited to the framing of numerical attributes like levels of risk. So the force of this argument in supporting the Likelihood Argument would be limited. Nevertheless, given that statistical information of this sort is very common in consent decisions, especially medical consent, it still has a lot of potential to threaten practices of valid consent, if it's true.

[83] Specifically, while 62.7% said they would recommend the mastectomy given the mortality framing, only 45.8% did so if given the survival framing.

[84] E.g., see Kahneman, 2011.

If susceptibility to framing effects were explained by an inability to correctly reason about risk, we would expect susceptibility to framing to be relatively strongly related to numeracy. This is because we would expect those with high numeracy to have a high degree of understanding of risks presented to them whilst those with low numeracy to have a low degree of understanding. Therefore, we would expect that the latter group, but not the former, would be affected by positive vs. negative framing of risks.

A number of studies have investigated the relationship between numeracy and survival/mortality framing.[85] However, findings have not provided strong support for the hypothesis that susceptibility to framing is strongly related to poor numeracy; associations between numeracy and susceptibility to framing are inconsistent and limited. For instance, Peters, Sol Hart and Fraenkel (2011) found that participants' assessment of a medication with a risk of side-effects was affected by whether that risk was framed as the percentage of persons who experience side-effects or the percentage of persons who do not experience side-effects. But whether participants were low or high in numeracy did not affect their susceptibility to this framing effect. And although a different kind of framing effect—framing risk as a frequency or as a percentage—*was* stronger in those with low numeracy, it was still present in those with high numeracy.[86]

[85] E.g., see Peters, 2012.

[86] Similarly, while Gamliel and Kriener (2017) found greater attribute framing effects for low numeracy participants in some cases (in four out of six conditions across two experiments), they still found medium and large effects (according to standard interpretations of Cohen's *d*) for high numeracy participants. In a different paper, Kreiner & Gamliel 2017 found that numeracy moderated framing effects for number contrasts (e.g., 30% effective vs. 70% ineffective) with framing only occurring for low numerate individuals, but numeracy had no impact on framing effects for graphical contrasts (pie charts of effective vs. ineffective portions with one or the other portion highlighted red).

This suggests that although difficulty understanding numerical information can *exacerbate* certain kinds of statistical or numerical framing effects, it's not central to framing effects in general.[87] This means that being susceptible to a framing effect does not make it especially likely that the agent is low in numeracy (even if it raises the odds).

Furthermore, even in cases of low numeracy, we require a further step to make it likely that the agent has an inadequate ability to understand the risks or probabilities that are pertinent to their decision. Indeed, to the extent that low numeracy is correlated with being more susceptible to framing effects in certain instances, this is not generally because low-numeracy subjects misunderstand the framed numbers and base their decisions on those misunderstandings. One hypothesis, for instance, is that those with low numeracy are simply more likely to use non-numerical cues to guide their decision, including the frames. Another hypothesis is that those with low numeracy are simply less likely to spontaneously and explicitly transform a numerical magnitude into positive and negative versions when reasoning about their decision (e.g., explicitly reasoning that a procedure with a 90% survival rate means it has a 10% mortality rate), and so are less likely to trigger the same process of reasoning and weighing of values. Consistent with the latter hypothesis, Gamliel and Kreiner (2017) found that although low numeracy participants showed stronger framing effects, they were equally sensitive to the magnitudes of numbers—for example, just like high numeracy participants, they evaluated a hotel with a 90% rate of favorable reviews more positively than one with 75% favorable reviews, which in turn was evaluated more positively than one with a 60% rate. This suggests that the susceptibility of low numeracy participants to framing effects does not rest on an inability to understand the

[87] Reyna, 2018.

import of the numerical quantities for the decision at hand, even though it may stem from different ways of reasoning about the numerical information.

11.d. Prediction 4: Comparing Frames

I have argued that evidence suggests susceptibility to framing is not particularly associated with decisions that are out of line with an agent's values, nor with an incapacity to make decisions that are reasonable in light of the agent's values.

But a final reason that you might think that framing interferes with the sufficient autonomy of a decision is if *one frame* in particular leads to insufficiently autonomous decisions, while the other frame leads to more autonomous decisions. For example, if framing risks in terms of mortality were especially frightening or anxiety-inducing in a way that tends to lead to unreasonable or inaccurate weighing of different considerations, but survival framing led to especially thorough, measured and sensible decisions, then susceptibility to framing effects *as such* would not be associated with poor decision-making on average, looking at the quality of people's decision-making across both frames; nevertheless, framing effects *would* be indicative that a particular category of decisions is especially poor and perhaps insufficiently autonomous—namely, decisions made after being exposed to a mortality frame.

Again, this does not seem to be borne out by the evidence. While there has been a lot of work on possible processing differences stimulated by being presented with negative rather than positive frames, most of this work concerns the hypothesis that negative framing leads to *more* thorough, deliberative reasoning (perhaps because negative words like "death" and "mortality" act as an additional prompt that the decision is something that

needs to be taken seriously and undertaken cautiously; c.f. Kuvaas & Selart, 2004), or, according to a different school of thought, merely a different style of processing, but one that is no less thorough or accurate (e.g., Putrevu, 2014).

Furthermore, consider a study by Carling and colleagues (2010) that examined participant values and hypothetical choices about taking antihypertensive medication to manage high blood pressure. Preferences about taking the medication was affected by whether information about rates of heart disease was framed positively or negatively. However, irrespective of whether participants had been presented with a positive or a negative frame, their choices were strongly associated with the balance of their values—that is, with the extent to which they rated avoiding heart disease as more important than avoiding side-effects and avoiding inconveniences associated with the medication. Moreover, irrespective of what frame they were exposed to, participants were able to reason sensibly and alter their decision based on new information about the medication (in particular, after being given new information suggesting that taking the medication was not advisable for them after all, many participants changed their minds about taking the medication, irrespective of framing). These results are not what would be expected if one kind of frame (either the negative or the positive one) induced poor understanding or poor decision-making; instead, it suggests that whether participants were exposed to a positive or a negative frame, they made choices that were consistent with their values, and irrespective of whether they were presented with a positive or negative frame they were equal able to take new information into account in their decision-making.

11.e. Summary: The End of the Likelihood Argument

Recall the Likelihood Argument for the claim that framing effects threaten consent:

The Likelihood Argument

1. **Generalization:** Many cases of consent (that otherwise seemed valid) depend on framing.
2. **Likelihood Principle:** If S's consent depends on framing, then it's likely that S's consent is not valid.
3. **Conclusion:** Therefore, many cases of consent (that otherwise seemed valid) are likely not valid.

A crucial step in this argument was the idea that being subject to framing effects makes it likely that consent is not valid (the Likelihood Principle). For this argument to succeed, therefore, being decisively affected by framing should make it likely that one has not decided in a sufficiently autonomous way. I argued earlier that this could be for either of two reasons: because one is likely to have made a decision that is not reasonable in light of one's values, or because one is impaired in one's ability to make decisions that are reasonable in light of one's values.

Earlier, we saw that such claims were not predicted by existing psychological theories of the decision-making processes that give rise to framing effects. Now, we have seen that the available empirical evidence suggests that these claims are false. Decisions that depend on framing are not thereby likely to involve important misunderstanding, an inability to make reasonable decisions, or a significant departure from one's values. Indeed, framing is more likely to affect decisions when either option is sufficiently reasonable in light of the agent's values, and less likely to affect decisions when the agent's values speak decisively in favor of one option over the other. This suggests that framing effects do not make it likely that decisions depart from a reasonable application of the agent's values. Instead, the evidence supports an account of framing effects according to which frame-dependency

itself does not pose a threat to the sufficient autonomy of consent. I conclude that the evidence does not support the Likelihood Argument, and that we should reject it.

With the rejection of the Likelihood Argument, I conclude that framing effects do not threaten the validity of consent. If it's neither the case that dependency on framing *entails* that consent is invalid, nor the case that dependency on framing makes it *likely* that consent is invalid, then framing effects do not, in general, threaten the validity of consent. Of course, dependency on framing is *consistent* with the possibility that any given token of consent is not valid, but this is just what we would expect of any ordinary decision to consent. After all, consent *is* sometimes invalidly given; one of the goals of theorizing about the nature of valid consent is to capture and understand the difference between those cases in which consent is morally transformative and those cases in which it is not. But the phenomenon of framing effects, therefore—however widespread—is not a good reason for rejecting the belief that most ordinary, seemingly-valid decisions to consent are just that: sufficiently autonomous, morally transformative decisions.

12. Conclusion: Sufficiently Autonomous, Frame-Dependent Consent

Recall the example with which we began:

> **Framing-Induced Surgery:** A patient has been diagnosed with lung cancer. Their doctor explains to them that surgery is one treatment option. He explains what this would involve and the likely impact it would have on the cancer. He also explains that all surgeries carry some risk, and states that 80% of patients survive surgery of this kind. The patient says he wants to have the surgery, signs the relevant consent forms, and, eventually, he is taken in for the surgery to be performed. However, had the doctor presented this patient with the *mortality rate* of the surgery—had the doctor stated that 20% of patients do *not* survive surgery of this kind—the patient would not have

> consented to surgery, and would have ended up pursuing a different treatment option, such as radiotherapy.

This example seemed troubling. In particular, it motivated a worry that consent decisions that are dependent on framing effects may not be sufficiently autonomous to be valid. Moreover, combined with a worry that frame-dependent decisions may be widespread, this sort of case seemed to motivate radically revisionist views of consent—either the view that valid consent is not, in fact, required for interventions like surgery to be permissible (Consent Skepticism), or the view that many very ordinary-seeming consent-based interactions are, in fact, morally wrong (Consent Pessimism).

These worries brought two larger questions to the foreground, one normative, and one descriptive: Firstly, what kind of decision-making could consent require? Secondly, why and how does framing affect people's decision-making?

In answer to the normative question, I argued that consent cannot require optimally autonomous decision-making. Ordinary agents often fall short of ideal standards of autonomous decision-making, and yet are still able to give valid consent. At most, then, decisions need only be sufficiently autonomous, and many ordinary yet suboptimal ways of resolving decisions can give rise to valid consent. To develop this claim further, I argued that suboptimal agents make sufficiently autonomous choices if two requirements are met: the agent's decision is sufficiently reasonable in light of their values (i.e. based on a sufficiently weighting of different attributes in light of the agent's values), and the agent is not significantly hampered in their ability to make sufficiently reasonable decisions in light of their values.

For ordinary human agents—who are unable to discern all relevant information and to weigh all relevant reasons when making a decision—meeting these requirements is consistent with some arbitrary prioritization of some subset of their values over others. Although this implies that some values will be over- or underweighted, this is consistent with a sufficiently autonomous decision so long as this does not depart too greatly from the ideal—that is, so long as the agent does not neglect or severely underweight values that are very important to them (or severely overweight others). Furthermore, and because of this, it's often the case that ordinary agents face more than one possible way of resolving a choice in a suboptimal yet sufficiently autonomous way (I called this the Options thesis), and thus it's often the case that people face genuine choices about what they can validly consent to. It follows that valid consent can be variable: a consenter may validly consent to something even though it is also true that they could have dissented (and given valid consent to some other option) instead—for example, because they pay attention to certain pieces of information over others, or because they treat this or that value as slightly more important.

This normative picture could not tell us, on its own, whether *framing effects* threaten consent. To answer this question, we had to ask descriptive questions concerning the way in which frames come to exert a decisive influence on whether or not agents give consent. In light of the descriptive facts combined with this normative framework, we would be able to ask whether frame-dependent decisions always fail to meet sufficiently autonomous standards of decision-making (the Entailment Argument) or at least are likely to be insufficiently autonomous (the Likelihood Argument).

I argued against both possibilities. More specifically, I argued that frames affect decision-making either by serving as informational cues—in which case framing effects do not

undermine the autonomy of decisions at all—or by slightly altering the weights that agents assign to different attributes—in which case they are not likely to cause an agent to choose options that are significantly out of line with their values, too out of line to qualify as sufficiently autonomous. I argued that this account of how frames affect decision-making is consistent with several leading empirical theories, and that several predictions of this account are supported by empirical evidence. By contrast, hypotheses associated with the claim that frame-dependent decisions are unlikely to be sufficiently autonomous—such as the hypothesis that frame-dependent decisions rest on misunderstanding, are associated with poor reasoning, or lead to impulsive decisions that bypass important values—are neither predicted by leading theories nor consistent with empirical studies on framing effects that have been conducted to date.

So, to return to the example of Frame-Induced Surgery: this example seemed highly troubling. But I suggest that we should no longer be troubled by this case—not now that we have considered what valid consent requires of us (suboptimal decision-makers that we are), and now that we have come to understand why framing might affect a person's decision in this way. The frame-dependence of this choice need not reflect any severe departure from rational thought; it need not reflect any misunderstanding of what is being consented to; and it does not imply that the patient is making a decision based on "mere words", bypassing consideration of all that is truly important for them. Indeed, the frame-dependence of the choice does not even make any of these worrying failures likely—that is, more likely than if it had been an ordinary decision that was not frame-dependent.

Instead, the patient in question may simply be understood as facing a difficult choice amongst multiple options that each have good reasons in their favor—recommended, tried and tested medical options to treat a deadly disease. The agent may value preserving their

life, having the opportunity to finish projects that are important to them, being able to spend valuable time with loved ones, making decisions in line with modern medical recommendations. But in light of these values, which option is superior? Available evidence and average levels of introspection may not give a strongly decisive answer here, and, if so, a frame may be sufficient to prompt them one way or the other—not because they base their decision on "mere wording", but because the frame is a mere nudge, something sufficient to prompt him or her towards emphasizing some genuinely valued considerations more than others, thus tipping the balance this way or that.

A PREFACE TO

"AUTONOMY AND THE FOLK CONCEPT OF VALID CONSENT"

In "Framing Effects, Suboptimal Agents, and the Standard View of Consent", I examined the intuitive idea that framing effects threaten the validity of consent. Given that framing effects are a psychological phenomenon, and given that autonomy is widely thought to be necessary for valid consent, I diagnosed the putative threat of framing in terms of the possibility that framing effects call into question something about the agent's decision-making processes or psychological competencies that is necessary for consent to be autonomous and therefore valid.

Ultimately, I argued that even if an agent's consent is contingent on framing, this does not imply that their consent is invalid, nor indeed does it make it likely. Consequently, I argued, ethicists should not be moved towards revisionist, skeptical or morally pessimistic positions with regards to consent on the basis of phenomena like framing effects (even if framing effects are found to be very widespread indeed).

To build my argument, I did not dispute the idea that autonomous decision-making *is* necessary for valid consent; instead, I argued that, in any case, framing effects do not make it likely that consent would fail to meet such standards. So although a crucial part of my argument was that *optimal* autonomy could not be required by any plausible theory of consent—that, at most, valid consent could only require some sufficient level of autonomy where that level falls short of optimality—my argument was otherwise built in a way that allowed for a variety of relatively demanding interpretations of what exactly the requirements of autonomy amount to.

An important way in which the definition of autonomous consent used in my argument was relatively demanding was that it allowed for the idea that the consenter must not only be in *possession* of autonomous decision-making capacities (e.g., have the ability to base their decision on reasoned consideration of values), but that the decision to consent must itself be a product of the more or less successful exercise of those capacities (e.g., consent must be based on the reasoned consideration of values with accurate understanding of what is being consented to). Thus I not only argued that frame-dependence doesn't threaten the agent's *capacity* for autonomous decision-making, but that, in fact, frame-dependence doesn't threaten the likelihood that the agent's decision to consent is itself based on a successful exercise of these capacities—successful not only in the sense of a decision that is based on the agent's own values, but also in the objective sense that the agent's decision is based on a sufficiently accurate understanding of whether highly valued attributes are in fact possessed by the option in question. Furthermore, this argument was neutral with regards to which specific kinds of psychological capacities and resulting processes are necessary for valid consent—for example, whether the necessary aspects of the agent's psychology are cognitive (e.g., understanding, theoretical rationality, etc.), conative (e.g., authenticity, basing decisions on important and non-alien values, etc.), or indeed both. I aimed to show that framing effects needn't threaten valid consent, no matter which version of this view is correct.

Making sure that my argument allowed for demanding views of what is necessary for valid consent, whether or not I accept such views, was a dialectically important move. Most contemporary theories of consent in philosophy and bioethics assume that valid consent requires that the decision to consent must be made autonomously, and that it is not sufficient merely for the agent to possess the capacity to decide autonomously. So showing that framing effects not only fail to threaten an agent's capacity to decide autonomously,

but moreover even fail to show that the agent has decided in an insufficiently autonomous way, provided a more robust rebuttal of revisionism based on framing effects. Consequently, my argument was able to encompass and dispel many versions of the idea that framing effects threaten autonomous, valid consent that have appeared in philosophical writing and thought (whether that threat is spelled out in terms of misunderstanding, irrationality, inability to reason, the bypassing of value-based decision-making, and so on).

In what follows, however, I subject the very idea that valid consent is autonomous consent to critical scrutiny, asking: what exactly is the relationship between the psychological capacities involved in autonomous agency, on the one hand, and the validity of consent, on the other? And *should* we accept the claim that autonomous agents in possession of these capacities only give valid consent if they exercise these capacities—that is, give their consent autonomously?

The following chapter, entitled "Autonomy and the Folk Concept of Valid Consent", takes an empirical approach to this question. In it, my co-author, Roseanna Sommers, and I make the case that the exercise of autonomous decision-making capacities, over and above the mere possession of those capacities, does not, in fact, play an important role in ordinary ascriptions of valid consent. The chapter takes the form of an experimental research paper, as submitted for consideration at a psychology journal, that examines what kind of autonomy, if any, underlies intuitive, ordinary ascriptions of valid consent across a variety of cases. Afterwards, I will discuss some of the philosophical implications of these findings for the ethics of consent.

"Autonomy and the Folk Concept of Valid Consent" is a co-authored paper. I originally came up with the main idea for this project—to examine the relationship between the autonomy of the consenter and ordinary judgments of the validity of consent, focusing specifically on the respective roles of the exercise and the possession of autonomous decision-making capacities in such judgments—in 2019. After conducting a pilot version of Study 1, I contacted Sommers about collaborating on the project. We both contributed equally to the design, development and implementation of the further studies reported here, to the interpretation of the studies, and to the writing of the paper itself.[88]

[88] The project was conducted in a highly collaborative way in that both authors contributed in some way to all aspects of the paper, though I took more of a lead on some aspects and Sommers on others. For example, Sommers was primarily responsible for the statistical analyses reported here, and produced all of the graphs; I took primary responsibility for parts of the paper connecting the work to philosophical debates and literature, and for providing initial drafts of most sections of the paper. All sections were subsequently edited and re-written together, and decisions about content, structure, incorporating feedback etc. were made collaboratively.

AUTONOMY AND THE FOLK CONCEPT OF VALID CONSENT

Joanna Demaree-Cotton

(Yale University)

Roseanna Sommers

(University of Michigan)

1. Introduction

Consent is morally transformative and suffuses our everyday moral and social lives. Valid consent makes the difference between permissible sex and rape; between a medical exam and assault; between entering a person's home and trespass; between an economic transaction and theft. We need consent to include participants in research, to collect private information, to borrow things, to exchange money, to perform medical procedures, to cut someone's hair, and to enter into legally binding contracts. The importance of consent is reflected in extensive treatments of valid consent in moral philosophy and the law, as well as in biomedical ethics and psychiatry (Beauchamp & Childress, 2013; Faden & Beauchamp, 1986; Appelbaum & Roth, 1982).

These fields emphasize that a person who assents—for example, by saying "yes"—does not necessarily give morally transformative or legally valid consent. For instance, most philosophers, legal scholars and medical ethicists do not consider the assent of people who are coerced, under duress, severely intoxicated, underage, intellectually disabled, or otherwise incapacitated to constitute valid consent (Beauchamp & Childress, 2013, Chapter 3; Dougherty, 2019; Hurd, 1996; Pugh, 2020; Wertheimer, 2003). Thus, the

received view in these fields is that consent is valid only if it is *autonomous* (Beauchamp, 2010).

Yet, the idea that consent must be autonomous is inherently ambiguous. It could mean that the consenter must be autonomous in the sense of having the *capacity* to make autonomous decisions. Alternatively, it could mean that the consenter must *exercise* this capacity and in fact make an autonomous decision. Note that one can *possess* a capacity, though one does not *exercise* it. Thus, someone might possess the mathematical capacity to work out the answer to "24 x 7," but that doesn't mean they will exercise it: perhaps, instead of working out the answer, they might simply guess. Similarly, a competent adult can possess the capacity to engage in autonomous decision-making, but their decision to consent could fail to be autonomous if they don't exercise this capacity: they could act whimsically or impulsively; give in to a fleeting emotional reaction; not think things through properly; or give in to pressure from others to say 'yes' to something they don't really want to do.

The present studies address the understudied question of how people reason about the validity of consent by investigating whether autonomy plays an important role in ordinary reasoning, and, if so, in what sense consent must be autonomous for it to be considered valid and thus morally transformative. Specifically, must consent be the product of *exercising autonomous decision-making capacities* (that is, be the product of rational and authentic processes)? Or is it sufficient for consent to be given by an agent who *possesses autonomous decision-making capacities,* even if they fail to exercise these capacities, resulting in a decision that may be neither rational nor what the agent really wants?

Surprisingly little is known about how ordinary people reason about valid consent, including its relationship to autonomy. Despite the pervasive importance of consent to

social life, its role in moral cognition has been relatively understudied. While ample social science research investigates how people *communicate* about consent through both verbal and non-verbal means (e.g., Beres, 2014; Wignall, Stirling & Scoats, 2020), and how people reason about violations involving nonconsensual interactions, as when a person verbally objects (e.g., Gravelin, Biernat, & Bucher, 2019; Hammond, Berry, & Rodriguez, 2011; Niemi & Young, 2016; Peace & Valois, 2014; Whatley, 1996; Yndo & Zawacki, 2020; see Muehlenhard et al., 2016, for an overview of social scientific research on sexual consent), few studies have investigated the factors that underlie ordinary judgments concerning whether consent is valid. Moreover, much of the existing psychological research on consent examines specialized topics—most prominently, sexual consent (e.g., Beres, 2014; Jozkowski et al., 2014) and informed consent (Bohns, in press)—rather than investigating consent as a domain-general moral concept. Put simply, while much work in contemporary moral psychology has focused on when "no" is taken to mean "no," little research has focused on when "yes" is taken to mean "yes."

One exception is Sommers (2020). Sommers asked participants to evaluate scenarios describing agents who give consent only because they have been intentionally deceived about important facts (e.g., an agent gives sexual consent after the partner lies to them about not having HIV; a buyer signs a sales contract after the seller lies about the product). Surprisingly—and in contrast to treatments of deception in moral philosophy (Dougherty, 2013) and the law (Blum et al., 2021) according to which deception vitiates consent—Sommers found that participants tend to judge that such interactions are consensual across a variety of contexts. This finding raises the possibility that the ordinary concept of valid consent, and its role in moral reasoning, may deviate starkly from academic treatments.

Studying consent judgments is important for several reasons. First, prior theorizing suggests that judgments of valid consent should play an important role in moral reasoning. Literature in philosophy, law, and bioethics suggest that valid consent is normally required for interactions with other people's lives, bodies, or property to be considered permissible. For instance, it's very morally wrong to have sex with someone without their valid consent. Moral philosophy and legal theory explain the importance of consent in terms of *transformation*: valid consent transforms moral rights and corresponding obligations (Hurd, 1996). For instance, if A gives valid consent to sex with B, then A (temporarily) waives their autonomy-based right to not be touched by B in this manner, and B is no longer under a corresponding moral obligation to refrain from having sex with A. Consequently, if B has sex with A, they don't wrong A in virtue of violating this right. Of course, consensual sex can be considered wrong or inappropriate for other reasons, such as adultery or incest (e.g., Haidt, Bjorklund, & Murphy, 2000; Lim and Roloff, 1999). But consent is considered a necessary element for sex to be morally permissible. Furthermore, the granting of valid consent is thought to affect the allocation of other rights and obligations, including those of innocent third parties. For instance, whether or not police are obligated to arrest someone for sexual assault should depend crucially on the validity of the sexual partner's consent—not, for instance, on the moral wrongness of violating moral norms against adultery. Thus, the ordinary concept of valid consent may play an important role in reasoning about moral permissions and moral rights.

Second, ordinary reasoning about consent is likely to have practical implications for consent policy in areas such as law, medicine, and educational campaigns (e.g., Beres, 2014; Humphreys & Herold, 2003; Marg, 2020). If people's ordinary concept of valid consent misaligns with expert treatments and official policy, it may make it more likely that people misunderstand, misapply, or simply disagree with policy, with implications for compliance

and trust in institutions that uphold these policies (Hosmer, 1995; Humphreys & Herold, 2003; see also discussion in Muehlenhard et al., 2016).

Thirdly, the folk conception of consent may carry legal ramifications. Legal scholarship and practice rely on understanding the "ordinary meaning" of legally relevant concepts, including consent; thus there is increasing interest in the contribution of the cognitive scientific study of ordinary concepts to legal theory in the emerging field of "experimental jurisprudence" (Struchiner, Hannikainen, & de Almeida, 2020; Tobia, 2022, forthcoming). Additionally, ethical attitudes influence juror verdicts (e.g., Peter-Hagene & Ratliff, 2020), juries are empowered to decide whether valid consent has been granted in legal cases (Kahan, 2010; Rerick, Livingston, & Davis, 2019).

1.1. Prior empirical work on consent and moral rights

While few studies have investigated the folk concept of valid consent, prior psychological research on adjacent concepts, such as ownership, suggests that consent may play an important role in moral reasoning about people's rights and obligations. For example, when people attribute ownership of an object, they first determine whether it was acquired consensually (e.g., through purchasing it or receiving it as a gift) or non-consensually (e.g., by stealing), and only in the former cases do they judge the possessor to be the owner of the object (Friedman, Neary, Defeyter & Malcolm, 2011). Once ownership is established, people infer certain rights over the object, such as the right to keep it or the right to determine what happens to it. From age 3, children discern that consent is important for determining ownership: they assume that the person who forbids (or allows) others to use an object by withholding (or giving) permission is the owner of the object (Neary, Friedman, & Burnstein, 2009). Correspondingly, non-owners are morally obliged to refrain

from using an object owned by someone else (Neary, Friedman, & Burnstein, 2009), taking it, or destroying it (Millar, Turri, and Friedman, 2014)—unless they have the owner's consent. Similarly, if Jack protests against Tom playing with a toy, children object to Jack's interference if the owner had given Tom permission. Thus, children appear to think that consent affects whether or not a third party may interfere (Schmidt, Rakozcy & Tomasello, 2013).

The foregoing research suggests that ownership rights are allocated as a function of consent, but whether such allocations are affected by the *validity* of the consent has not been directly investigated. If an owner grants permission in a non-autonomous fashion, will people still draw inferences about people's rights to use or interfere with the property?

1.2. The role of autonomy in judgments of valid consent: Two hypotheses

The concept of autonomy has been studied in the context of psychological wellbeing (Deci & Ryan, 2009) and it has received widespread attention in the study of moral cognition: perceptions of autonomy play an important role in folk reasoning about moral responsibility (Feltz & Cova, 2014), free will (Vonasch, Baumeister, & Mele, 2018), ownership (Starmans and Friedman, 2016), and rights of personal choice (Nucci & Lee, 1993). More generally, autonomy-based reasoning may capture a distinct and cross-cultural domain of moral thought (Graham et al, 2013; Neff, 2001; Rozin, Lowery, Imada, & Haidt, 1999; Shweder et al, 1997).

According to the literature, individuals are autonomous if they can make decisions freely and shape their lives according to their own values (e.g., Mele, 1995). Prior empirical work reveals two components thought to be required for autonomy: (a) freedom from external

interference or constraints (e.g., Deci & Ryan, 2000; Espinosa & Starmans, 2020); and (b) the possession of certain types of decision-making capacities (Baumeister & Monroe, 2014; Beauchamp & Childress, 2013, Chapter 3; Starmans & Friedman, 2016; see also Gray, Young & Waytz, 2012). The present studies focus on the latter: the decision-making capacities involved in autonomous consent, setting aside external constraints or interference by others.

Autonomous decision-making capacities are thought to include capacities for *rational* decision-making (e.g., the capacity to reason properly, to understand one's options, to appreciate the implications of a decision, and to make decisions on the basis of relevant reasons) and *authentic* decision-making (e.g., the capacity to guide one's decisions according to personal values and desires that are truly one's own; Baumeister & Monroe, 2014; Moye et al., 2006; Starmans & Friedman, 2016). These twin concepts are reflected in the measures of capacity (also known as "competence") used in psychiatry and clinical ethics to assess an individual's ability to give valid consent to medical treatment or participation in research (e.g., Appelbaum & Roth, 1982).

Thus, the present studies investigate the following two hypotheses about the relationship between the consenter's autonomy and judgments of valid consent:

> **The Exercises Capacity Hypothesis**: Whether the decision to consent is made in an autonomous (rational, authentic) way determines whether a consenter is judged to have given valid consent.
>
> **The Mere Capacity Hypothesis**: Whether or not a consenter possesses the capacity to make autonomous (rational, authentic) decisions determines whether they are judged to have given valid consent, irrespective of whether the decision to consent is in fact made in an autonomous (rational, authentic) way.

According to the Exercises Capacity Hypothesis, what goes on in the agent's mind when they are making their decision is crucial: it matters that they in fact make their decision in an autonomous way—not merely that they possess the capacity to do so. This hypothesis mirrors the importance of an agent's mental state for other kinds of moral reasoning (Chakroff & Young, 2015).

If the Exercises Capacity Hypothesis is right, the relevance of autonomy to consent is naturally explained in terms of the way it allows agents to make autonomous (i.e., rational, authentic) decisions. For instance, a straightforward explanation for why people might care whether a consenter is intoxicated would be that these impairments make it likely that the person is in fact making a bad decision—perhaps they are doing something they don't understand (an irrational choice) or don't want to do (an inauthentic choice).

The Exercises Capacity hypothesis aligns with contemporary philosophical views of valid consent, according to which a decision to consent must in fact be rational (Savulescu & Momeyer, 1995), well informed, voluntary, or reflective of appropriate values in order to be considered relevantly autonomous and valid (Beauchamp & Childress, 2013; Pugh 2020). While philosophical and bioethical views of valid consent do not require that consenters make the objectively best choice, many philosophical and bioethical accounts state that the autonomous quality of the agent's decision-making—specifically, the extent to which it is rational or based on the consenter's own preferences and values—plays some kind of necessary role in determining whether consent is valid. More broadly, legal and institutional requirements of consent are thought to help protect this philosophical ideal of autonomous decision-making, in which individuals are free to promote their own well-being as defined by their own, personal values (see Berg et al., 2001, Chapter 2).

By contrast, the Mere Capacity Hypothesis contends that it matters little whether agents *in fact* decide autonomously; it matters only whether they have the *capacity* to do so. Thus, this hypothesis has the somewhat surprising implication that being incapacitated undermines consent, but not because of how this state in fact affects the agent's decision-making process; if a sober person made the exact same choice in an equally irrational or impulsive way, their decision would constitute valid consent. If the Mere Capacity Hypothesis is right, then, autonomy matters for valid consent even when the decision to consent is not an expression of the agent's autonomy; instead, it matters only that it was the decision of an autonomous agent.

1.3. The present research

The present studies investigate the relationship between attributions of valid consent and the autonomy of the consenter. Studies 1 and 2 investigate whether the folk concept of valid consent requires that the consenter possesses autonomous capacities, and whether it additionally requires that the consenter's choice be the product of autonomous decision-making. Study 3 investigates the effect of autonomous decision-making capacities on hypothesized downstream consequences of consent, including moral judgments and the allocation of rights to third parties.

Open science. Reports of all measures, manipulations, and exclusions, as well as all data, analysis code, and experimental materials are available for download at https://osf.io/z5cdh.

2. Study 1

Study 1 was designed to assess whether the mere possession of autonomous decision-making capacities—or the exercise of those capacities—matters for the folk concept of valid concept.

2.1. Methods

Sample size, predictions and analyses were pre-registered (https://aspredicted.org/blind.php?x=zv7mm4), and our analyses adhere closely to our preregistered plans.

2.1.1 Participants

In line with our pre-registration, we recruited 450 participants on Amazon Mechanical Turk. After excluding participants who answered at least one of the binary-choice attention checks incorrectly, we were left with a sample of 364 participants (52.6% male, 47.4% female; median age 36 years).[89]

[89] Consistent with our pre-registration, we present all results with and without exclusions in Appendix A, where we also note where these exclusions made a substantial difference to results. We follow this procedure for all studies.

2.1.2 Design

Participants were randomly assigned to read one of nine vignettes in a 3 (autonomy: Exercises Capacity; Mere Capacity; Lacks Capacity) by 3 (domain: medical consent; sexual consent; consent to police entry) between-subjects design.

2.1.3 Procedure and materials

The vignettes begin by explaining that an agent is facing a decision about whether to consent to something: an elective surgery (medical consent condition), sex after a date (sexual consent condition), or consenting to police entering and searching the person's home (police entry condition). The vignettes were adapted from materials used by Sommers, 2020. The full materials are available in Appendix B; here we illustrate the three conditions using the vignette from the medical domain (see Table 1). In the Exercises Capacity condition, the agent both possesses and exercises the capacity to make autonomous decisions and says "yes" based on their personal values and thinking things through rationally; in the Mere Capacity condition, the agent possesses the capacity to do this but does not exercise it, and fails to think things through rationally or base their decision on their personal values; and in the Lacks Capacity condition, the agent does not possess these autonomous decision-making capacities at all.

After reading the vignette, participants rated their agreement with a series of statements, presented in a random order, on seven-point Likert scales that ranged from *strongly disagree* to *strongly agree*.

Medical Consent Vignette		
Exercises Capacity	Mere Capacity	Lacks Capacity
Marvin has been in physical therapy for ankle pain. One day his doctor asks him whether he wishes to undergo elective surgery to repair the tendon. The doctor explains that surgery would carry some risks, as all surgeries do, but if all goes well it could potentially completely cure his ankle pain.		
Marvin is an intelligent, able adult. He is perfectly capable of weighing up pros and cons; thinking through the choice he faces; and making decisions based on what is best for him, which options align with his personal values, and what he really wants. And he does so in this instance. After thinking things through very carefully—and with careful regard for the pros and cons, and whether it aligns with his personal values and what he really wants—Marvin says 'yes' to the surgery.	Marvin is an intelligent, able adult. He is perfectly capable of weighing up pros and cons; thinking through the choice he faces; and making decisions based on what is best for him, which options align with his personal values, and what he really wants. **But he doesn't do so in this instance. Without** thinking things through even a little bit—and with **absolutely no regard** for the pros and cons, or whether it aligns with his personal values and what he really wants—Marvin says 'yes' to the surgery.	**Marvin is not able and intelligent like most adults**. He is completely **incapable** of weighing up pros and cons; thinking through the choice he faces; and making decisions based on what is best for him, which options align with his personal values, and what he really wants. **So he doesn't do so in this instance**. **Without** thinking things through even a little bit—and with **absolutely no regard** for the pros and cons, or whether it aligns with his personal values and what he really wants—Marvin says 'yes' to the surgery.

Table 1. Study 1 vignette used in the medical consent condition, varied by autonomy condition. Boldface type is used here for emphasis; it was not used in the stimuli presented to participants.

Because "valid consent" is a technical term that may not reflect ordinary reasoning, we used three measures to assess judgments of valid consent, as follows (adapted according to vignette):

Consent 1: The doctor had Marvin's permission to proceed with the surgery.

Consent 2: If the doctor proceeds with the surgery now, he'll be acting without Marvin's consent.[90] (reverse-scored)

Consent 3: Marvin's 'yes' didn't count as consent. (reverse-scored)

According to the Mere Capacity Hypothesis, lacking capacity will be perceived as undermining valid consent, whereas failing to exercise capacity will not. Thus, the Lacks Capacity agent will be rated as lower in consent than the Exercises Capacity agent, while the Mere Capacity agent will not. By contrast, according to the Exercises Capacity Hypothesis, failing to exercise capacity will be perceived as undermining consent. Thus, both the Mere Capacity and the Lacks Capacity agent will be rated as lower in consent than the Exercises Capacity agent.

We also measured participants' judgments of the extent to which the agent was making the right choice by saying "yes":

Right Choice: Having surgery was probably the right choice for Marvin.

To ensure that our manipulations had the intended effect, we included four measures to check whether participants thought the agent had the general capacity to make decisions rationally and authentically (phrased as "the ability to be true to himself when making decisions"), as well as whether they thought the agent had done so in this particular

[90] A typo was discovered in the measure for Consent 2 for participants in the Police Entry condition. Although the character in the vignette is called "Johnny", this measure read, "If the police officers enter and search Frank's home now, they will be acting without Johnny's consent". However, this error did not appear to affect the results, which did not change substantially when Consent 2 was included in the overall consent composite.

instance (e.g., "Marvin made this particular decision rationally"; "When Marvin said 'yes' to having surgery, he was not being true to himself").

Following these manipulation checks, participants answered four binary-choice attention checks (e.g., "At the time the doctor suggested surgery, Marvin was *capable/incapable* of thinking through his choices and deciding based on the pros and cons")." We pre-registered that we would exclude participants who failed one or more of these attention checks. Finally, participants completed an exploratory measure that asked them to describe the reasoning behind their consent judgments, and a demographic survey in which they reported, in fixed order, their political views, bilingual status, age, gender, education, income, and race.

2.2. Results

2.2.1 Manipulation checks

As intended, the Autonomy manipulation was perceived as affecting capacities for rational and authentic decision-making (see Appendix A for full details). The agents in both the Exercises Capacity and Mere Capacity conditions were judged to possess capacities for rational and authentic decision-making, but the agent in Lacks Capacity was not. When it came to judgments of whether the agent made this particular decision rationally and authentically, by contrast, the agent in Mere Capacity garnered significantly lower ratings than did the agent in Exercises Capacity.

2.2.2 Judgments of Valid Consent and Right Choice

The three consent items created a reliable scale (α = .74); thus, they were averaged together to create a composite measure of consent. We analyzed this composite using the lme4 and lmerTest packages in R (Bates, Maechler, Bolker, & Walker et al. 2014; Kuznetsova, Brockhoff, & Christensen, 2017). Data were fit to a linear mixed model with autonomy condition included as a fixed factor and domain included as a random factor (random intercepts only) in all models. Significance of fixed effects was assessed via *t*-tests using Satterthwaite's method.

As predicted, participants' judgments conformed to the Mere Capacity hypothesis: lacking capacity had a large undermining effect on judgments of valid consent, whereas mere failure to exercise capacity did not. Compared to the Exercises Capacity baseline (M = 5.98, SD = 1.08), the Lacks Capacity condition yielded significantly lower agreement that the agent gave valid consent (M = 4.78, SD = 1.41), b = -1.21, SE = 0.14, t = -8.81, p < .001, 95% CI [-1.48, -0.94]. The Mere Capacity condition, by contrast, failed to yield lower agreement that the agent gave valid consent. In fact, participants gave *higher* ratings of valid consent in the Mere Capacity condition (M = 6.38, SD = 0.84) than in the Exercises Capacity condition, b = 0.36, SE = 0.15, t = 2.35, p = .019, CI [0.06, 0.65]. See Figure 1. An exploratory ANOVA revealed no interaction between domain and autonomy condition, $F(4, 355) = 0.83$, $p = .51$.[91] There was, however, a main effect of domain, $F(2, 355) = 9.99$, $p < .001$: participants gave overall lower ratings of consent in the police entry vignette.

[91] In line with this, exploratory pairwise comparisons indicated that within each domain, the Lacks Capacity condition led to significantly lower ratings of valid consent.

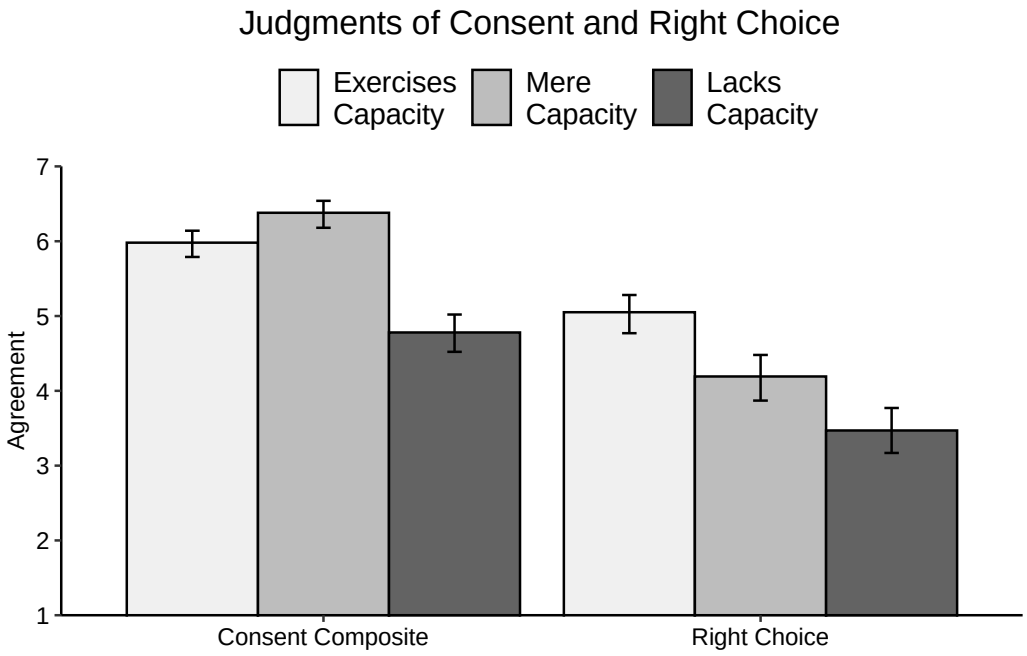


Figure 1. Respondents (n = 364) judged an agent who possesses autonomous decision-making capacities but failed to exercise them to having given no less valid consent than an agent who both possessed and exercised such capacities. An agent who lacked autonomous decision-making capacities was viewed as giving less valid consent. This pattern was observed despite participants believing that saying 'yes' was more likely to be the right choice for the fully autonomous agent compared to the Mere Capacity agent, who in turn was more likely to have made the right choice compared to the Lacks Capacity agent. Error bars represent bootstrapped 95% confidence intervals.

Judgments of Right Choice yielded a different pattern to judgments of valid consent (Fig. 1). We predicted that assessments of whether having surgery (or having sex, allowing the police to enter) was the right choice for the agent *would* be affected by both the possession and the exercise of autonomous capacities. As predicted, participants exhibited lower agreement that the Mere Capacity agent (M = 4.19, SD = 1.50) made the right choice compared to the Exercises Capacity agent (M = 5.05, SD = 1.57), b = -1.01, SE = 0.19, t = -5.43, p < .001, CI [-1.38, -0.65]. Judgments were even lower among the Lacks Capacity condition (M = 3.47, SD = 1.73) compared to the Mere Capacity condition, b = -0.58, SE = 0.19, t = -3.05, p = .002, CI [-0.96, -0.21].[92]

[92] An ANOVA revealed no significant interaction between condition and domain on judgments of whether the agent made the right choice, $F(4, 355) = 1.76$, p = .14. There

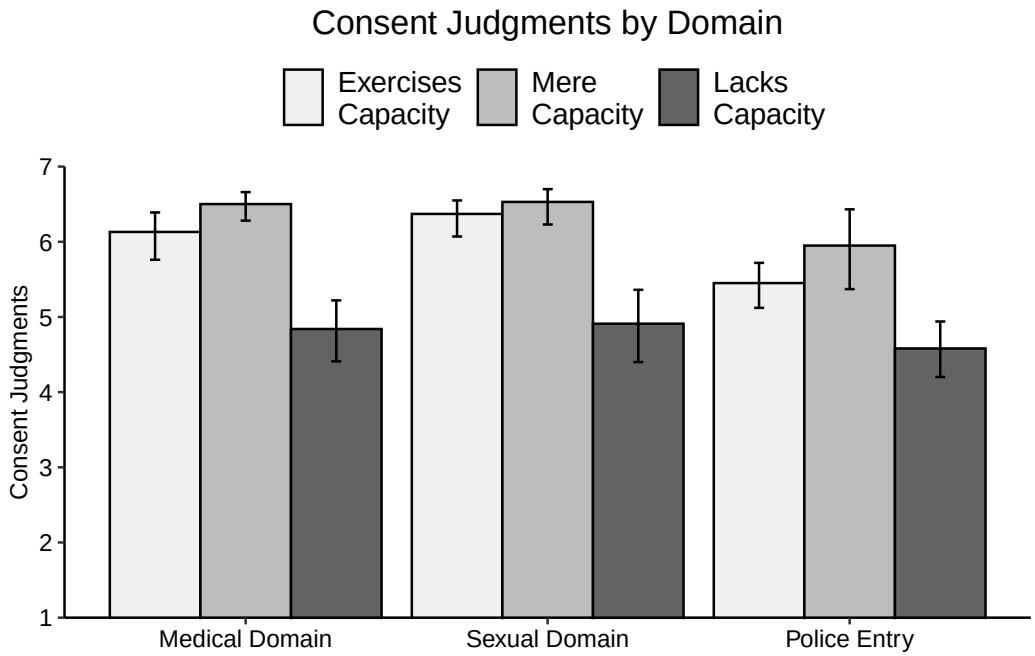


Figure 2. Across three domains, participants (n = 364) judged an agent who possesses autonomous decision-making capacity but fails to exercise it as validly consenting just as much as an agent who both possesses and exercises such capacities, while an agent who lacks autonomous decision-making capacity was viewed as giving less valid consent. Error bars represent bootstrapped 95% confidence intervals.

2.3. Discussion

The results of Study 1 suggest that autonomy plays an important role in the folk concept of valid consent: when agents lack the capacity to make autonomous decisions, as in the Lacks Capacity condition, judgments of the validity of consent are reduced. Notably, this relationship between autonomy and consent was consistent across different domains, including medical consent, sexual consent, and consent to police entering the home, suggesting that a domain-general concept of valid consent may be operative in reasoning

was a main effect of scenario, $F(2, 355) = 62.49$, $p < .001$: participants more strongly believed that the agent in the medical scenario (M = 5.36, SD = 1.12, n = 128) made the right choice as compared to the agents in the policing (M = 3.56, SD = 2.03, n = 112) or the sexual consent (M = 3.82, SD = 1.44, n = 124) scenarios.

about consent across these very different contexts.[93] Further, the results suggest people use a concept of valid consent that goes beyond the mere question of whether the agent said "yes" or "no", since all our conditions involved agents saying "yes".

However, folk judgments were not consistent with the Exercises Capacity view of valid consent, according to which the making of a decision in an autonomous manner is crucial for determining the validity of consent. Participants did not rate the consent of the Mere Capacity agent as any less valid compared to the Exercises Capacity agent, even though the Mere Capacity agent was described as saying "yes" without any regard whatsoever for the pros or cons or their own values.

This finding was observed despite the fact that participants recognized that failing to exercise autonomous decision-making has a strong impact on quality of choice, as indicated by our manipulation checks and our "Right Choice" measure: participants tended to disagree that the Mere Capacity agents were deciding rationally, and tended to disagree or neither agree nor disagree that they were being true to themselves and making the right

[93] Exploratory analyses indicated that the Police Entry vignette yielded lower ratings of valid consent overall compared to the Sexual and Medical consent vignettes. A number of features of the vignette may have contributed to this. Firstly, the consent in the Police Entry vignette may have been perceived to be less free due to stronger power dynamics: unlike in the other vignettes, the people requesting consent in Police Entry were police officers, and thus may have been perceived to be in a greater position of power and authority over the consenter compared to the power and authority possessed by a doctor or sexual partner. Moreover, there were two police requesting consent in that vignette, whereas the others described just one person seeking consent. Consistent with this speculation, an exploratory measure found that the consenter was perceived to be less free in the Police Entry scenario compared to the other vignettes. Second, participants may have been motivated to give lower ratings of consent because they disapprove of certain kinds of searches by police, irrespective of the citizen's consent. Indeed, an exploratory measure of moral judgment suggested that participants were less likely to agree that it was morally permissible for the officers to proceed compared to the doctor or the sexual partner. Further research would need to confirm whether these domain differences are robust or simply an artefact of this vignette.

choice. By contrast, participants overwhelmingly agreed that the Exercises Capacity agents were deciding rationally and being true to themselves, and tended to agree that they were making the right choice. Nevertheless, this assessment did not lead participants to judge that their consent was more valid. It therefore seems that while participants appear to view autonomous decision-making as *valuable,* they do not view it as required for valid consent.

Thus, participants' judgments accorded with the Mere Capacity view of valid consent, according to which it is only required that the consenter possesses capacities for autonomous decision-making, even if they do not exercise them.

Surprisingly, we found that judgments of the validity of consent were slightly *higher* for the Mere Capacity agent. Because the vignette in this condition emphasized that the agent in question failed to make his decision in a rational manner that was sensitive to his own values, even though they had the capacity to do so, participants may have felt that the vignette was implicitly suggesting that failing to make the decision autonomously can undermine consent. We suspect that participants may then have given even stronger ratings of valid consent (close to ceiling) to express disagreement with this implicit suggestion.

3. Study 2

Study 2 sought to determine the robustness of the finding that whether an agent possesses autonomous decision-making capacities, but not whether the agent makes a decision in an autonomous manner, is crucial for judgments of valid consent, while overcoming some limitations of Study 1.

First, Study 2 investigates whether the finding is observed even if participants are given descriptions of more concrete features of irrational or inauthentic decision-making. In Study 1, participants were presented with a description of a decision-making process at a high level of abstraction. Furthermore, the vignettes offered no information about *why* the Mere Capacity agent failed to make the decision autonomously. It remains possible, then, that participants would take the failures to engage in autonomous decision-making to reduce the validity of consent if offered more concrete descriptions of irrationality or inauthenticity. If so, this would suggest that the folk concept coheres with the Exercises Capacity view of valid consent after all. Consequently, Study 2 was designed to test whether our findings would extend to more concrete, realistic, and varied failures of autonomy. We did this by using vignettes involving a variety of ways in which an agent might fail to make their decision autonomously: by acting impulsively, giving in to peer pressure, basing a decision on irrational beliefs, or making an uninformed choice despite the availability of crucial information.

Second, in Study 1 participants were given no explicit information about whether saying "yes" was the right outcome for the agent. Because of this, even if the Mere Capacity agent did not make the decision in a rational or authenticity-preserving *way,* the scenario left open the possibility that the activity in question was in fact the *option* that best cohered with their reasons and values. For instance, participants might judge that having corrective surgery is likely to cohere with the values of a competent agent seeing the doctor about their ankle pain. Study 2 sought to use a more stringent test of whether the folk concept of valid consent merely requires autonomous decision-making capacities or whether it additionally requires the making of *autonomous decisions* by using vignettes that specify that the agent in the Mere Capacity and Lacks Capacity conditions makes the wrong decision relative to their reasons and values, while the Exercises Capacity agent makes the right decision.

Finally, we also included an additional pair of manipulation checks to ensure that our autonomy conditions mirrored folk judgments about autonomy—namely the ability to shape one's life freely according to one's values. For instance, we asked participants to rate their agreement with "The way Marvin made this particular decision expressed an ability to shape his life freely according to his own values and what is right for him." This allowed us to rule out the possibility that participants think the Mere Capacity agent is deciding in a way that is equally as autonomous as the Exercises Capacity agent (e.g., because participants surmise that the Mere Capacity agent is making an autonomous decision *not* to exercise his capabilities, thus expressing a kind of meta-autonomous desire to make the choice in a nonautonomous way).

3.1. Methods

Sample size, predictions and analyses were pre-registered (https://aspredicted.org/blind.php?x=gs43rp).

3.1.1. Participants

In line with our pre-registration, we recruited 600 participants on Amazon Mechanical Turk. After excluding participants who failed at least one of the attention checks, we were left with a sample of 384 participants (49.6% male, 49.6% female, 0.8% other gender; median age 37 years).

3.1.2. Design

Participants were randomly allocated to one of twelve conditions in a 3 (Autonomy: Exercises Capacity; Mere Capacity; Lacks Capacity) by 4 (Failure Type: Impulse; Peer Pressure; Uninformed; Irrational Superstition) between-subjects design.

3.1.3. Procedure and materials

Each participant was presented with a vignette in which an agent, Marvin, is facing a choice about whether to undergo an elective surgery and ends up saying "yes." Before saying "yes," he faces the possibility of having his decision non-autonomously determined by impulse, peer pressure, lack of information, or irrational superstition (depending on Failure Type). Autonomy conditions determined whether he possessed the capacity to make the decision according to his values and reasons, and whether or not he in fact did so.

As in Study 1, the Exercises Capacity agent is described as having the capacity to make the decision in an autonomous, rational manner (e.g., able to make decisions for himself and override impulses when they are inappropriate), as in fact doing so (e.g., resisting an initial, impulsive reaction to say "no", and instead thinking things through in a rational and authentic way), and as a result saying "yes." This is described as the right choice for him.

The Mere Capacity agent is described as having these very same capacities, but as not exercising them (e.g., giving in to an initial impulsive reaction to say "yes" without thinking things through in a rational or authentic way) and as a result saying "yes" even though that is *not* the right choice for him. The Lacks Capacity agent lacks these capacities altogether (e.g., he is not able to make decisions for himself and resist inappropriate impulses) and

thus says "yes" even though it is not the right choice for him. See Table 2 for an illustration of the vignettes used in the Impulse condition; full text for all conditions is available in Appendix B.

As in Study 1, participants rated their agreement with a number of statements presented in a random order on a seven-point Likert scale. We used the same three measures to assess judgments of valid consent (e.g., "Marvin's 'yes' didn't count as consent").

In addition to the manipulation check measures from Study 1, we added two manipulation check questions eliciting judgments of autonomy:

> **Capacity for autonomy:** Marvin has the ability to shape his own life freely according to his own values and what is right for him.
>
> **Decided autonomously:** The way Marvin made this particular decision expressed an ability to shape his life freely according to his own values and what is right for him.

Following the Likert-scale measures, we again included attention checks. Participants were asked three binary-choice questions, presented in random order, concerning (1) the agent's capacities; (2) the way they made the decision, and (3) whether they made the right choice. (In other words, in Study 2 "right choice" was included as a manipulation check rather than as a main dependent measure.) For instance, the attention checks in the Impulse condition were: "Marvin is/is not able to resist and overcome impulses."; "Marvin made this particular decision by thinking it through properly/on an impulse."; "Having surgery was/was not the right choice for Marvin."

Intro paragraph (all conditions):		
Marvin has been in physical therapy for ankle pain. One day his doctor asks him whether he wishes to undergo elective surgery to repair the tendon. The doctor explains that the surgery carries some risks, as all surgeries do, but if all goes well it could potentially completely cure his ankle pain.		
Exercises Capacity:	Mere Capacity:	Lacks Capacity:
Marvin feels an initial impulse to simply say 'no' to surgery.	Marvin feels an initial impulse to simply say 'yes' surgery.	Marvin feels an initial impulse to simply say 'yes' surgery.
Marvin is an intelligent, able adult, fully capable of making decisions for himself and controlling impulses when they are inappropriate.	Marvin is an intelligent, able adult, fully capable of making decisions for himself and controlling impulses when they are inappropriate.	Marvin is **not able and intelligent like most adults who are fully capable of making decisions for themselves**: he is completely **incapable** of controlling impulses, even when they are inappropriate.
And he does so in this instance. Although he feels an initial impulse to avoid surgery, he thinks things through carefully, and makes his decision with careful regard for the pros and cons, and whether surgery aligns with his personal values and what he really wants. Because of this, Marvin says 'yes' to the surgery.	**But he does not do so in this instance**. **Acting on an initial impulse** to have the surgery, he **doesn't think things through even a little bit**, and **pays absolutely no attention** to the pros and cons, or whether surgery aligns with his personal values and what he really wants. He simply says 'yes' to the surgery **on an impulse**.	**So he does not do so in this instance**. **Acting on an initial impulse** to have the surgery, he **doesn't think things through even a little bit**, and **pays absolutely no attention** to the pros and cons, or whether surgery aligns with his personal values and what he really wants. He simply says 'yes' to the surgery **on an impulse**.
If he had not resisted his initial impulse and made a decision based on thinking things through properly, Marvin would have said 'no', despite surgery being the right choice for him.	If he had resisted his initial impulse and made a decision based on thinking things through properly, Marvin would have said 'no', as it is **not the right choice for him**.	If he had resisted his initial impulse and made a decision based on thinking things through properly, Marvin would have said 'no', as it is **not the right choice for** him.

Table 2. Vignette used in the Impulse failure type condition, with variations according to Autonomy condition. Boldface type is used here for emphasis; it was not used in the stimuli presented to participants.

Finally, participants completed an exploratory question explaining their answer and a demographic survey as in Study 1.

3.2. Results

3.2.1. Manipulation checks and autonomy judgments

Again, participants' conceptual judgments exhibited the predicted patterns (see Appendix A for full details). As intended, both the Exercises Capacity and Mere Capacity agent were judged to possess capacities for rational, authentic, and autonomous decision-making, but the Lacks Capacity agent was not. Also as intended, the agent in Mere Capacity was rated significantly lower on having made this particular decision rationally, authentically and autonomously compared to the Exercises Capacity agent.

3.2.2. Judgments of Valid Consent

Again, we created a composite measure of judgments of valid consent (α =.75). We fit a linear mixed model with autonomy condition included as a fixed factor and failure type included as a random factor. The model was a singular fit because of an estimate of zero variance for the intercept, suggesting that the model did not warrant a random effect of failure type (e.g., Henne et al., 2019). Hence, we simplified the model, using a linear model with no random effects.

In line with our hypotheses, participants rated consent as higher in the Exercises Capacity condition (M = 6.15, SD = 1.23) than in the Lacks Capacity condition (M = 5.21, SD = 1.28), b = -0.94, SE = 0.15, t = -6.42, p < .001, CI [-1.22, -0.65], but no higher than in the Mere Capacity condition (M = 5.93, SD = 1.17), b = -0.22, SE = 0.16, t = -1.38, p = .17, CI [-0.54, 0.09]. See Figure 3. From a Bayesian perspective, these results provided support for the absence of an effect of Exercising Capacity vs. Mere Capacity on consent

judgments, though the evidence for this null result is "weak" or "anecdotal," falling short of "positive" or "substantial" ($BF_{10} = 0.36$).[94]

In line with our pre-registration, we confirmed via an exploratory two-way ANOVA that the interaction between autonomy condition and failure type was not significant, $F(6, 372) = .58$, $p = .75$.[95] Despite this, we conducted exploratory post hoc pairwise comparisons within the Uninformed failure type to explore the possibility that this vignette yielded a different effect of condition (see Figure 3). These post hoc tests revealed that consent judgments did not differ significantly between Exercises Capacity and Mere Capacity conditions in the Uninformed vignette ($p = .104$), but they did between the Exercises Capacity and Lacks Capacity conditions ($p = .001$).

[94] This Bayesian *t*-test was not pre-registered, but was helpfully suggested by a reviewer.

[95] This finding further justified collapsing across Failure Type. The ANOVA revealed a main effect of autonomy condition, $F(2, 362) = 21.36$, $p < .001$. There was no main effect of Failure Type, $F(3, 372) = .54$, $p = .66$.

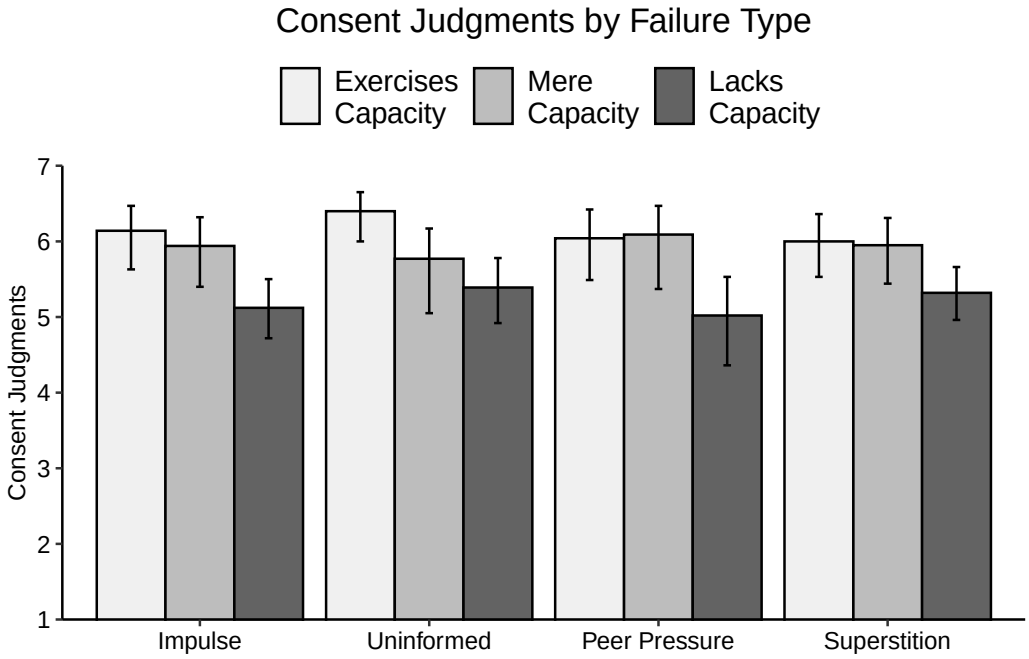


Figure 3. Results from Study 2 for mean agreement that the agent gave valid consent. Compared to an agent who possesses and exercises autonomous decision-making capacities, participants (n = 381) were not less likely to judge that an agent gave valid consent when they possessed autonomous decision-making capacity but failed to exercise it in some way (through impulse, being uninformed, peer pressure, or irrational superstition). Participants were significantly less inclined to agree that an agent who lacked autonomous decision-making capacity gave valid consent. Error bars represent bootstrapped 95% confidence intervals.

3.3. Discussion

Study 2 replicated and extended our main findings from Study 1. Once again, judgments of valid consent were related to the possession of autonomous capacities, but not to the exercise of those capacities. As in Study 1, it didn't matter whether the consenter actually made the decision for himself in an autonomous way: participants regarded the consent of an agent with autonomous decision-making capacities to be equally valid even if he did not make this particular decision autonomously. This was so even though we explicitly specified that the Mere Capacity agent would have said "no" had he made the decision

properly, because it wasn't the right choice for him, whereas the Exercises Capacity agent would have said "yes."

We found this pattern for a number of different "Failure Types" that invoked particular, concrete capacities involved in autonomous decision-making. For instance, judgments of consent were reduced when the agent lacked the general *ability* to control impulses, to resist peer pressure, to get relevant information, or to distinguish reasonable from unreasonable ways of making decisions. Judgments of the validity of the agent's consent were not lower when the agent retained those capacities but said "yes" on impulse, because of peer pressure to say "yes," because of a failure to inform themselves, or because of an irrational and unreasonable superstition.

These findings thus provide stronger evidence that whether an agent makes an autonomous decision is not important for judgments that consent was valid, even though having the capacity to make an autonomous decision does matter.

While this main finding was consistent across different Failure Types, the Uninformed category appeared to creep towards subtly different results. While post hoc analyses indicated that only the Lacks Capacity condition yielded significantly lower judgments of valid consent compared to Exercises capacity, in line with our main findings, the Uninformed Mere Capacity agent appeared closer to the Lacks Capacity agent than the Exercises Capacity agent.

One possible explanation for the appearance of a different pattern for the Uninformed category is that the vignette described a case of consent to a medical procedure. Unlike other consent contexts, medicine is associated with explicit institutional and cultural norms

of the importance of so-called "*informed consent*." Awareness of this norm could have reduced ratings in the Mere Capacity condition, either because participants agreed with this domain-specific norm, or because awareness of this norm affected participants' judgments of consent.

Another possibility, of course, is that the apparent deviation from the main pattern in this condition is simply noise due to random error. Future research could confirm whether attempting to inform oneself is important for judgments of valid consent, preferably using a larger array of vignettes including non-medical contexts.[96]

4. Study 3

Studies 1 and 2 establish that judgments of valid consent vary according to autonomy. But in what respect does it matter whether a "yes" is judged to be valid or invalid? Are these judgments superficial, merely reflecting how words like "consent" are used, without playing an important role in moral reasoning? We designed Study 3 to investigate whether participants intuitively embrace a normative concept of valid consent akin to that employed

[96] It is of course possible that there is a specific exception to the general rule that exercising autonomy does not affect judgments of valid consent, namely that attempting to inform oneself adequately is considered crucial for valid consent, even in non-medical contexts. This would be a surprising result, but further studies could certainly investigate it, preferably using multiple vignettes and inclusion of non-medical contexts. However, we do not believe that *knowledge* is crucial for judgments of valid consent. In other words, even if attempting to inform oneself is important for judgments of valid consent, it's unlikely that people believe this attempt would need to be successful, resulting in accurate information and beliefs, to perceive consent as valid. In studies reported by Sommers (2020), participants overwhelmingly treated cases of material deception as consistent with valid consent; these are cases in which a person only gives valid consent because they have been lied to and thus have false beliefs or are ignorant about crucial information (e.g., only consenting to sex because one does not believe the sexual partner has HIV, or is married).

in philosophical and legal spheres by examining whether their judgments of valid consent carry important downstream consequences for moral reasoning.

How might consent affect moral reasoning? One obvious answer is that it's more morally wrong to perform an action (e.g., take something that belongs to someone else) if it was not validly consented to. For example, recent work by Rodríguez-Arias and colleagues (2020) on the topic of physician-assisted suicide suggests that consent has important implications for moral judgments. Specifically, their studies showed that consent is responsible for the morally motivated causal distinction between 'killing' and 'letting die': if a patient consents to ending their life, then the cause of death is judged to be the patient's illness, but if the patient does not consent, the doctor is seen as 'killing' the patient. It appears that here the giving of consent affects whether the doctor's action is construed as permissible assistance or as something morally wrong.

However, according to the philosophical and legal concept, the effect of valid autonomous consent should be even more far-reaching than its effect on judgments of the moral permissibility of performing the consented-to act: consent should additionally determine various parties' rights and obligations.

As described earlier, the presence or absence of valid consent is thought to be crucial even if the consented-to act is morally wrong. Adulterous sex might be morally wrong regardless of consent, for instance, but the presence or absence of consent is thought to determine both the severity and criminality of the wrong. Crucially, the presence or absence of consent carries consequences for rights and obligations, such as the right or duty of third parties to forcibly intervene in the case of rape but not adultery; of police to arrest the rapist; the right of the sexual partner to seek redress and justice in the case of rape; and so

on. Consent similarly functions to alter rights and obligations in other domains: consent is thought to determine whether or not a signer is bound by the terms of a contract. in the case of ownership, consent determines if the taking of an item constitutes a sale, and thus the successful transfer of ownership rights to a new owner (rather than, e.g., a theft).

In Study 3, we chose to study whether the effect of autonomy on judgments of valid consent carries downstream consequences for judgments of *ownership transfer*. As discussed in the introduction, prior research suggests that the presence or absence of consent plays a role in the ascription of ownership rights. Prior research suggests that adults and children reason about violations of ownership rights in a very similar way to how they reason about violations of bodily rights (Van de Vondervoort & Friedman, 2015), which supports our hypothesis that the same concept of valid consent should extend to this domain. Studying ownership also allowed us to extend the findings from Study 1 and Study 2 to a new context.

Study 3 employed a vignette in which the owner of an item gives consent to a second party (the seller) to sell that item; the item is then bought by an innocent third party (the buyer). Introducing a blameless third party provided a clean way to assess the potential downstream consequences of the validity of consent. For instance, if an owner consents to selling something only because they haven't thought it through properly, and the seller knows this, participants might reasonably judge that the seller is doing something morally wrong, since they are knowingly doing something that's bad for the consenter/original owner. Nevertheless, based on Studies 1 and 2, we predicted that participants would judge such a transaction to be consensual. Thus, assessing the allocation of rights to an innocent third party (the buyer) allowed us to differentiate the effect of autonomous consent on the

morality of the consented-to action, on the one hand, and its impact on rights, on the other.

We predicted that whether participants judge that ownership rights have transferred to the buyer of an item would depend on whether they judged that the original owner gave valid consent to the sale, which itself would depend on the possession of autonomous decision-making capacities. Thus, we hypothesized that if the consenter (the original owner) has autonomous decision-making capacities, then consent will be judged to be valid, and the buyer will be judged to have been conferred ownership rights. By contrast, if the consenter lacks autonomous decision-making capacities, then participants will judge consent to have been invalid; in turn, we expect them to judge that the consenter retains ownership rights and that the buyer has failed to gain ownership rights. Importantly, we expected that whether the consenter *exercises* autonomous decision-making capacities would not have an impact on valid consent or the transfer of ownership rights.

4.1. Methods

Sample size, predictions and analyses were pre-registered (https://aspredicted.org/blind.php?x=2fv78s).

4.1.1. Participants

In line with our pre-registration, we set out to recruit 300 participants on Amazon Mechanical Turk; 303 were recruited due to random M-Turk software error. After excluding participants who failed at least one of the attention checks (N=143) and additional participants who wrote gibberish (N=7), we were left with a sample of 153

participants (59.5% male, 40.5% female; median age 34 years). Responses classified as gibberish included nonsense, responses that were identical to those of other participants, copy-and-pasted material from the experiment, and irrelevant copy-and-pasted material from the internet.

4.1.2. Design

Participants were randomly allocated to one of three Autonomy conditions in a fully between-subjects design: Exercises Capacity; Mere Capacity; Lacks Capacity.

4.1.3. Procedure and materials

All participants were presented with a vignette in which an agent, Jessica, is in hospital recovering from a procedure. The full text is available in Appendix B. Sam comes to visit Jessica and asks for consent to sell her diamond bracelet—a bracelet which is very precious to her but which he thinks could make them a lot of money.

In all conditions, Jessica says "yes", but her autonomy differs according to condition. In the Exercises Capacity and Mere Capacity conditions, Jessica is not on heavy medication. She is described as being "perfectly capable of weighing up pros and cons, thinking through choices she faces, and making decisions based on what is best for her, which options align with her personal values, and what she really wants," and although she is on medication, "it's only ibuprofen and some antibiotics. In fact, she feels calm and lucid, and nothing is interfering in any way with her ability to think or make decisions."

In the Exercises Capacity condition, she goes on to use these capacities to make her decision to consent in an autonomous way: "using her ability to make decisions according to her own values and what is best for her, Jessica says 'yes' after thinking things through very carefully, with careful regard for the pros and cons and whether it's what she really wants." By contrast, in the Mere Capacity condition, "despite her ability to make decisions according to her own values and what is best for her, Jessica just says 'yes' to the sale without thinking things through even a little bit, and with absolutely no regard for the pros and cons or whether it's what she really wants."

Finally, in the Lacks Capacity condition, Jessica is on heavy medication that undermines her decision-making capacities: "the medication she is on is incredibly powerful and is severely interfering with her ability to think and make decisions. Indeed, in her current state she is completely incapable of weighing up pros and cons, thinking through choices she faces, or making decisions based on what is best for her, which options align with her personal values, or what she really wants." Consequently, she does not make her decision in her autonomous way: "so Jessica says 'yes' to the sale without thinking things through even a little bit, and with absolutely no regard for the pros and cons, or whether it's what she really wants."

For all participants, Sam is described as being only motivated by making money, and not caring about which decision will make Jessica happy; in fact, he proceeds even though he suspects that Jessica will regret her decision.

The vignette then goes on to describe the completion of the sale by Sam to a blameless third party, Melanie, who sees the bracelet advertised as for sale online, pays for it, and receives the bracelet in the mail two days later.

After reading the vignette, participants rated their agreement with a series of statements, presented in a random order, on seven-point Likert scales. The vignette remained visible to participants for their reference.

We measured judgments of valid consent, morality, and ownership rights using agreement with a seven-point Likert scale. The three consent measures closely matched those used in Studies 1 and 2 (e.g., "Jessica's 'yes' didn't count as consent"). Moral judgments of the consented-to action were measured as follows:

> **Morality:** Under these circumstances, it was morally wrong for Sam to proceed with selling the bracelet. (reverse-scored)

We also used five new "ownership transfer" measures using a seven-point Likert scale from "strongly disagree" to "strongly agree." These were designed to assess judgments of the rights and obligations of the buyer, Melanie.

> **Ownership Transfer 1:** The bracelet does not truly belong to Melanie. (reverse-scored)

> **Ownership Transfer 2:** Even if Melanie was told about the bracelet's true history, it would be morally acceptable for her to keep the bracelet if that's what she wanted to do.

Ownership Transfer 3: If Melanie was told about the bracelet's true history, it wouldn't just be nice of her to give the bracelet back to Jessica: it would be her *moral duty* to give it back. (reverse-scored)

Ownership Transfer 4: Melanie should be forced to return the bracelet. (reverse-scored)

Ownership Transfer 5: A good law would require the bracelet to be returned to Jessica under these circumstances. (reverse-scored)

At the start of Ownership Transfer 2-5 measures, the question instructions clarified, "Assume that Melanie could return the bracelet to Jessica and get her money back. Do you agree with the following statement?" This was to make sure that participants' answers reflected judgments about whether ownership rights to the bracelet transferred to Melanie, and not concerns about Melanie's money. For this reason, Ownership Transfer measures 2-5 were presented one after the other (in a random order) to aid participant comprehension, instead of fully randomizing the order of all measures.

In addition, we included one binary multiple-choice measure of ownership judgments as follows:

Ownership, binary: Who is the rightful owner of the bracelet? [Options: Jessica/Melanie]

We additionally included manipulation checks as in Study 1.

Participants then completed two multiple-choice comprehension checks ("Which is correct? Jessica's medication interfered with her ability to think/Jessica's medication DID NOT interfere with her ability to think" and "Which is correct? Jessica said 'yes' WITH regard for whether she really wanted to sell the bracelet / Jessica said 'yes' WITHOUT regard for whether she really wanted to sell the bracelet").

Finally, participants answered an exploratory open-ended question explaining their reasoning, provided demographic information, and were debriefed.

4.2. Results

4.2.1 Manipulation checks

Our manipulations checks confirmed that the autonomy conditions successfully manipulated participants' judgments of Jessica's autonomy in largely the desired way, with minor exceptions.

Judgments of whether Jessica *made the decision* to consent rationally and authentically showed the predicted pattern. While agreement that she made this decision rationally and authentically was high in the Exercises Capacity condition, participants tended to disagree that she made this decision rationally and authentically in the Mere Capacity and Lacks Capacity conditions.

Judgments of whether Jessica possessed the capacity for autonomous decision-making largely, but not entirely, conformed to predicted patterns. As expected, participants tended to disagree that Jessica had the capacity to make decisions rationally ($M = 2.11$, $SD = 1.60$)

and authentically (M =2.25, SD = 1.70) in the Lacks Capacity condition, while participants tended to agree that she had the capacity in both the Exercises Capacity agent and the Mere Capacity conditions. However, contrary to expectations, agreement with these measures was significantly lower in the Mere Capacity condition compared to the Exercises Capacity condition. Jessica was regarded as less capable of rationality in the Mere Capacity condition (M = 4.96, SD = 1.73) compared to the Exercises Capacity condition (M = 5.88, SD = 1.12), t_{Welch} (83.00) = 3.08, p = .003. Additionally, she was regarded as having lower capacity for authenticity in the Mere Capacity condition (M = 4.80, SD = 1.86) compared to the Exercises Capacity condition (M = 5.74, SD = 1.18), t_{Welch} (82.25) = 2.96, p = .004.

4.2.2. Judgments of Valid Consent

A composite measure of judgments of valid consent was created by averaging together Consent 1, Consent 2 (reverse-scored), and Consent 3 (reverse-scored) (α =.89).

A one-way ANOVA revealed a significant effect of autonomy condition on judgments of valid consent, F(2, 150) = 59.82, p < .001 (Fig. 4). Pairwise comparisons of estimated marginal means indicated that, as predicted, participants showed lower agreement that the agent in Lacks Capacity gave valid consent (M = 2.75, SD = 1.53) relative to the Mere Capacity agent (M = 4.80, SD = 1.56), b = 2.05, SE = .28, t = 7.25, p < .001, CI [-2.61, -1.49].

Participants also showed slightly lower agreement that the agent in the Mere Capacity condition gave valid consent compared to the Exercises Capacity condition (M = 5.84, SD = 1.27), b = 1.03, SE = 0.31, t = 3.36, p = .001, CI [-1.64, -0.43].

Nevertheless, mean ratings of valid consent in the Mere Capacity condition (t (48) = 3.60, p < .001, d = .51) and the Exercises Capacity condition (t (42) = 9.46, p < .001, d = 1.44) were both significantly above midpoint. By contrast, mean rating of valid consent in the Lacks Capacity condition was significantly below midpoint (t (60) = -6.35, p < .001, d = .81).[97]

4.2.3. Judgments of Ownership Rights

The Likert ownership rights measures 1 and 3-5 were reverse-scored, so that for all measures, higher score indicates greater agreement that ownership rights have been successfully transferred (i.e. that the buyer, Melanie, is now the rightful owner of the bracelet, and that the previous owner, Jessica, no longer has rights to it). The five measures of ownership transfer showed very high scale reliability (α =.89), so a composite measure of ownership transfer was created by averaging together all five variables.

The effect of condition on judgments of ownership transfer mirrored the pattern found for judgments of valid consent, F(2, 150) = 35.60, p < .001 (Fig. 4). Pairwise comparisons of estimated marginal means indicated that participants showed lower agreement that ownership had been transferred in the Lacks Capacity condition (M = 3.18, SD = 1.40) relative to the Mere Capacity agent (M = 4.71, SD = 1.58), b = 2.35, SE = .29, t = 8.14, p < .001, CI [-2.92, -1.78]. Participants also showed slightly lower agreement that ownership transferred in the Mere Capacity condition compared to the Exercises Capacity condition (M = 5.53, SD = 1.36), b = 0.82, SE = .31, t = 2.69, p = .008, CI [-1.41, -0.22].

[97] We thank a reviewer for suggesting these additional analyses. These comparisons to midpoint were not pre-registered.

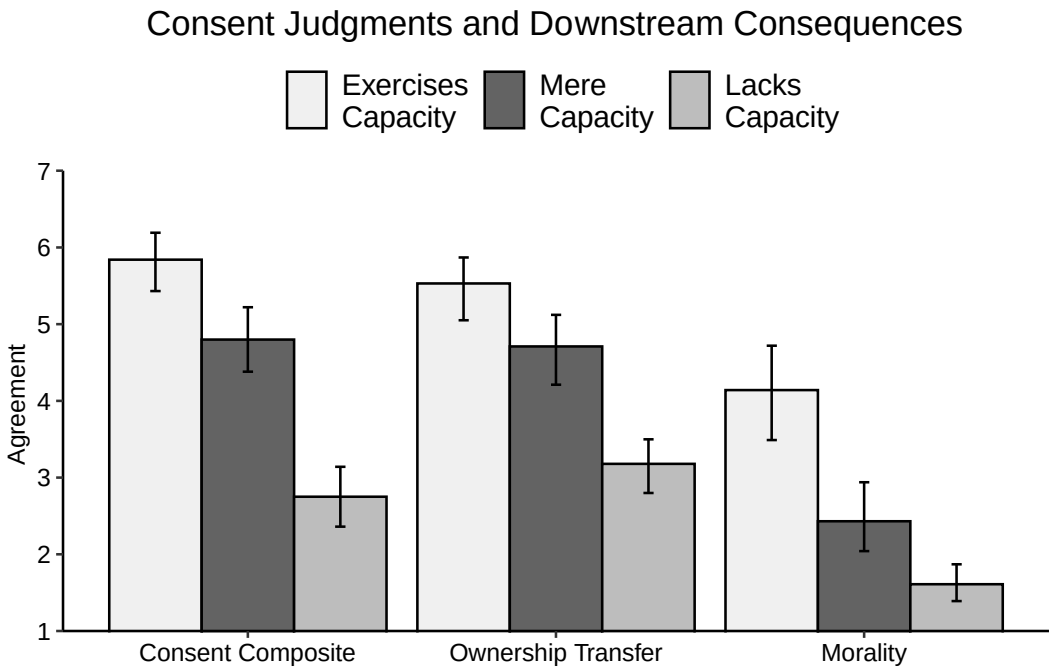


Figure 4. Results for Study 3 showing the effect of Autonomy condition on mean agreement that (i) there was valid consent, (ii) ownership rights transferred to the buyer, (iii) it was morally permissible for Sam to sell the bracelet. Error bars represent bootstrapped 95% confidence intervals.

The binary choice measure of ownership rights showed a similar pattern. In the Exercises Capacity and Mere Capacity conditions, the majority of participants said that the bracelet now belonged to Melanie (81.40% and 57.14% respectively), but in the Lacks Capacity condition only a minority chose Melanie as the rightful owner (21.31%). Fisher's exact tests confirmed that in the Lacks Capacity condition, significantly fewer participants said that the bracelet now belonged to Melanie compared to the Mere Capacity condition (Odds ratio (OR) = 0.21, $p < .001$, CI [0.08, 0.50]), and compared to the Exercises Capacity condition (OR = 0.06, $p < .001$, CI [0.02, 0.18]). The difference in ownership transfer judgments between Mere Capacity and Exercises Capacity was also significant (OR = 0.31, $p = .014$, CI [0.10, 0.86]).

4.2.4. Judgments of Morality

Moral judgments were reverse-scored so that higher scores indicate higher agreement that it was morally permissible to sell the bracelet, whereas lower scores indicate that it was not morally permissible to sell the bracelet. We observed a significant effect of Autonomy condition on moral judgments, $F(2, 150) = 35.27$, $p < .001$ (Fig. 4). Pairwise comparisons indicated that, as predicted, participants tended to disagree more that this was morally permissible in the Lacks Capacity condition, ($M = 1.61$, $SD = .94$) compared to the Mere Capacity condition ($M = 2.43$, $SD = 1.62$), $b = 0.82$, $SE = .29$, $t = -2.81$, $p = .005$, CI [-1.40, -0.24]), where participants also exhibited strong disagreement that Sam's conduct was morally permissible. The agent in Exercises Capacity ($M = 4.14$, $SD = 2.01$) was viewed as acting more permissibly as compared to the agent in Mere Capacity, $b = -1.71$, $SE = 0.32$, $t = -5.38$, $p < .001$, CI [-2.34, -1.08]).

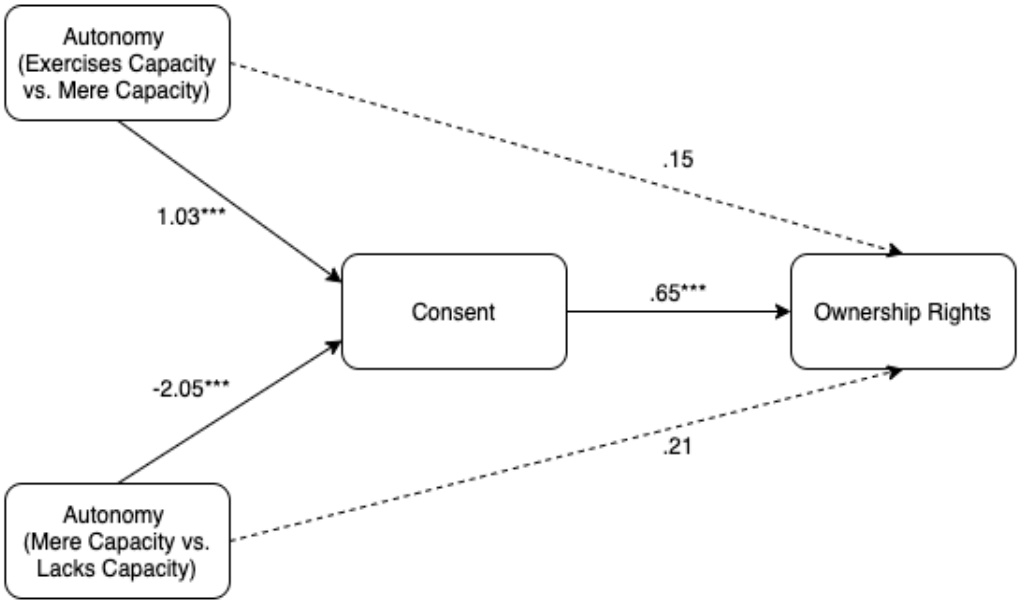


Figure 5. Mediation model showing that the effect of autonomy condition on judgments of ownership rights was fully mediated by the effects of autonomy condition on judgments of valid consent. Unstandardized regression coefficients are shown for each path. Asterisks indicate statistically significant relationships, *** = p <.001. Dotted arrows indicate non-significant relationships.

4.2.5. Relationship Between Judgments of Consent and Ownership

Consistent with our pre-registered analyses, we used linear regression to assess whether judgments of valid consent predicted transfer of ownership rights independently of the effect of autonomy condition. The regression confirmed that they did, b = -0.80, SE = .10, t = -8.16, p < .001, CI [-0.99, -0.61].

To further explore the relationship between judgments of valid consent and judgments of ownership rights, we conducted an exploratory multicategorical mediation analysis using Process v. 3.5, Model 4, with 10,000 bootstrapped samples (see Hayes, 2012; Hayes & Preacher, 2008) and 95% bias-corrected confidence intervals for indirect effects. Autonomy condition was entered as the independent variable and ownership rights judgments as the dependent variable, with consent judgments as the mediator. Results suggested that judgments of valid consent fully mediated the effect of Autonomy on judgments of ownership rights (see Table 3 and Figure 5).

Autonomy condition		*b*	S.E.	95% C.I.	*p*
Exercises Capacity vs. Mere Capacity	Indirect	.67	.20	.28, 1.06	-
	Direct	.15	.24	-.32, .62	.538
	Total	.82	.30	.22, 1.41	.008
Mere Capacity vs. Lacks Capacity	Indirect	-1.32	.24	-1.83, -.88	-
	Direct	-.21	.24	-.69, .27	.394
	Total	-1.53	.28	-2.08, -.98	<.001

Table 3. *b* shows unstandardized indirect, direct, and total effects of Autonomy condition on ownership rights via consent, with the Mere Capacity condition as the baseline for comparison, with standard errors and 95% confidence intervals.

4.3. Discussion

Study 3 demonstrated that the impact of autonomy on judgments of valid consent has important downstream consequences for the allocation of rights and obligations to others, here under the category of "ownership rights." When participants judged that consent to a sale was not valid, they judged that the buyer had no right to the item, that it was her moral duty to give it back, and that she should be forced to return it. By contrast, when participants judged that consent to a sale was valid, they judged that the buyer was morally permitted to keep the item, had no duty to return it, and should not be forced to return it.

Importantly, participants' perceptions of ownership rights appeared to be independent of their general moral judgments of the rightness or wrongness of the consented-to action. In this case, whether or not consent was valid transformed the perceived rights and obligations of an innocent third party (the buyer, Melanie), even though she was not involved in the acquisition of consent, had no reason to suspect that the relevant consent was compromised, was not culpable for any wrongdoing in the original consent transaction, and indeed did not have any reason to suspect any wrongdoing.

Furthermore, moral assessment of the actions of the consent-obtainer (Sam, the seller) came apart from judgments of the validity of the consent: moral judgments, but not judgments of valid consent, were strongly affected by whether the consenter exercised her capacities to make her decision in an autonomous manner. So, while Sam's selling the bracelet was considered somewhat morally permissible in the Exercises Capacity condition, it was considered morally wrong in the Mere Capacity condition, where the consenter was rash and failed to decide in a rational or authentic manner even though she had the mental

capacity to do so. By contrast, agreement that the consent was valid remained well above midpoint in both the Exercises Capacity and Mere Capacity conditions.

As in Studies 1 and 2, judgments of whether consent was valid in the first place were largely driven by the extent to which the agent possessed the capacity to make autonomous decisions at the time of deciding. While participants showed slightly lower levels of agreement that the Mere Capacity agent gave valid consent compared to the Exercises Capacity agent—unlike in Studies 1 and 2—in both conditions, agreement remained above midpoint. By contrast, ratings of valid consent dropped well below midpoint in the Lacks Capacity condition.

The finding the Mere Capacity agent's failure to decide in an autonomous manner slightly lowered perceptions of valid consent differed from predictions and from the results of Studies 1 and 2. Two possibilities could account for this inconsistency.

On the one hand, it's possible that the exercise of autonomous decision-making directly promotes perceptions of the degree to which consent is valid some circumstances, even though it is not normally regarded as necessary for moral transformation to take place (the decision of the Mere Capacity agent was regarded as morally transformative). We speculate that this could especially be the case when participants are less confident overall about whether the interaction is consensual—perhaps because the person asking for consent appears to be exploiting or coercing the consenter.

On the other hand, it's likely that whether an agent exercises their autonomous decision-making capacities is itself used as a cue to make judgments about whether the agent possesses autonomous capacities, which in turn drives judgments of the validity of their

consent. For example, whether a person behaves rationally affects the extent to which they are inferred to have the *capacity* to behave rationally. Supporting this interpretation, we observed the same pattern both for judgments of valid consent and for judgments of the agent's capacities: unlike in Studies 1 and 2, our manipulation checks showed that participants were less inclined to agree that the Mere Capacity agent really did have the capacity for rational and authentic choice to the same extent as they agreed that Exercises Capacity agent did. Although the vignette stated that the Mere Capacity agent possessed these capacities, participants may have been more uncertain about the true extent of these capacities, given that the agent was in a vulnerable context (i.e., recovering from surgery in hospital) and there was no further explanation for why they did not make a better decision. This was reflected in exploratory data we collected in which we asked participants to justify their answers, as a number of participants who were in the Mere Capacity condition but disagreed that the agent gave valid consent said they did not truly believe that the agent's capacity to choose was unimpaired. For example, one participant wrote that "[r]egardless of what was said about Jessica being in her right mins [sic] it was obvious that she wasn't". A number of others suggested that situational factors would have undermined the agent's capacity. For example, one participant wrote, "I think that Jessica was in a stressful experience…I know that she is not on any mind altering medications, but surgery itself is a lot to handle"; another wrote, "Jessica is in the hospital recovering from surgery. Even if her medication isn't getting in the way of her thinking, her situation will certainly have worn her down and exhausted her."

5. General Discussion

The present studies investigated the folk concept of valid consent and its relationship to autonomy. Study 1 showed that when agents lack autonomous decision-making capacities,

participants are less likely to view their consent as valid, while simply failing to exercise these capacities does not undermine perceived valid consent. Study 2 showed that failing to exercise autonomous decision-making in various concrete ways (e.g., giving in to impulse) similarly failed to reduce judgments of valid consent, even when it led the agent to make a choice that was not right for them. Again, however, lacking the capacity to make autonomous decisions (e.g., lacking the capacity to resist impulses) did reduce judgments that consent was valid. Finally, Study 3 showed that these consent judgments carry significant downstream consequences for judgments of ownership rights.

Our studies therefore suggest that the folk concept, like academic treatments, links the validity of consent to the autonomy of the agent. Specifically, across three studies, we found consistent evidence that consent is regarded as less valid when the agent's autonomous decision-making capacities are impaired or lacking.

By contrast, the observed relationship between judgments of valid consent and the exercise of autonomous capacities was inconsistent. While we did not find strong evidence for the claim that there is no relationship between exercising autonomy and judgments of valid consent, participants tended to treat the Mere Capacity agent's consent as equally valid (Study 2) or slightly more valid (Study 1) compared to someone deciding in a fully autonomous manner; this was so even when the Mere Capacity agent was regarded as making their decision irrationally and without regard for their own values, and as doing something that wasn't right for them.

An exception was Study 3, where we found that participants treated a competent agent's consent as slightly less valid when the vignette stated that they failed to decide in an autonomous way. However, this result coincided with doubt that the Mere Capacity agent

in this scenario really did have the capacity to decide autonomously, as revealed by participants' responses to the manipulation-check questions. Furthermore, the agent's possession of autonomous capacities played a stronger role in driving consent judgments: while the Mere Capacity agent was still viewed, on average, as giving valid consent overall, judgments of consent were dramatically reduced when it was stated that the agent lacked these capacities altogether, in which case participants no longer tended to judge that consent was valid.

On balance, then, we found little support for the Exercises Capacity Hypothesis, according to which the validity of consent is generally determined by whether an agent decides in an autonomous (rational, authentic) manner. Although Study 3 suggests that there may be specific circumstances in which the exercise of autonomous decision-making enhances perceptions that consent is valid, our study findings provide stronger support for a moderate Mere Capacity Hypothesis, according to which morally transformative consent normally only depends on whether an agent possesses the capacity to make autonomous decisions, irrespective of whether the decision to consent is in fact made autonomously. Of course, this finding is consistent with the possibility, to be explored in further research, that whether an agent decides in a rational or authentic manner may well affect perceptions of valid consent indirectly, since irrational, inauthentic or otherwise poor decision-making may be a cue that an agent's capacity to decide autonomously is lacking or constrained.

Thus, we find that folk reasoning about valid consent differs from standard treatments in medical ethics and moral philosophy that require, not only that the agent has autonomous capacities, but also that consent decisions be formed in an autonomous way. At the same time, our findings suggest that the folk concept of consent mirrors expert treatments in the sense that the transformation of rights and obligations depends on the provision of consent by an autonomous agent.

5.1. The Role of Autonomous Consent in Moral Cognition

Our findings suggest that consent judgments based on the possession of autonomous capacities come apart from judgments of the moral wrongness of acting on that consent, instead playing a nuanced role in downstream moral reasoning about rights and obligations.

We therefore suggest that making autonomous decisions (i.e., exercising autonomous capacities), on the one hand, and possessing autonomous capacities, on the other, play importantly different roles in moral reasoning about consent transactions.

On the one hand, our studies suggest that whether an agent makes a decision in an autonomous way affects whether their choice is thought likely to be good for them. It is thus relevant to assessing whether others are doing something morally wrong when they act on this decision. For instance, in Study 3, Sam was judged to have done something wrong when he acted on Jessica's consent to sell the bracelet even though he knew she didn't make the decision in a rational, authentic way and would regret it. This finding is consistent with a large literature showing that breaking moral norms, harming others, and knowingly causing harmful consequences affect judgments that a person is doing something morally wrong. More broadly, this result suggests that the exercising autonomous decision-making—while not relevant to participants' assessments of valid consent—is relevant to their morally evaluative appraisals. This result coheres with the dominant framework of research in moral cognition, which focuses on how people generate morally evaluative appraisals of moral agents (especially potential wrongdoers) within agent-victim or agent-beneficiary dyads. This wide literature has explored how

intentions, consequences and norms affect judgments of the rightness or wrongness of actions (Crockett, 2013; Cushman & Greene, 2012; Greene, 2008; Graham et al 2013; Gray, Young & Waytz 2012), the agent's blameworthiness or praiseworthiness (Alicke, 2000; Malle, Guglielmo, & Monroe, 2014), the moral agent's character (Uhlmann, Pizarro & Diermeier, 2015), and the agent's relation to good or bad outcomes (Knobe, 2003).

Our findings underscore that these types of moral appraisals are psychologically distinct from judgments of valid consent. Judgments of the validity of consent came apart from appraisals of whether the consenter and consent-obtainer were acting rightly or wrongly, making good or poor decisions, or acting in praiseworthy or blameworthy ways. Consent judgments had a different function: a core finding of Study 3 was that the provision of autonomous, valid consent crucially affected judgments of what other third parties were and were not permitted to do. Specifically, the character who selfishly got a friend to agree to selling her bracelet was perceived to be doing something morally wrong across all conditions. Nevertheless, variations in the validity of consent determined whether ownership was in fact transferred, whether the buyer was regarded as morally permitted to keep the item, and whether the state was permitted to forcibly stop or reverse the transaction. This finding not only contributes to prior work on reasoning about ownership (Friedman, Neary, Defeyter & Malcolm, 2011), it also illuminates the importance of autonomous capacities and valid consent to moral reasoning about the rights and duties of third parties. This is a moral dimension not captured by existing work on moral appraisals of actions or character within the agent-patient dyad.

We suggest, therefore, that the possession of autonomous capacities, and corresponding judgments of valid consent, may play an important role in whether people's decisions are treated as authoritative. If agents possess autonomous capacities, they have not only the

ability but the right to make their own decisions;[98] this status means that others are required to respect their decisions and refrain from intervening when they consent, even if intervening would be in the consenter's best interest. On the other hand, if an agent lacks autonomous capacities, third parties may regard the consenter's "decision" as something to be ignored, allowing them to paternalistically intervene for the consenter's own good.

Future research might investigate whether this phenomenon extends to domains other than ownership. For instance, imagine that A is immorally trying to get B to have unprotected sex, where this isn't what B wants, and A doesn't care about B's wishes, desires, or welfare. We would expect participants to judge that A is doing something morally wrong. However, a crucial difference may emerge between the case where B has the capacity to make an autonomous decision but fails to exercise these capacities and the case where B lacks these capacities altogether (e.g., because they are cognitively impaired, too young, too intoxicated, or A exerts coercive pressure). In the latter case, we might expect participants to judge that B's consent is not valid, and that it would therefore be permissible, or perhaps obligatory, for third parties to attempt to protect B by physically intervening to prevent them from having sex with A. By contrast, in the case where B is competent but gives consent irrationally, without accounting for the consequences or their personal desires, we might expect participants to judge that third parties must respect B's decision—even if it's the wrong decision, and even if A is immorally taking advantage of

[98] This phenomenon raises interesting questions about possible limits on the kinds of things to which autonomous agents are able to give valid consent, such as whether an autonomous agent is able to make a morally authoritative decision to *give up* their right to make decisions for themselves—as in consensually opting into slavery. Studies by Starmans and Friedman (2016) suggest that agents are not viewed as ownable to the extent that they are autonomous, but that an autonomous agent may be viewed as somewhat "owned" if they consent to being owned (Starmans and Friedman, 2016, Study 4). Future research could examine to what extent, and in what way precisely, such consent is viewed as valid and morally transformative, if at all. We thank an anonymous reviewer for this suggestion.

B. Moreover, we would expect participants to judge that it would be immoral for third parties to forcibly prevent B from having sex with A (although they may attempt to warn B or persuade B to change their mind).

The disjuncture between moral permissibility and valid consent that we documented in Study 3 may illuminate prior research about consent-by-deception which shows that even when someone is intentionally deceived into saying "yes" to things like sex or surgery, participants tend to judge the interaction to be consensual (Sommers, 2020). Although the consenters in Sommers's vignettes were deceived, they retained normal capacities to reason and make decisions. Furthermore, they were not forced to believe the deceiver's testimony or prevented from checking the information they were being given. While further research is needed, it's plausible that deception did not affect participants' perception that the consenter was in full possession of their psychological capacity to make rational and authentic choices—even though the deception meant that the choice they made did not objectively satisfy their desires. If so, our finding that only the capacity to make autonomous decisions is important for valid consent could partly explain participants' judgments that such interactions are consensual. Furthermore, our findings suggest that participants did not necessarily think that deception was morally permissible just because they judged the interaction to be consensual. Instead, it's plausible that participants judged that consent-by-deception was valid even though they also judged that the deceiver's actions were morally wrong, where the former judgment may have affected downstream moral reasoning (e.g., about what kinds of punishment are appropriate for the deceiver).

It may be that consent is regarded as valid only if the consenter is *responsible* for their decision to consent. Just as agents are not regarded as forfeiting rights (i.e. being subject to punishment) when they are not regarded as morally responsible for harmful behavior,

so, perhaps, agents are not regarded as transforming their own rights when they are not responsible for the decision to do so. Previous work has shown that a central component of moral responsibility is the agent's ability to choose, which is determined in part by the possession of cognitive and psychological capacities for rational decision-making (Alicke, 2000; Malle, Guglielmo, & Monroe, 2014; Schlenker et al., 1994; Shaver, 1985; Weiner, 1995). Accordingly, lacking autonomous decision-making capacities reduces judgments that an agent is morally responsible (Blakey & Kremsmayer, 2018; Daigle & Demaree-Cotton, 2021; Monroe, Brady & Malle, 2017, Study 4; Rise & Halkjelsvik 2019). Similar patterns are seen in lay judgments of criminal responsibility (e.g., Allen et al., 2019). Future research could explore the link between judgments of autonomous capacity consent and judgments of responsibility.

Finally, a core finding of the present studies was that the relationship between autonomous decision-making capacities and judgments of valid consent held across a variety of domains—including sexual consent, medical consent, consent to entry and search, and consent to property transfer. This provides novel evidence that people possess a domain-general moral concept of autonomous, valid consent that they employ across very different domains. The apparent domain-generality of reasoning about autonomous, valid consent further supports our contention that consent forms an important but heretofore relatively neglected component of moral cognition.

5.2. Limitations

We used vignettes that explicitly manipulated information about the consenter's autonomy. In real life, however, people often lack access to explicit information about an agent's decision-making capacities. Instead, it's likely that they infer such information from what

is known about the consenter's circumstances, behavior and environment. Relevant cues could include age, mental illness, developmental disability, intoxication, history of trauma or abuse, the consenter's past behaviors, displays of emotion or pain, or social group stereotypes (e.g., Blakey & Kremsmayer, 2018; Vonasch, Baumeister, & Mele, 2018). Future research could further investigate which kinds of real-life circumstances are regarded as autonomy- and consent-undermining. For example, future research could examine whether, and in what circumstances, intoxication is regarded as undermining an agent's capacity for autonomous decision-making and therefore their ability to give valid consent, as opposed to being perceived as a state that merely leads competent agents to make bad decisions, without affecting the validity of their consent.

Relatedly, Study 3, in which an agent was temporarily incapacitated due to medication, yielded much lower ratings of valid consent and autonomy than did Studies 1 and 2, in which the agent's lack of decision-making capacity was unexplained. The lower ratings in Study 3 are likely due to the fact that the incapacity was wholly general (they were unable to engage in any kind of autonomous reasoning or decision-making), and was explained by a concrete condition (heavy medication). In Study 2, the agent in Lacks Capacity lacked a specific component of autonomous decision-making capacity (for example, lacking the ability to control impulses), while in Study 1, the incapacity was generalized but not explained by any concrete condition. In addition, unlike in Studies 1 and 2, the incapacity in Study 3 was temporary, not a permanent condition.

Finally, while the present studies suggest that possessing autonomous decision-making capacities, but not exercising those capacities, is necessary for judgments that consent is valid, we do not expect that this is sufficient. For instance, in addition to the possession of autonomous capacities, valid consent likely requires that "yes" is explicitly or implicitly

communicated, and that the agent's decision is free from external interference such as coercion. Further research is needed to examine other necessary conditions for judgments of valid consent.

5.3. Implications for public health and policy

Our research suggests that consent education and anti-sexual-violence campaigns may be more successful to the extent that the emphasize the way that alcohol, pressure, or manipulation undermine the consenter's capacity to make an autonomous decision rather than focusing on how they impair the quality of decisions (cf. Beres, 2014; Bonnefon, Shariff, & Rahwan, 2020). Secondly, our findings suggest that sex may be regarded as consensual even if one party is wronging the other (Study 3), even if sex is not the right choice for the consenter (Studies 1 and 2), and even if sex is influenced by unwanted impulse, peer pressure, or irrational belief (Study 2). Therefore, policies and educational campaigns may be more successful in targeting harmful and immoral behaviors by not only appealing to "consent" but to additional moral concepts as well (e.g., respect, care, harm; see Carmody, 2005; Carmody & Ovenden, 2013). Finally, our findings may also inform research on victim-blaming (Niemi & Young, 2014, 2016), in that victim-blaming narratives may capitalize on the role of autonomous capacity in valid consent. Emphasizing the victim's capacities to make their own decisions may contribute to the assignment of moral obligations to victims to avoid sexual defilement, judgments that they have the capacity to avoid defilement, and judgments of causal responsibility for sexual assault, all of which contribute to ascriptions of moral responsibility to victims (Niemi & Young, 2014). More than that, however, our findings suggest that emphasis on the consenter's capacities may result in inappropriate judgments that the assault was *consensual*, protecting

the assaulter from third party punishment and interference *even if* they are regarded as morally responsible for wrongdoing.

6. Conclusion

Before these studies, it remained an open possibility that "valid consent" as a rich and normatively complex force existed only as a technical concept used in philosophical, legal and academic domains. We found, however, that the folk concept of consent involves normative distinctions between valid and invalid consent that are sensitive to the consenter's autonomy, even if the linguistic utterance of "yes" is held constant, and that this concept plays an important role in moral reasoning.

Specifically, the studies presented here examined the relationship between autonomy and intuitive judgments of valid consent in several domains: medical procedures, sexual relations, police searches, and agreements between buyers and sellers. Across scenarios, we found that judgments of valid consent carried a specific relationship to autonomy: whether an agent possesses the mental capacity to make decisions in an autonomous way has a consistent impact on whether their consent is regarded as valid, and thus whether it was regarded as morally transformative of the rights and obligations of the consenter and of third parties. Yet, whether the agent *in fact* makes their decision in an autonomous, rational way—based on their own authentic values and what is right for them—has little impact on perceptions of consent or associated rights, although it has relevance for whether the consent-obtainer is acting wrongly. Autonomy thus has a subtle role in the ordinary reasoning about morally transformative consent, where consent given by an agent with autonomous capacities has a distinctive role in downstream moral reasoning.

APPENDIX A: SUPPLEMENTAL ANALYSES

1. Study 1 results, with and without exclusions

Unless noted otherwise, shared superscripts indicate that means do not differ significantly from one another as revealed by fitting a linear mixed model with autonomy condition included as a fixed factor and domain included as a random factor. Significance of fixed effects was assessed via *t*-tests using Satterthwaite's method.

Where an analysis of a dependent variable indicated a significant difference ($p < .05$) only after exclusions but not before, or vice versa, the analysis is shaded grey.

a. N per condition

	Before exclusions	After exclusions
Exercises Capacity	N = 154	N = 146
Mere Capacity	N = 148	N = 91
Lacks Capacity	N = 153	N=127
No condition (e.g., participant did not complete survey)	N = 5	
Total	**N = 450**	**N = 364**
Attention check failure rates differ significantly by condition, $\chi^2(6, N = 445) = 69.65, p < .001$.		

b. Manipulation checks

	N = 450 (before exclusions) Mean (SD)	**N = 364 (after exclusions) Mean (SD)**
Mean <u>rational capacity</u> rating in each of 3 conditions E.g., "Marvin has the ability to make rational decisions." Our pre-registration states that the Lacks Capacity condition will be lower than the other two conditions on this measure.		
Exercises Capacity	6.16 (1.14), *n* = 155[a]	6.29 (0.96), *n* = 146[a]
Mere Capacity	6.03 (1.09), *n* = 149[a]	6.38 (0.85), *n* = 91[a]
Lacks Capacity	2.71 (1.72), *n* = 146[b]	2.35 (1.41), *n* = 127[b]
Mean <u>authentic capacity</u> rating in each of 3 conditions E.g., "Marvin has the ability to be true to himself when making decisions." Our pre-registration states that the Lacks Capacity condition will be lower than the other two conditions on this measure.		
Exercises Capacity	6.19 (1.02), *n* = 155[a]	6.32 (0.77), *n* = 146[a]
Mere Capacity	5.85 (1.12), *n* = 149[b]	6.15 (0.93), *n* = 91[a]
Lacks Capacity	3.36 (1.82), *n* = 146[c]	2.98 (1.56), *n* = 127[b]
Mean <u>rational exercise</u> rating in each of 3 conditions E.g., "Marvin made this particular decision rationally." Our pre-registration states that Exercises Capacity condition will be higher than the other two conditions on this measure.		
Exercises Capacity	5.88 (1.27), *n* = 155[a]	5.99 (1.15), *n* = 146[a]
Mere Capacity	3.48 (1.84), *n* = 149[b]	3.19 (1.81), *n* = 91[b]
Lacks Capacity	2.70 (1.63), *n* = 146[c]	2.35 (1.38), *n* = 127[c]
Mean <u>authentic exercise</u> rating in each of 3 conditions E.g., "When Marvin said 'yes' to having surgery, he was not being true to himself." (reverse-scored) Our pre-registration states that Exercises Capacity condition will be higher than the other two conditions on this measure (once it is reverse-scored).		
Exercises Capacity	5.67 (1.45), *n* = 155[a]	5.81 (1.31), *n* = 146[a]
Mere Capacity	3.95 (1.67), *n* = 149[b]	4.04 (1.68), *n* = 91[b]
Lacks Capacity	3.58 (1.38), *n* = 146[c]	3.69 (1.34), *n* = 127[b]

c. Judgments of valid consent

Our preregistration states that we will test whether condition predicts consent, and if so, will conduct pairwise comparisons comparing Mere Capacity to Exercises Capacity, and Mere Capacity to Lacks Capacity.

	N = 450 (before exclusions)	**N = 364 (after exclusions)**
Cronbach's alpha for three consent items	0.71	0.74
Mean rating of valid consent (composite measure) in each of 3 conditions		
Exercises Capacity	5.87 (1.18), n = 155[a]	5.98 (1.08), n = 146[a]
Mere Capacity	5.91 (1.16), n = 149[a]	6.38 (0.84), n = 91[b]
Lacks Capacity	4.71 (1.39), n = 146[b]	4.78 (1.41), n = 127[c]
Difference between mean consent ratings in each of 3 conditions, adjusting for vignette domain (random factor)		
Exercises Capacity vs. Mere Capacity	b = 0.05, SE = 0.14, t = 0.35, p = .73, CI [-0.23, 0.32]	b = 0.36, SE = 0.15, t = 2.35, p = .019, CI [0.06, 0.65]
Exercises Capacity vs. Lacks Capacity	b = -1.16, SE = 0.14, t = -8.32, p < .001, CI [-1.44, -0.89]	b = -1.21, SE = 0.14, t = -8.81, p < .001, CI [-1.48, -0.94]
Mere Capacity vs. Lacks Capacity	b = -1.21, SE = 0.14, t = -8.58, p < .001, CI [-1.49, -0.93]	b = -1.56, SE = 0.16, t = -10.05, p < .001, CI [-1.87, -1.26]

d. Judgments of right choice: e.g., "Having surgery was probably the right choice for Marvin."/"Having sex with Frank was probably the right choice for Ellen."/"Allowing the police officers to search his home was probably the right choice for Johnny."

Our pre-registration states that we predict that comparison between Mere Capacity and Exercises Capacity to be significant, as well as the comparison between Mere Capacity and Lacks Capacity.

	N = 450 (before exclusions)	**N = 364 (after exclusions)**
Mean <u>right choice</u> rating in each of 3 conditions		
Exercises Capacity	5.03 (1.61), $n = 155^{a}$	5.05 (1.57), $n = 146^{a}$
Mere Capacity	4.19 (1.57), $n = 149^{b}$	4.19 (1.50), $n = 91^{b}$
Lacks Capacity	3.71 (1.78), $n = 146^{c}$	3.47 (1.73), $n = 127^{c}$
Difference between mean <u>right choice</u> ratings in each of 3 conditions, adjusting for vignette domain (random factor)		
Exercises Capacity vs. Mere Capacity	b = -.82, SE = 0.17, t = -4.79, $p < .001$, CI [-1.16, -0.49]	b = -1.01, SE = 0.19, t = -5.43, $p < .001$, CI [-1.38, -0.65]
Exercises Capacity vs. Lacks Capacity	b = -1.31, SE = 0.17, t = -7.58, $p < .001$, CI [-1.65, -0.97]	b = -1.60, SE = 0.17, t = -9.47, $p < .001$, CI [-1.93, -1.26]
Mere Capacity vs. Lacks Capacity	b = -0.49, SE = 0.17, t = -2.79, $p = .005$, CI [-0.83, -0.14]	b = -0.58, SE = 0.19, t = -3.05, $p = .002$, CI [-0.96, -0.21]

e. Judgments of <u>morality</u>: "Under these circumstances, it would be morally wrong for the doctor to proceed with the surgery."/ "Under these circumstances, it would be morally wrong for Frank to have sex with Ellen."/ "Under these circumstances it would be morally wrong for the police officers to enter and search Johnny's home. home."

Our pre-registration states that as a secondary analysis, we will test whether morality judgments exhibit the same pattern as consent.

	N = 450 (before exclusions)	**N = 364 (after exclusions)**
Mean <u>morality</u> rating in each of 3 conditions		
Exercises Capacity	2.44 (1.79), $n = 155^{a}$	2.35 (1.75), $n = 146^{a}$
Mere Capacity	2.75 (1.75), $n = 149^{a}$	2.14 (1.44), $n = 91^{a}$
Lacks Capacity	4.59 (1.86), $n = 146^{b}$	4.54 (1.89), $n = 127^{b}$
Difference between mean <u>morality</u> ratings in each of 3 conditions, adjusting for vignette domain (random factor)		
Exercises Capacity vs. Mere Capacity	$b = .31$, $SE = 0.20$, $t = 1.57$, $p = 0.12$, CI [-0.08, 0.07]	$b = -0.09$, $SE = 0.22$, $t = -0.42$, $p = .68$, CI [-0.53, 0.34]
Exercises Capacity vs. Lacks Capacity	$b = 2.17$, $SE = 0.20$, $t = 10.88$, $p < .001$, CI [1.78, 2.56]	$b = 2.20$, $SE = 0.20$, $t = 10.99$, $p < .001$, CI [1.81, 2.60]
Mere Capacity vs. Lacks Capacity	$b = 1.86$, $SE = 0.20$, $t = 9.23$, $p < .001$, CI [1.46, 2.25]	$b = 2.30$, $SE = 0.23$, $t = 10.09$, $p < .001$, CI [1.85, 2.75]

f. Judgments of <u>freedom</u>: "Marvin freely chose to have surgery"/"Ellen freely chose to have sex with Frank"/"Johnny freely chose to allow the police officers to enter and search his home."

Our pre-registration states that as a secondary analysis, we will test whether freedom judgments exhibit the same pattern as consent.

	N = 450 (before exclusions)	**N = 364 (after exclusions)**
Mean <u>freedom</u> rating in each of 3 conditions		
Exercises Capacity	6.32 (0.95), $n = 155^{a}$	6.46 (0.68), $n = 146^{a}$
Mere Capacity	6.28 (0.85), $n = 149^{a}$	6.44 (0.73), $n = 91^{a}$
Lacks Capacity	4.80 (1.63), $n = 146^{b}$	4.68 (1.67), $n = 127^{b}$
Difference between mean <u>freedom</u> ratings in each of 3 conditions, adjusting for vignette domain (random factor)		
Exercises Capacity vs. Mere Capacity	$b = -.04$, $SE = 0.14$, $t = -0.30$, $p = .77$, CI [-0.31, 0.23]	$b = -0.05$, $SE = 0.15$, $t = -0.31$, $p = .75$, CI [-0.34, 0.25]
Exercises Capacity vs. Lacks Capacity	$b = -1.52$, $SE = 0.14$, $t = -11.15$, $p < .001$, CI [-1.79, -1.25]	$b = 1.79$, $SE = 0.14$, $t = -13.13$, $p < .001$, CI [-2.05, -1.52]
Mere Capacity vs. Lacks Capacity	$b = -1.48$, $SE = 0.14$, $t = -10.75$, $p < .001$, CI [-1.75, -1.21]	$b = -1.74$, $SE = 0.15$, $t = -11.28$, $p < .001$, CI [-2.04, -1.44]

2. Study 2 results, with and without exclusions

Unless noted otherwise, shared superscripts indicate that means do not differ significantly from one another as revealed by fitting a linear mixed model with autonomy condition included as a fixed factor and domain included as a random factor. Significance of fixed effects was assessed via *t*-tests using Satterthwaite's method.

Where an analysis of a dependent variable indicated a significant difference ($p < .05$) only after exclusions but not before, or vice versa, the analysis is shaded grey.

a. N per condition

	Before exclusions	After exclusions
Exercises Capacity	N = 212	N = 157
Mere Capacity	N = 211	N = 95
Lacks Capacity	N = 212	N =132
No condition (e.g., participant did not complete survey)	N = 4	
Total	**N = 639**	**N = 384**
Attention check failure rates differ significantly by condition, $\chi^2(2, N = 631) = 38.32, p < .001$.		

b. Manipulation checks

	N = 639 (before exclusions) Mean (SD)	**N = 384 (after exclusions) Mean (SD)**
Mean <u>rational capacity</u> rating in each of 3 conditions "Marvin has the ability to make rational decisions." Our pre-registration states that Lacks Capacity condition will be lower than the other two conditions on this measure.		
Exercises Capacity	6.21 (0.98), $n = 205^{a}$	6.46 (0.73), $n = 157^{a}$
Mere Capacity	5.83 (1.17), $n = 206^{b}$	6.09 (0.91), $n = 95^{b}$
Lacks Capacity	3.77 (1.91), $n = 212^{c}$	2.91 (1.58), $n = 132^{c}$
Mean <u>authentic capacity</u> rating in each of 3 conditions "Marvin has the ability to be true to himself when making decisions." Our pre-registration states that Lacks Capacity condition will be lower than the other two conditions on this measure.		
Exercises Capacity	6.00 (1.09), $n = 208^{a}$	6.30 (0.80), $n = 157^{a}$
Mere Capacity	5.85 (1.14), $n = 208^{a}$	6.03 (0.95), $n = 95^{a}$
Lacks Capacity	4.02 (1.73), $n = 212^{b}$	3.29 (1.56), $n = 132^{b}$
Mean <u>general autonomous capacity</u> rating in each of 3 conditions "Marvin has the ability to shape his own life freely according to his own values and what is right for him." Our pre-registration states that Lacks Capacity condition will be lower than the other two conditions on this measure.		
Exercises Capacity	6.16 (1.07), $n = 207^{a}$	6.47 (0.71), $n = 157^{a}$
Mere Capacity	5.97 (0.95), $n = 207^{a}$	6.06 (0.88), $n = 95^{b}$
Lacks Capacity	4.16 (1.79), $n = 212^{b}$	3.45 (1.65), $n = 132^{c}$
Mean <u>rational exercise</u> rating in each of 3 conditions "Marvin made this particular decision rationally." Our pre-registration states that Exercises Capacity condition will be higher than the other two conditions on this measure.		
Exercises Capacity	6.09 (1.16), $n = 208^{a}$	6.41 (0.85), $n = 156^{a}$
Mere Capacity	3.91 (1.94), $n = 207^{b}$	2.92 (1.60), $n = 95^{b}$
Lacks Capacity	3.43 (2.03), $n = 212^{c}$	2.39 (1.50), $n = 132^{c}$
Mean <u>authentic exercise</u> rating in each of 3 conditions "When Marvin said 'yes' to having surgery, he was not being true to himself." (reverse-scored) Our pre-registration states that Exercises Capacity condition will be higher than the other two conditions on this measure (once it is reverse-scored).		
Exercises Capacity	5.04 (2.00), $n = 209^{a}$	5.69 (1.66), $n = 157^{a}$
Mere Capacity	2.89 (1.56), $n = 207^{b}$	3.11 (1.61), $n = 95^{b}$
Lacks Capacity	2.97 (1.51), $n = 212^{b}$	3.05 (1.60), $n = 132^{b}$

Mean general autonomous exercise rating in each of 3 conditions "The way Marvin made this particular decision expressed an ability to shape his life freely according to his own values and what is right for him." Our pre-registration states that Exercises Capacity condition will be higher than the other two conditions on this measure.		
Exercises Capacity	6.08 (1.05), $n = 209^{a}$	6.32 (0.82), $n = 157^{a}$
Mere Capacity	4.85 (1.70), $n = 207^{b}$	4.43 (1.80), $n = 95^{b}$
Lacks Capacity	4.00 (1.90), $n = 212^{c}$	3.19 (1.76), $n = 132^{c}$

c. Judgments of valid consent (composite measure)

Our preregistration states that we expected consent judgments to be higher in Mere Capacity than Lacks Capacity, and to find no significant difference between Exercises Capacity and Mere Capacity.

	N = 639 (before exclusions)	**N = 384 (after exclusions)**
Cronbach's alpha for three consent items	0.75	0.73
Mean rating of valid consent (composite measure) in each of 3 conditions		
Exercises Capacity	5.67 (1.49), $n = 210^{a}$	6.15 (1.23), $n = 157^{a}$
Mere Capacity	5.18 (1.55), $n = 207^{b}$	5.93 (1.17), $n = 95^{a}$
Lacks Capacity	4.74 (1.37), $n = 212^{c}$	5.21 (1.28), $n = 132^{b}$
Difference between mean consent ratings in each of 3 conditions, adjusting for failure type (random factor)		
Exercises Capacity vs. Mere Capacity	b = -0.49, SE = 0.14, t = -3.41, $p < .001$, CI [-0.77, -0.21]	b = -0.22, SE = 0.16, t = -1.38, $p = .167$, CI [-0.54, -0.09]
Exercises Capacity vs. Lacks Capacity	b = -0.93, SE = 0.14, t = -6.48, $p < .001$, CI [-1.21, -0.65]	b = -0.94, SE = 0.15, t = -6.42, $p < .001$, CI [-1.22, -0.65]
Mere Capacity vs. Lacks Capacity	b = -0.44, SE = 0.14, t = -3.04, $p = .002$, CI [-0.72, -0.15]	b = -0.71 SE = 0.16, t = -4.30, $p < .001$, CI [-1.04, -0.39]

d. Judgments of morality: "Under these circumstances, it would be morally wrong for the doctor to proceed with the surgery."

Our pre-registration states that as a secondary analysis, we will test whether morality judgments exhibit the same pattern as consent.

	N = 639 (before exclusions)	**N = 384 (after exclusions)**
Mean <u>morality</u> rating in each of 3 conditions		
Exercises Capacity	5.38 (2.00), $n = 208^{a}$	6.06 (1.58), $n = 157^{a}$
Mere Capacity	4.38 (2.03), $n = 208^{b}$	5.27 (1.67), $n = 95^{b}$
Lacks Capacity	3.67 (1.79), $n = 212^{c}$	4.02 (1.75), $n = 132^{c}$
Difference between mean <u>morality</u> ratings in each of 3 conditions, adjusting for failure type (random factor)		
Exercises Capacity vs. Mere Capacity	b = -0.99, SE = 0.19, t = -5.22, $p < .001$, CI [-1.36, -0.62]	b = -0.77, SE = 0.21, t = -3.56, $p < .001$, CI [-1.19, -0.34]
Exercises Capacity vs. Lacks Capacity	b = -1.70, SE = 0.19, t = -9.01, $p < .001$, CI [-2.07, -1.33]	b = -2.02, SE = 0.20, t = -10.34, $p < .001$, CI [-2.40, -1.64]
Mere Capacity vs. Lacks Capacity	b = -0.71, SE = 0.19, t = -3.76, $p < .001$, CI [-1.08, -0.34]	b = -1.26, SE = 0.22, t = -5.64, $p < .001$, CI [-1.69, -0.82]

e. Judgments of <u>freedom</u>: "Marvin freely chose to have surgery."

Our pre-registration states that as a secondary analysis, we will test whether freedom judgments exhibit the same pattern as consent.

	N = 639 (before exclusions)	**N = 384 (after exclusions)**
Mean <u>freedom</u> rating in each of 3 conditions		
Exercises Capacity	6.20 (1.18), $n = 208^{a}$	6.53 (0.84), $n = 157^{a}$
Mere Capacity	5.74 (1.21), $n = 207^{b}$	5.87 (1.31), $n = 95^{a}$
Lacks Capacity	4.92 (1.67), $n = 212^{c}$	4.67 (1.75), $n = 132^{b}$
Difference between mean <u>freedom</u> ratings in each of 3 conditions, adjusting for failure type (random factor)		
Exercises Capacity vs. Mere Capacity	b = -0.45, SE = 0.13, t = -3.45, $p < .001$, CI [-0.71, -0.20]	b = -0.69, SE = 0.16, t = -4.22, $p < .001$, CI [-1.01, -0.37]
Exercises Capacity vs. Lacks Capacity	b = -1.28, SE = 0.13, t = -9.85, $p < .001$, CI [-1.54, -1.03]	b = -1.88, SE = 0.15, t = -12.70, $p < .001$, CI [-2.18, -1.59]
Mere Capacity vs. Lacks Capacity	b = -0.83, SE = 0.13, t = -6.37, $p < .001$, CI [-1.09, -0.58]	b = -1.19, SE = 0.17, t = -7.08, $p < .001$, CI [-1.53, -0.87]

3. Study 3 results with and without exclusions

Unless noted otherwise, shared superscripts indicate that means do not differ significantly from one another as revealed by fitting a linear model with autonomy condition as a fixed factor. Significance of fixed effects was assessed via *t*-tests using Satterthwaite's method.

Where an analysis of a dependent variable indicated a significant difference ($p < .05$) only after exclusions but not before, or vice versa, the analysis is shaded grey.

a. N per condition

	Before exclusions	After exclusions
Exercises Capacity	N = 101	N = 43
Mere Capacity	N = 100	N = 49
Lacks Capacity	N = 102	N = 61
Total	**N = 303**	**N = 153**
Attention check failure rates differ significantly by condition, $\chi^2(2, N = 303) = 11.35, p = .003$.		

b. Manipulation checks

	N = 303 (before exclusions) Mean (SD)	**N = 153 (after exclusions) Mean (SD)**
Mean <u>rational capacity</u> rating in each of 3 conditions "When Sam asked whether he could sell the bracelet, Jessica had the ability to make a rational decision."		
Exercises Capacity	5.58 (1.25), $n = 101^{a}$	5.88 (1.12), $n = 43^{a}$
Mere Capacity	5.02 (1.56), $n = 100^{b}$	4.96 (1.73), $n = 49^{b}$
Lacks Capacity	3.49 (2.28), $n = 102^{c}$	2.11 (1.60), $n = 61^{c}$
Mean <u>authentic capacity</u> rating in each of 3 conditions "Jessica had the ability to be true to herself when making a decision about whether to let Sam sell the bracelet."		
Exercises Capacity	5.49 (1.30), $n = 101^{a}$	5.74 (1.18), $n = 43^{a}$
Mere Capacity	5.02 (1.64), $n = 100^{a}$	4.80 (1.86), $n = 49^{b}$
Lacks Capacity	3.63 (2.27), $n = 102^{b}$	2.25 (1.70), $n = 61^{c}$
Mean <u>rational exercise</u> rating in each of 3 conditions "Jessica made the decision to say 'yes' rationally."		
Exercises Capacity	5.34 (1.45), $n = 101^{a}$	5.84 (1.17), $n = 43^{a}$
Mere Capacity	4.48 (1.75), $n = 100^{b}$	3.88 (1.79), $n = 49^{b}$
Lacks Capacity	3.77 (2.32), $n = 102^{c}$	2.41 (1.85), $n = 61^{c}$
Mean <u>authentic exercise</u> rating in each of 3 conditions "When Jessica said 'yes' to Sam's selling the bracelet, she was not being true to herself." (reverse-scored)		
Exercises Capacity	3.65 (1.79), $n = 101^{a}$	4.86 (1.68), $n = 43^{a}$
Mere Capacity	2.84 (1.33), $n = 100^{b}$	3.12 (1.49), $n = 49^{b}$
Lacks Capacity	2.11 (1.07), $n = 102^{c}$	1.93 (0.98), $n = 61^{c}$

c. Judgments of valid consent

Our pre-registration predicted that the Lacks Capacity condition will yield lower consent ratings than the Mere Capacity condition.

	N = 303 (before exclusions) Mean (SD)	**N = 153 (after exclusions) Mean (SD)**
Cronbach's alpha for three consent items	0.76	0.89
Mean rating of <u>valid consent</u> (composite measure) in each of 3 conditions		
Exercises Capacity	4.60 (1.51), $n = 101^{a}$	5.84 (1.27), $n = 43^{a}$
Mere Capacity	4.25 (1.34), $n = 100^{a}$	4.80 (1.56), $n = 49^{b}$
Lacks Capacity	3.02 (1.27), $n = 102^{b}$	2.75 (1.53), $n = 61^{c}$
Omnibus F Test examining whether mean <u>consent</u> ratings differ by condition		
	$F(2, 300) = 37.29$, $p < .001$	$F(2, 150) = 59.82$, $p < .001$
Difference between mean <u>consent</u> ratings in each of 3 conditions		
Exercises Capacity vs. Mere Capacity	$b = 0.34$, $SE = 0.19$, $t = 1.82$, $p = .069$, CI [-0.74, -0.03]	$b = 1.03$, $SE = 0.31$, $t = 3.36$, $p = .001$, CI [-1.64, -0.43]
Exercises Capacity vs. Lacks Capacity	$b = 1.59$, $SE = 0.19$, $t = 8.22$, $p < .001$, CI [-1.97, -1.21]	$b = 3.08$, $SE = 0.29$, $t = 10.51$, $p < .001$, CI [-3.66, -2.50]
Mere Capacity vs. Lacks Capacity	$b = 1.23$, $SE = 0.19$, $t = 6.37$, $p < .001$, CI [-1.61, -0.85]	$b = 2.05$, $SE = 0.28$, $t = 7.25$, $p < .001$, CI [-2.61, -1.49]

d. Judgments of (1) <u>ownership transfer</u>: composite continuous measure and (2) <u>binary ownership measure</u> ("Who is the rightful owner of the bracelet?" Jessica/Melanie)

For the continuous measure, our pre-registration states that we predict the Lacks Capacity condition will yield lower ratings of ownership transfer than the Mere Capacity condition. For the binary measure, the pre-registration states that we will conduct Fisher's exact tests analyzing whether the percentage of participants answering "Melanie" differs between the autonomy conditions. We predicted lower rates in the Lacks Capacity condition.

	N = 303 (before exclusions) Mean (SD)	**N = 153 (after exclusions) Mean (SD)**
Cronbach's alpha for five ownership items	0.83	0.89
Mean rating of <u>ownership transfer</u> (composite continuous measure) in each of 3 conditions		
Exercises Capacity	4.29 (1.54), $n = 101^a$	5.53 (1.36), $n = 43^a$
Mere Capacity	4.00 (1.44), $n = 100^a$	4.71 (1.58), $n = 49^b$
Lacks Capacity	3.11 (1.14), $n = 102^b$	3.18 (1.40), $n = 61^c$
Omnibus F Test examining whether mean <u>ownership transfer</u> ratings differ by condition		
	$F(2, 300) = 19.90$, $p < .001$	$F(2, 150) = 35.60$, $p < .001$
Difference between mean <u>ownership transfer</u> ratings in each of 3 conditions		
Exercises Capacity vs. Mere Capacity	$b = 0.29$ $SE = 0.20$, $t = 1.46$, $p = .15$, CI [-0.67, 0.10]	$b = 0.82$, $SE = 0.31$, $t = 2.69$, $p = .008$, CI [-1.41, -0.22]
Exercises Capacity vs. Lacks Capacity	$b = 1.18$, $SE = 0.19$, $t = 6.05$, $p < .001$, CI [-1.56, -0.79]	$b = 2.35$, $SE = 0.29$, $t = 8.14$, $p < .001$, CI [-2.92, -1.78]
Mere Capacity vs. Lacks Capacity	$b = 0.98$, $SE = 0.20$, $t = 4.57$, $p < .001$, CI [-1.27, -0.51]	$b = 1.53$, $SE = 0.28$, $t = 5.51$, $p < .001$, CI [-2.08, -0.98]
Percentage concluding that ownership transferred to the recipient Melanie (binary measure) in each of 3 conditions		
Exercises Capacity	46.53%, $n = 101^a$	81.40%, $n = 43^a$
Mere Capacity	40.00%, $n = 100^a$	57.14%, $n = 49^b$
Lacks Capacity	20.59%, $n = 102^b$	21.31%, $n = 61^c$
Fisher's exact tests comparing rates at which participants concluded ownership had transferred to Melanie across conditions. We report chi-square test statistics as well.		
Exercises Capacity vs. Mere Capacity	OR = 0.77, $p = .39$, CI [.42, 1.39] $X^2(1) = .628$, $p = .43$	OR = 0.31, $p = .014$, CI [0.10, 0.86] $X^2(1) = 5.17$, $p = .023$
Exercises Capacity vs. Lacks Capacity	OR = 0.30, $p < .001$, CI [0.15, 0.58]; $X^2(1) = 14.19$, $p < .001$	OR = 0.06, $p < .001$, CI [0.02, 0.18]; $X^2(1) = 34.26$, $p < .001$
Mere Capacity vs. Lacks Capacity	OR = 0.39, $p = .003$, CI [0.19, 0.76]; $X^2(1) = 8.13$, $p = .004$	OR = 0.21, $p < .001$, CI [0.08, 0.50]; $X^2(1) = 13.43$, $p < .001$
Linear regression predicting <u>ownership transfer</u> judgments from <u>consent</u> judgments, adjusting for condition. The relationship between ownership transfer and consent judgments varied by condition; thus we interact consent and condition.		
	$b = -0.83$, $SE = .07$, $t = -11.68$, $p < .001$, CI [-0.97, -0.69]	$b = -0.80$, $SE = .10$, $t = -8.16$, $p < .001$, CI [-0.99, -0.61]

e. Judgments of morality: "Under these circumstances, it was morally wrong for Sam to proceed with selling the bracelet." (reverse-scored)

Our pre-registration stated that as a secondary analysis we would examine the effect of condition on morality judgments.

	N = 303 (before exclusions) Mean (SD)	**N = 153 (after exclusions) Mean (SD)**
Mean morality rating in each of 3 conditions		
Exercises Capacity	3.16 (1.82), $n = 101$[a]	4.14 (2.01), $n = 43$[a]
Mere Capacity	2.44 (1.47), $n = 100$[b]	2.43 (1.62), $n = 49$[b]
Lacks Capacity	1.84 (0.96), $n = 102$[c]	1.61 (0.94), $n = 61$[c]
Omnibus F Test examining whether mean morality ratings differ by condition		
	$F(2, 300) = 20.71$, $p < .001$	$F(2, 150) = 35.27$, $p < .001$
Difference between mean morality ratings in each of 3 conditions		
Exercises Capacity vs. Mere Capacity	$b = -0.72$, $SE = 0.21$, $t = -3.49$, $p < .001$, CI [-1.12, -0.31]	$b = -1.71$, $SE = 0.32$, $t = -5.38$, $p < .001$, CI [-2.34, -1.08]
Exercises Capacity vs. Lacks Capacity	$b = -1.32$, $SE = 0.20$, $t = -6.43$, $p < .001$, CI [-1.72, -0.91]	$b = -2.53$, $SE = 0.30$, $t = -8.36$, $p < .001$, CI [-3.13, -1.93]
Mere Capacity vs. Lacks Capacity	$b = -0.60$, $SE = 0.21$, $t = -2.91$, $p = .003$, CI [-1.00, -0.19]	$b = -0.82$, $SE = 0.29$, $t = -2.81$, $p = .006$, CI [-1.40, -0.24]

f. Judgments of right choice: "Saying 'yes' to selling the bracelet was probably the right choice for Jessica at the time."

Our pre-registration stated that as a secondary analysis we would examine the effect of condition on judgments of right choice.

	N = 303 (before exclusions) Mean (SD)	**N = 153 (after exclusions) Mean (SD)**
Mean <u>right choice</u> rating in each of 3 conditions		
Exercises Capacity	4.71 (1.63), n = 101[a]	4.23 (1.66), n = 43[a]
Mere Capacity	3.95 (1.84), n = 100[b]	3.20 (1.68), n = 49[b]
Lacks Capacity	3.79 (2.28), n = 102[b]	2.39 (1.76), n = 61[c]
Difference between mean <u>right choice</u> ratings in each of 3 conditions		
Exercises Capacity vs. Mere Capacity	b = -0.76, SE = 0.27, t = -2.79, p = .006, CI [-1.30, -0.22]	b = -1.03, SE = 0.36, t = -2.88, p = .004, CI [-1.73, -0.32]
Exercises Capacity vs. Lacks Capacity	b = -0.92, SE = 0.27, t = -3.38, p < .001, CI [-1.45, -0.38]	b = -1.84, SE = 0.34, t = -5.40, p < .001, CI [-2.51, -1.17]
Mere Capacity vs. Lacks Capacity	b = -0.16, SE = 0.27, t = -0.57, p = .57, CI [-0.69, -0.38]	b = -0.81, SE = 0.33, t = -2.47, p = .015, CI [-1.46, -0.16]

g. Judgments of <u>freedom</u>: "Jessica freely chose to sell her bracelet."

Our pre-registration stated that as a secondary analysis we would examine the effect of condition on freedom judgments.

	N = 303 (before exclusions) Mean (SD)	**N = 153 (after exclusions) Mean (SD)**
Mean <u>freedom</u> rating in each of 3 conditions		
Exercises Capacity	5.15 (1.53), n = 101[a]	5.58 (1.42), n = 43[a]
Mere Capacity	4.79 (1.74), n = 100[a]	4.67 (1.84), n = 49[b]
Lacks Capacity	3.69 (2.37), n = 102[b]	2.30 (1.74), n = 61[c]
Difference between mean <u>freedom</u> ratings in each of 3 conditions		
Exercises Capacity vs. Mere Capacity	B = -0.36, SE = 0.27, t = -1.33, p = .19, CI [-0.89, -0.17]	b = -0.91, SE = 0.35, t = -2.57, p = .011, CI [-1.61, -0.21]
Exercises Capacity vs. Lacks Capacity	b = -1.46, SE = 0.27, t = -5.44, p < .001, CI [-1.99, -0.93]	b = -3.29, SE = 0.34, t = -9.77, p < .001, CI [-3.95, -2.62]
Mere Capacity vs. Lacks Capacity	b = -1.10, SE = 0.27, t = -4.09, p < .001, CI [-1.63, -0.57]	b = -2.38, SE = 0.32, t = -7.34, p < .001, CI [-3.02, -1.74]

APPENDIX B: STIMULI

1. Study 1 Stimuli

a. Study 1 vignettes by condition

MEDICAL		
Marvin has been in physical therapy for ankle pain. One day his doctor asks him whether he wishes to undergo elective surgery to repair the tendon. The doctor explains that the surgery carries some risks, as all surgeries do, but if all goes well it could potentially completely cure his ankle pain.		
Exercises Capacity	Mere Capacity	Lacks Capacity
Marvin is an intelligent, able adult. He is perfectly capable of weighing up pros and cons; thinking through the choice he faces; and making decisions based on what is best for him, which options align with his personal values, and what he really wants. And he does so in this instance. After thinking things through very carefully--and with careful regard for the pros and cons, and whether it aligns with his personal values and what he really wants--Marvin says 'yes' to the surgery.	Marvin is an intelligent, able adult. He is perfectly capable of weighing up pros and cons; thinking through the choice he faces; and making decisions based on what is best for him, which options align with his personal values, and what he really wants. But he does not do so in this instance. Without thinking things through even a little bit--and with absolutely no regard for the pros and cons, or whether it aligns with his personal values and what he really wants--Marvin says 'yes' to the surgery.	Marvin is not able and intelligent like most adults. He is completely incapable of weighing up pros and cons; thinking through the choice he faces; or making decisions based on what is best for him, which options align with his personal values, or what he really wants. So he does not do so in this instance. Without thinking things through even a little bit--and with absolutely no regard for the pros and cons, or whether it aligns with his personal values and what he really wants--Marvin says 'yes' to the surgery.

SEXUAL		
Ellen and Frank meet in a night class. They go on several dates. At the end of their last date, Frank asks Ellen to have sex with him. He adds, "I know we haven't known each other very long, but I've been enjoying our time together, and this feels right to me."		
Exercises Capacity	Mere Capacity	Lacks Capacity
Ellen is an intelligent, able adult. She is perfectly capable of weighing up pros and cons; thinking through the choice she faces; and making decisions based on what is best for her, which options align with her personal values, and what she really wants. And she does so in this instance. After thinking things through very carefully--and with careful regard for the pros and cons, and whether it aligns with her personal values and what she really wants--Ellen says 'yes' to having sex with Frank.	Ellen is an intelligent, able adult. She is perfectly capable of weighing up pros and cons; thinking through the choice she faces; and making decisions based on what is best for her, which options align with her personal values, and what she really wants. But she does not do so in this instance. Without thinking things through even a little bit--and with absolutely no regard for the pros and cons, or whether it aligns with her personal values and what she really wants--Ellen says 'yes' to having sex with Frank.	Ellen is not able and intelligent like most adults. She is completely incapable of weighing up pros and cons; thinking through the choice she faces; or making decisions based on what is best for her, which options align with her personal values, or what she really wants. So she does not do so in this instance. Without thinking things through even a little bit--and with absolutely no regard for the pros and cons, or whether it aligns with her personal values and what she really wants--Ellen says 'yes' to having sex with Frank.

POLICE ENTRY		
Johnny hears someone knock on his apartment door. Two men are standing outside. They say, "Police here. Can we come in?" Johnny asks through the door, "What for?" One of the police officers says, "We are looking for drugs and drug paraphernalia. We got an anonymous call reporting drug dealing in this building. So can we come in?"		
Exercises Capacity	Mere Capacity	Lacks Capacity
Johnny is an intelligent, able adult. He is perfectly capable of weighing up pros and cons; thinking through the choice he faces; and making decisions based on what is best for him, which options align with his personal values, and what he really wants. And he does so in this instance. After thinking things through very carefully--and with careful regard for the pros and cons, and whether it aligns with his personal values and what he really wants--Johnny says 'yes' to letting the police search his apartment.	Johnny is an intelligent, able adult. He is perfectly capable of weighing up pros and cons; thinking through the choice he faces; and making decisions based on what is best for him, which options align with his personal values, and what he really wants. But he does not do so in this instance. Without thinking things through even a little bit--and with absolutely no regard for the pros and cons, or whether it aligns with his personal values and what he really wants--Johnny says 'yes' to letting the police search his apartment.	Johnny is not able and intelligent like most adults. He is completely incapable of weighing up pros and cons; thinking through the choice he faces; or making decisions based on what is best for him, which options align with his personal values, or what he really wants. So he does not do so in this instance. Without thinking things through even a little bit--and with absolutely no regard for the pros and cons, or whether it aligns with his personal values and what he really wants--Johnny says 'yes' to letting the police search his apartment.

b. Study 1 measures

"X" and "Y" replaced to refer to the relevant agents and "z" replaced with the relevant action, according to vignette viewed by the participant.

Consent

Consent 1: X had Y's permission to proceed with z.

Consent 2: If X proceeds with z-ing now, he'll be acting without Y's consent.

Consent 3: Y's 'yes' didn't count as consent.

Morality: Under these circumstances, it would be morally wrong for Y to proceed with z-ing.

Freedom: Y freely chose to z.

Right Choice: Z-ing was probably the right choice for X.

Manipulation checks

Rational capacity: Y has the ability to make rational decisions.

Rational exercise: Y made this particular decision rationally.

Authentic capacity: Y has the ability to be true to him[her]self when making decisions.

Authentic exercise: When Y said 'yes' to z-ing, [s]he was not being true to him[her]self.

c. Study 1 attention checks

Which is correct?

a. Y **did** think through the pros and cons of z-ing.

b. Y **did not** think through the pros and cons of z-ing.

Y made this particular decision . . .

a. **with** regard for whether z-ing aligned with his[her] personal values.

b. **without** regard for whether z-ing aligned with his[her] personal values.

At the time X suggested z-ing, Y was . . .

a. **capable** of thinking through his[her] choice and deciding based on the pros and cons

b. **incapable** of thinking through his[her] choice and deciding based on the pros and cons

At the time X suggested z-ing, Y was . . .

a. **able** to make a decision based on which option aligned with his[her] personal values.

b. **unable** to make a decision based on which option aligned with his[her] personal values.

2. Study 2 Stimuli

a. Study 2 vignettes by condition

All vignettes begin:

> Please read the following story carefully.
>
> Marvin has been in physical therapy for ankle pain. One day his doctor asks him whether he wishes to undergo elective surgery to repair the tendon. The doctor explains that the surgery carries some risks, as all surgeries do, but if all goes well it could potentially completely cure his ankle pain.

The vignettes continue according to autonomy and failure type condition:

IMPULSE		
Exercises Capacity	Mere Capacity	Lacks Capacity
Marvin feels an initial impulse to simply say 'no' surgery. Marvin is an intelligent, able adult, fully capable of making decisions for himself and controlling impulses when they are inappropriate. And he does so in this instance. Although he feels an initial impulse to avoid surgery, he thinks things through carefully, and makes his decision with careful regard for the pros and cons, and whether surgery aligns with his personal values and what he really wants. Because of this, Marvin says 'yes' to the surgery. If he had not resisted his initial impulse and made a decision based on thinking things through properly, Marvin would have said 'no', despite surgery being the right choice for him.	Marvin feels an initial impulse to simply say 'yes' to surgery. Marvin is an intelligent, able adult, fully capable of making decisions for himself and controlling impulses when they are inappropriate. But he does not do so in this instance. Acting on an initial impulse to have the surgery, he doesn't think things through even a little bit, and pays absolutely no attention to the pros and cons, or whether surgery aligns with his personal values and what he really wants. He simply says 'yes' to the surgery on an impulse. If he had resisted his initial impulse and made a decision based on thinking things through properly, Marvin would have said 'no', as surgery is not the right choice for him.	Marvin feels an initial impulse to simply say 'yes' to surgery. Marvin is not able and intelligent like most adults who are fully capable of making decisions for themselves: he is completely incapable of controlling impulses, even when they are inappropriate. So he does not do so in this instance. Acting on an initial impulse to have the surgery, he doesn't think things through even a little bit, and pays absolutely no attention to the pros and cons, or whether surgery aligns with his personal values and what he really wants. He simply says 'yes' to the surgery on an impulse. If he had been able to resist his initial impulse and make a decision based on thinking things through properly, Marvin would have said 'no', as surgery is not the right choice for him.

PEER PRESSURE		
Exercises Capacity	Mere Capacity	Lacks Capacity
Marvin's friends are really pushy and opinionated when it comes to medical matters, and they think that it would be really irresponsible for Marvin to have the surgery. Marvin is an intelligent, able adult, fully capable of making decisions for himself: even if he is under pressure from others to make a certain choice, he is still perfectly able to make his own decision and resist pressure to do things that he doesn't want to do or that aren't right for him. And he does so in this instance. Although he feels pressure from his friends to say 'no' to the surgery, he thinks things through carefully, and makes his decision with careful regard for the pros and cons, and whether surgery aligns with his personal values and what he really wants. Because of this, Marvin says 'yes' to the surgery. If he had not resisted the pressure from his friends and thought things through for himself, Marvin would have said 'no' even though surgery is what he really wants.	Marvin's friends are really pushy and opinionated when it comes to medical matters, and they think that it would be really irresponsible for Marvin not to have the surgery. Marvin is an intelligent, able adult, fully capable of making decisions for himself: even if he is under pressure from others to make a certain choice, he is still perfectly able to make his own decision and resist pressure to do things that he doesn't want to do or that aren't right for him. But he does not do so in this instance. Giving in to pressure from his friends to have the surgery, he doesn't think things through even a little bit, and pays absolutely no attention the pros and cons, or whether surgery aligns with his personal values and what he really wants. Marvin simply says 'yes' to the surgery. If he had resisted the pressure from his friends and thought things through for himself, Marvin would have said 'no', as surgery is not what he really wants.	Marvin's friends are really pushy and opinionated when it comes to medical matters, and they think that it would be really irresponsible for Marvin not to have the surgery. Marvin is not able and intelligent like most adults who are able to make decisions for themselves: if he is under pressure from others to make a certain choice, he is completely incapable of making his own decision and resisting pressure to do things that he doesn't want to do or that aren't right for him. So he does not do so in this instance. Giving in to pressure from his friends to have the surgery, he doesn't think things through even a little bit, and pays absolutely no attention the pros and cons, or whether surgery aligns with his personal values and what he really wants. Marvin simply says 'yes' to the surgery. If he had been able to resist pressure from his friends and think things through for himself, Marvin would have said 'no', as surgery is not what he really wants.

UNINFORMED		
Exercises Capacity	Mere Capacity	Lacks Capacity
While the doctor is explaining some of the critical risks and benefits of the procedure, however, Marvin becomes distracted by a text message, and doesn't hear what the doctor is saying. Marvin is an intelligent, able adult: he can tell that he doesn't know enough about the surgery to make an informed choice, so he is perfectly capable of making sure he gets the information he needs. And he does so in this instance. He tells the doctor he was distracted, and the doctor repeats the relevant information. Now that he has substantial information about the procedure, Marvin makes his decision with careful regard for the pros and cons, and whether surgery aligns with his personal values and what he really wants. Because of this, Marvin says 'yes' to the surgery. If he had not asked the doctor to repeat himself because he didn't know enough about the surgery to make an informed choice, Marvin would not have realized that surgery was right for him, and he would have said 'no'.	Whilst the doctor is explaining some of the critical risks and benefits of the procedure, however, Marvin becomes distracted by a text message, and doesn't hear what the doctor is saying. Marvin is an intelligent, able adult: he can tell that he doesn't know enough about the surgery to make an informed choice, so he is perfectly capable of making sure he gets the information he needs. But he does not do so in this instance. He doesn't tell the doctor he was distracted, and the doctor does not repeat the relevant information. With little information about the procedure, he does not make his decision with regard for the pros and cons or whether surgery aligns with his personal values and what he really wants. Marvin simply says 'yes' to the surgery. If he had asked the doctor to repeat himself because he didn't know enough about the surgery to make an informed choice, Marvin would have realized that surgery was not right for him, and he would have said 'no'.	Whilst the doctor is explaining some of the critical risks and benefits of the procedure, however, Marvin becomes distracted by a text message, and doesn't hear what the doctor is saying. Marvin is not able and intelligent like most adults; he is completely incapable of making informed decisions. So he is completely unable to make sure he has the information he needs, understand his options properly, and make an informed choice. So he does not do so in this instance. He doesn't tell the doctor he was distracted, and the doctor does not repeat the relevant information. With little information about the procedure, he does not make his decision with regard for the pros and cons or whether surgery aligns with his personal values and what he really wants. Marvin simply says 'yes' to the surgery. Marvin is not able and intelligent like most adults; he simply can't tell that he doesn't know enough about the surgery to make an informed choice, so he is not capable of making sure he gets the information he needs.

SUPERSTITION		
Exercises Capacity	Mere Capacity	Lacks Capacity
Today is the 14th of April. Marvin is superstitious about the number 14: he thinks that 14 is his unlucky number and that things that come up on the 14th of the month are liable to be bad. Marvin is an intelligent, able adult: he is perfectly capable of distinguishing reasonable from unreasonable ways of making decisions and of making decisions in a reasonable way. And he does so in this instance. Even though fourteen is his unlucky number, he thinks the surgery through carefully, and makes his decision with careful regard for the pros and cons, and whether surgery aligns with his personal values and what he really wants. Because of this, Marvin says 'yes' to the surgery. If he had made this decision in an unreasonable way by not thinking it through properly and instead basing it entirely on his superstition, Marvin would not have realized that surgery was right for him, and he would have said 'no'.	Today is the 14th of April. Marvin is superstitious about the number 14: he thinks that 14 is his lucky number and that things that come up on the 14th of the month are liable to be good. Marvin is an intelligent, able adult: he is perfectly capable of distinguishing reasonable from unreasonable ways of making decisions and of making decisions in a reasonable way. But he does not do so in this instance. Because fourteen is his lucky number, he simply says 'yes' to the surgery. He doesn't think the surgery through even a little bit, and pays absolutely no attention to the pros and cons, or whether surgery aligns with his personal values and what he really wants. If he had made this decision in a reasonable way by thinking it through properly instead of basing it entirely on superstition, Marvin would have realized that surgery was not right for him, and he would have said 'no'.	Today is the 14th of April. Marvin is superstitious about the number 14: he thinks that 14 is his lucky number and that things that come up on the 14th of the month are liable to be good. Marvin is not able and intelligent like most adults: he is completely incapable of distinguishing reasonable from unreasonable ways of making decisions and of making decisions in a reasonable way. So he does not do so in this instance. Because fourteen is his lucky number, he simply says 'yes' to the surgery. He doesn't think things through even a little bit, and pays absolutely no attention to the pros and cons, or whether surgery aligns with his personal values and what he really wants. If he had been able to distinguish reasonable from unreasonable ways of making decisions and had thought this decision through properly instead of basing it entirely on superstition, Marvin would have realized that surgery was not right for him, and he would have said 'no'.

b. <u>Study 2 measures</u>

Consent

Consent 1: The doctor had Marvin's permission to proceed with the surgery.

Consent 2: If the doctor proceeds with the surgery now, he'll be acting without Marvin's consent.

Consent 3: Marvin's 'yes' didn't count as consent.

Morality: Under these circumstances, it would be morally wrong for the doctor to proceed with the surgery.

Freedom: Marvin freely chose to have surgery.

Manipulation checks

Rational capacity: Marvin has the ability to make rational decisions.

Rational exercise: Marvin made this particular decision rationally.

Authentic capacity: Marvin has the ability to be true to himself when making decisions.

Authentic exercise: When Marvin said 'yes' to having surgery, he was not being true to himself.

Autonomy capacity: Marvin has the ability to shape his own life freely according to his own values and what is right for him.

Autonomy exercise: The way Marvin made this particular decision expressed an ability to shape his life freely according to his own values and what is right for him.

c. Study 2 attention checks

Attention Checks, Impulse Condition

1. Which is correct?

a. Marvin IS able to resist and overcome impulses.

b. Marvin IS NOT able to resist and overcome impulses.

2. Which is correct?

a. Marvin made this particular decision by thinking it through properly.

b. Marvin made this particular decision on an impulse.

3. Which is correct?

a. Having surgery WAS the right choice for Marvin.

b. Having surgery WAS NOT the right choice for Marvin.

Attention Checks, Peer Pressure Condition

1. Which is correct?

a. Marvin IS able to resist pressure from others and make decisions for himself.

b. Marvin IS NOT able to resist pressure from others and make decisions for himself.

2. Which is correct?

a. Marvin made this particular decision to have surgery based on pressure from his friends.

b. Marvin made this particular decision to have surgery based on thinking it through for himself.

3. Which is correct?

a. Having surgery WAS the right choice for Marvin.

b. Having surgery WAS NOT the right choice for Marvin.

Attention Checks, Uninformed Condition

1. Which is correct?

a. Marvin WAS ABLE to tell that he didn't have enough information to make an informed choice.
b. Marvin WAS NOT ABLE to tell that he didn't have enough information to make an informed choice.

2. Which is correct?
    a. In the end, Marvin made his decision with SUBSTANTIAL information about the surgery.
    b. In the end, Marvin made his decision with LITTLE information about the surgery.
3. Which is correct?
    a. Having surgery WAS the right choice for Marvin.
    b. Having surgery WAS NOT the right choice for Marvin.

<u>Attention Checks, Superstition</u>

1. Which is correct?
    a. Marvin IS capable of distinguishing reasonable from unreasonable ways of making decisions.
    b. Marvin IS NOT capable of distinguishing reasonable from unreasonable ways of making decisions.
2. Which is correct?
    a. Marvin made this particular decision by thinking it through properly.
    b. Marvin made this particular decision by relying entirely on superstition.
3. Which is correct?
    a. Having surgery WAS the right choice for Marvin.
    b. Having surgery WAS NOT the right choice for Marvin.

3. Study 3 Stimuli

a. Study 3 Vignettes by condition

Exercises Capacity	Mere Capacity	Lacks Capacity
Jessica is an adult. Although she is occasionally known to be rash or impulsive, she is very able and intelligent. Indeed, she is perfectly capable of weighing up pros and cons, thinking through choices she faces, and making decisions based on what is best for her, which options align with her personal values, and what she really wants. Right now she is in the hospital recovering from surgery. She is fully awake, and, although she is on medication, it's only ibuprofen and some antibiotics. In fact, she feels calm and lucid, and nothing is interfering in any way with her ability to think or make decisions. Sam is a friend of Jessica's. He knows that Jessica has a diamond bracelet. Sam knows the bracelet is very precious to Jessica, but he thinks it would sell for a lot of money. He goes to the hospital to ask Jessica if he can sell her bracelet and split the money with her. She is not rash on this occasion: using her ability to make decisions according to her own values and what is best for her, Jessica says 'yes' after thinking things through very carefully, with careful regard for the pros and cons and whether it's what she really wants.	Jessica is an adult. Although she is occasionally known to be rash or impulsive, she is very able and intelligent. Indeed, she is perfectly capable of weighing up pros and cons, thinking through choices she faces, and making decisions based on what is best for her, which options align with her personal values, and what she really wants. Right now she is in the hospital recovering from surgery. She is fully awake, and, although she is on medication, it's only ibuprofen and some antibiotics. In fact, she feels calm and lucid, and nothing is interfering in any way with her ability to think or make decisions. Sam is a friend of Jessica's. He knows that Jessica has a diamond bracelet. Sam knows the bracelet is very precious to Jessica, but he thinks it would sell for a lot of money. He goes to the hospital to ask Jessica if he can sell her bracelet and split the money with her. She is rash on this occasion: despite her ability to make decisions according to her own values and what is best for her, Jessica just says 'yes' to the sale without thinking things through even a little bit, and with absolutely no regard for the pros and cons or whether it's what she really wants.	Jessica is an adult. Although she is occasionally known to be rash or impulsive, she is very able and intelligent. Right now she is in the hospital recovering from surgery. She is fully awake, but the medication she is on is incredibly powerful and is severely interfering with her ability to think and make decisions. Indeed, in her current state she is completely incapable of weighing up pros and cons, thinking through choices she faces, or making decisions based on what is best for her, which options align with her personal values, or what she really wants. Sam is a friend of Jessica's. He knows that Jessica has a diamond bracelet. Sam knows the bracelet is very precious to Jessica, but he thinks it would sell for a lot of money. He goes to the hospital to ask Jessica if he can sell her bracelet and split the money with her. Because of her current medicated state, she completely lacks the ability to make decisions according to her own values and what is best for her, so Jessica says 'yes' to the sale without thinking things through even a little bit, and with absolutely no regard for the pros and cons, or whether it's what she really wants.
ALL CONDITIONS continue as follows:		

Even though she said 'yes', Sam suspects that Jessica will actually be really sad later about parting with the bracelet. But he just wants to make money for himself. So he takes the bracelet from Jessica and puts it up for sale online.

A woman named Melanie, a complete stranger, sees the ad. She tells Sam that she's interested in buying the bracelet and asks how he came by it. Sam doesn't tell her the real story. Instead, he gives a really convincing story about purchasing it years ago from a jewelry store. Melanie has no reason to doubt his story.

Melanie decides to buy the bracelet. She transfers the money to Sam. Sam receives the payment and then mails Melanie the bracelet, which arrives safely by courier two days later.

b. Study 3 measures

Consent

Consent 1: Sam had Jessica's permission to proceed with the sale.

Consent 2: When Sam proceeded with the sale, he was acting without Jessica's consent.

Consent 3: Jessica's 'yes' didn't count as consent.

Morality: Under these circumstances, it was morally wrong for Sam to proceed with selling the bracelet.

Ownership, binary: Who is the rightful owner of the bracelet? Multiple Choice: Jessica/Melanie

Ownership Rights

Assume that Melanie could return the bracelet to Jessica and get her money back. Do you agree with the following statement?

Ownership 1: The bracelet does not truly belong to Melanie.

Ownership 2: Even if Melanie was told about the bracelet's true history, it would be morally acceptable for her to keep the bracelet if that's what she wanted to do.

Ownership 3: If Melanie was told about the bracelet's true history, it wouldn't just be nice of her to give the bracelet back to Jessica: it would be her *moral duty* to give it back.

Ownership 4: Melanie should be forced to return the bracelet.

Ownership 5: A good law would require the bracelet to be returned to Jessica under these circumstances.

Right choice: Saying 'yes' to selling the bracelet was probably the right choice for Jessica at the time.

Freedom: Jessica freely chose to sell her bracelet.

Manipulation checks

Rational capacity: When Sam asked whether he could sell the bracelet, Jessica had the ability to make a rational decision.

Rational exercise: Jessica made the decision to say 'yes' rationally.

Authentic capacity: Jessica had the ability to be true to herself when making a decision about whether to let Sam sell the bracelet.

Authentic exercise: When Jessica said 'yes' to Sam's selling the bracelet, she was not being true to herself.

Autonomy capacity: Marvin has the ability to shape his own life freely according to his own values and what is right for him.

Autonomy exercise: The way Marvin made this particular decision expressed an ability to shape his life freely according to his own values and what is right for him.

c. Study 3 attention checks

Capacity Check: Which is correct?

a. Jessica's medication interfered with her ability to think.
b. Jessica's medication DID NOT interfere with her ability to think.

Exercise Check: Which is correct?

a. Jessica said 'yes' WITH regard for whether she really wanted to sell the bracelet.
b. Jessica said 'yes' WITHOUT regard for whether she really wanted to sell the bracelet

AFTERWORD:

THE MERE CAPACITY VIEW

OF AUTONOMOUS CONSENT

In the previous chapter, I presented three studies on the ordinary concept of valid consent. These studies showed that this concept distinguishes between agents who possess autonomous decision-making capacities (whom we might call "competent" agents), and those who do not; specifically, people are less likely to judge that a "yes" from an agent who lacks autonomous decision-making capacities constitutes valid consent, whether that is consent to sex, consent to surgery, consent to entry, or consent to the sale of a belonging. However, we found little evidence that this concept is sensitive to whether an agent in fact exercises these capacities in coming to give consent—in other words, whether an agent *decides in an autonomous way*. So long as the agent possessed these capacities, participants tended to treat the agent's consent as at least as valid as someone deciding in a fully autonomous manner—even if the agent made their decision irrationally and without regard for their own values, and even if they were assenting to something that they did not, in fact, want and was not good for them.

Although Study 3 found that participants treated a competent agent's consent as slightly less valid when the vignette stated that they failed to decide in an autonomous way, participants were also less confident that the agent in this scenario really *did* have the capacity to decide autonomously. Their reluctance to accept the terms of the vignette may have been influenced by features of the scenario: given that the agent was in a vulnerable context (i.e. recovering from surgery in hospital), the agent made a very irrational and inauthentic decision, and there was no further explanation for why they did not make a better decision, participants may have inferred that their capacities were impaired after all.

This was reflected in exploratory data we collected in which we asked participants to justify their answers: a number of participants who were assigned to the Mere Capacity condition but who disagreed that the agent gave valid consent also raised doubts about the consenter's capacity to make an autonomous decision. For example, one participant wrote that "[r]egardless of what was said about Jessica being in her right mins [sic] it was obvious that she wasn't". A number of others suggested that situational factors would have undermined the agent's capacity. For example, one participant wrote, "I think that Jessica was in a stressful experience…I know that she is not on any mind altering medications, but surgery itself is a lot to handle"; another wrote, "Jessica is in the hospital recovering from surgery. Even if her medication isn't getting in the way of her thinking, her situation will certainly have worn her down and exhausted her."

So it's not clear that the lower ratings of consent were really due to the failure of the agent to decide autonomously, rather than the perception that her capacities were somewhat limited or impaired. Moreover, in this study, judgments of consent were dramatically reduced when it was stated that the agent lacked these capacities altogether. So, all things considered, we found consistent evidence that ordinary judgments of the validity of consent are sensitive to the possession of autonomous decision-making capacities, but little evidence that they are sensitive to the exercise of those capacities.

Many empirical questions remain about the ordinary concept. Moreover, we should not interpret these findings as indicating that people possess anything like a "theory" of valid consent. Still, these findings suggest that the ordinary concept of valid consent coheres much better with certain kinds of philosophical theories of the relationship between autonomy and valid consent over others. Specifically, the ordinary concept coheres better

with what I am calling the "Mere Capacity" view of valid consent than with the "Exercised Capacity" view.

According to the "Exercised Capacity" view, consent is valid only if agents come to consent on the basis of exercising their capacity for autonomous decision-making (we might say that, on such a view, valid consent requires autonomous agents to *make their decision autonomously*, or to *decide in an autonomous manner*). By contrast, according to the "Mere Capacity" view, it is not required that the agent in fact makes their decision autonomously—that is, on the basis of exercising their capacity for autonomous decision-making; it is only required that the agent possesses the capacity to make decisions autonomously (we might say, consent is valid only if the consent is given by an autonomous agent).

Of course, the terms "Exercised Capacity view" and "Mere Capacity view" really refer to *families* of views that are united in terms of their commitment to a claim regarding whether or not the exercise of autonomy is required for valid consent on top of the mere possession of autonomous decision-making capacities. First of all, particular views in either family might differ as to the exact nature of the capacities that must at least be possessed in order for consent to be valid, both in terms of the types of capacities required and in terms of their quality or extent. (For example, particular views might disagree about whether or not the capacity for rational reflection is necessary, and, amongst those that agree that it is necessary, they might disagree about whether a capacity for a more rudimentary form of rational reflection would suffice, or whether a capacity for more robust rational reflection is required.)

Secondly, particular versions of the Exercised Capacity view will have specific commitments regarding the way and extent to which these capacities must be exercised in order for consent to be valid. Consequently, different versions of the Exercised Capacity view may differ in their demandingness, depending on how robust and ideal this exercise of autonomy must be in order to yield valid consent. Because of this, depending on how exactly "autonomous decision-making" is defined, the empirical evidence outlined in the previous chapter is compatible with the possibility that the folk concept requires the exercise of autonomous decision-making in some more minimal sense than that tested. Still, the findings suggest that the folk concept coheres more closely to theories of consent that forego more robust Exercised Capacity requirements in favor of those that resemble Mere Capacity requirements.

In contrast to the folk concept, many existing accounts of valid consent build in Exercised Capacity requirements of various kinds. For example, in addition to capacity (or "competence"), dominant theories of consent in bioethics and in moral philosophy require that consent decisions be "informed" in order to be valid. This goes beyond merely requiring that the consenter possesses the capacity to understand their choice, as it precludes consent that is in fact based on significant ignorance, misunderstanding or false beliefs (e.g., Feinberg, 1986, p.152; O'Neill, 2003, p.4; Hanna, 2011, p.524; Faden & Beauchamp, 1986, Chapter 9).

Another standard requirement is that consent be "voluntary". "Voluntariness", in this literature, is variously interpreted as involving a more or less rich exercise of autonomy, such as: being intentional, or intentionally willed in accordance with a plan (Faden & Beauchamp, 1986); not being subject to unwelcome or manipulative influences, like scare tactics (e.g., Bullock, 2018, p.86); and, according to some, being in some sense

"authentic"—perhaps being given on the basis of desires or values with which the agent identifies (e.g., Dworkin, 1988, pp.24-5; for discussion, see Faden & Beauchamp, 1986, pp.262-8). According to influential accounts, inadvertent or habitual decisions fail to meet the relevant standards of voluntariness (e.g., Faden & Beauchamp, 1986, p.264). Other philosophers have argued that valid consent requires that decision-making conform to norms of theoretical rationality (Savulescu & Momeyer, 1997; Pugh, 2020). Others have emphasized the requirement that the consenter's decision-making be practically rational in light of their values, ruling out decisions that are based on a distortion of the consenter's own interests (Wertheimer, 2011, pp.149-156). To the extent that all of these views claim that a failure of autonomous decision-making to some specified degree (even in the presence of capacity) critically undermines the validity of consent, these views present a clear departure from the ordinary concept.

Similarly, the empirical findings suggest that ordinary judgments do not cohere well with the picture of valid consent that is often presupposed by ethical debates about the influence of 'nudges', framing effects, and various decision-making biases on valid consent, as such debates often assume that valid consent requires autonomous decision-making. As we saw earlier in this dissertation, many ethicists have worried that such influences invalidate consent on the basis that those influences are believed to: reduce understanding (Beauchamp & Childress, p.2013, p.135; Blumenthal-Barby, 2016); lead to decisions that are irrational or based on bad reasoning (Holm & Ploug, 2013); or lead to decisions that fail to reflect authentic preferences (Chwang, 2016; Mills, 2013), settled values (Hanna, 2011), or autonomous desires (Schwab, 2006). In a different vein, the thought that the only decisions that are relevantly expressive of an agent's autonomy are those that result from the *successful* exercise of rational, authentic decision-making capacities has led some theorists to defend paternalistic interference into agents' choices when those choices are

liable to be made irrationally or otherwise non-autonomously, even in the absence of impaired capacity (e.g., Thaler & Sunstein, 2003).

In contrast, the arguments I presented earlier in this dissertation (against the view that framing effects undermine the validity of consent) were consistent with Mere Capacity views of valid consent and thus with this feature of ordinary judgment, although they were also consistent with Exercised Capacity views. While I worked with some principles that the Mere Capacity view would reject as *necessary* conditions of valid consent, my argument only depended on treating them as *sufficient* conditions—a claim that both families of views could support. For example, I argued that an agent gives valid consent if the following conditions are jointly met (setting aside possible external control or interference that does not affect the agent's psychology):

1. **Reasonable weighing of values:** the agent accords sufficiently reasonable weights to different attributes in making their decision (where the reasonableness of the weight is determined by how well it corresponds to how much the agent values the attribute in question).

2. **Absence of incapacity:** in making their decision, the agent was not significantly hampered in their ability to consider and weigh values in a reasonable way.

Both the Exercised Capacity view and the Mere Capacity view are consistent with treating these conditions as jointly sufficient for valid consent. And both the Exercised Capacity view and the Mere Capacity view could agree that the second condition constitutes an individually necessary condition on valid consent. But the Exercised Capacity view will posit that meeting some interpretation of the first condition is also necessary for consent to be valid, while the Mere Capacity view denies this. For the purposes of my argument, I only relied on treating these conditions as jointly sufficient, remaining neutral about which of the conditions are necessary, and thus about the disagreement between Exercised Capacity and Mere Capacity views of valid consent. It was important to remain neutral

between these views, in order to provide a stronger rebuttal of the claim that framing leads to a violation of a necessary condition of valid consent.

That is not to say that my anti-revisionist arguments were neutral between all possible theories of the relationship between autonomy and valid consent. Notably, my arguments were not consistent with more idealized versions of the Mere Capacity or Exercised Capacity views, as my argument depended on the idea that valid consent is consistent with various suboptimalities of decision-making, such as a failure or inability to consider all of one's reasons, and a failure or inability to have perfect understanding of one's options; I argued that more optimal autonomy requirements would be overdemanding. These claims, however, also appear to be consistent with the ordinary concept, which appears to allow for decisions to constitute valid consent even if they are suboptimal in various ways, such as being irrational, impulsive, or dependent on peer pressure, resulting in the selection of options that are worse with respect to the agent's desires and values.

Coherence with ordinary judgment provides, at best, one source of defeasible and *prima facie* support for a theory of valid consent. Yet, there may be independent reasons for favoring the Mere Capacity view over the Exercised Capacity view. In the last few pages of this dissertation, I want to sketch some promising lines of argument that might be developed in support of the Mere Capacity view, while highlighting some of the questions that would need further exploration if a full defense of the view were to be successful. While I will not have the opportunity to fully develop any such defense here, I hope that these preliminary comments convince the reader that the Mere Capacity view is not merely a statement of a position that coheres with ordinary judgment, but has sufficient promise to be worth exploring and taking seriously in its own right.

Firstly, because it demands less of an agent's psychology—requiring only capacities, but not their successful exercise—the Mere Capacity view reduces the room available for skeptical or otherwise revisionist threats to ordinary practices of valid consent based on empirical findings from psychological sciences. Although it does not eliminate such threats in principle, empirical findings would have to show not only that decisions to consent are made in ways that fail to express a successful exercise of autonomous decision-making—for example, because they are irrational— but that this failure issues from an *incapacity* to decide autonomously. But we often fail to decide in an autonomous way even when we have the capacity to do so—many cases of whimsical, impulsive, habitual, inadvertent, unreflective, irrational, or foolish decisions are like this. So the empirically-motivated skeptic or revisionist would have to shore up more surprising evidence than merely pointing to deficiencies in the quality of ordinary decision-making processes in order to convince us of the surprising claim that agents *generally* do not validly consent (although, of course, it's highly plausible that some particular agents in particular circumstances do not have sufficient capacity to validly consent). This sets up a more stringent evidential standard that must be met if sweeping revisionism or skepticism is to be justified—which, you might think, is the correct result. Consequently, the Mere Capacity view allows us more easily to maintain a greater degree of coherence between theory and ordinary practices surrounding consent, which assumes that standards of valid consent are frequently met in ordinary contexts by ordinary agents.

Of course, to the extent that a theory avoids charges of being overdemanding, it runs the risk of demanding too little, classifying certain cases of consent as valid even when it should not do so. Indeed, many of the claims that a certain kind of exercise of autonomy is required for valid consent are made at least partly on the basis of appealing to particular cases where (1) the agent fails to exercise autonomy in some way, and (2) intuitively, it

would be ethically wrong to act on that consent. This is taken to support the view that the exercise of autonomy is necessary for consent to be valid and thus morally transformative.

For example, consider a case discussed by Bullock (2018, p.86) in which a doctor manipulatively uses scare tactics to get a patient to consent to some procedure (say, they belabor the possible consequences of refusing the procedure beyond what the provision of accurate information requires, using long, gruesome anecdotes and frightening photographs of extreme cases). Say the patient consents because of these scare tactics. Intuitively, it would be morally wrong for the doctor to act on the patient's consent. Bullock takes the use of scare tactics to constitute a kind of influence on consent that bypasses rational deliberation, and so concludes that such a case (amongst others) provides support for the view that decisional influences that bypass rational deliberation are not compatible with valid consent. The Mere Capacity view cannot classify this case of consent as invalid on the grounds that the agent makes an impulsive, emotion-based decision that merely *bypasses* their capacity for deliberation. For the Mere Capacity view, classifying the consent as invalid would have to be based on the claim that scare tactics undermine the agent's very capacity for rational deliberation. But it seems that many kinds of scare tactics could influence the decisions of ordinary patients, in a way that makes it unethical for the doctor to proceed, without it being the case that the scare tactics are so overwhelming that they can reasonably be said to undermine the agent's *capacity* for rational deliberation, as opposed to merely bypassing it. This kind of case is thus taken as a reason for thinking that the exercise of rational capacities—that is, making a decision on some basis grounded within one's capacity for rational deliberation—is necessary for valid consent.

As another example, we might consider a case in which a competent patient completely unthinkingly gives consent to a relatively serious medical procedure that, had she reflected,

she would have realized she does not actually want to do and will regret. Intuitively, if the doctor nevertheless acts on such consent, she has acted wrongly. This is most clear in a version of the case in which the doctor acts on the patient's consent while knowing that the patient gave their consent completely unthinkingly. But even if it simply did not occur to the doctor that the patient had given the procedure no real thought at all, it's still plausible that the doctor wrongs her, though in such a case we might blame the doctor less. An Exercised Capacity view, it seems, can easily account for our ethical qualms about such a case, by claiming that the failure of the agent to engage their capacity for rational decision-making on the basis of their own desires means that their consent is not sufficiently autonomous to be valid, and as such it fails to be morally transformative.

However, if we combine the Mere Capacity view with a nuanced understanding of the way in which consent is morally transformative, we might find not only a promising strategy for replying to the charge that the Mere Capacity view demands too little in such cases, but the beginnings of a promising picture of the ethics of consent. Perhaps surprisingly, this nuanced understanding begins with an observation that is also reflected in the ordinary judgments observed in the previous chapter: that there is a distinction between the validity of consent, and the all-things-considered moral wrongness of acting on that consent.

In line with standard treatments of consent, I have defined the morally transformative nature of consent in terms of its propensity, *all else being equal,* to make an act that would be impermissible without consent permissible instead. While valid consent does do this, it is a mistake to think of the fundamental ethical function of consent in terms of making acts *all-things-considered* morally permissible. Instead, it is more plausible to think that valid consent functions to remove a specific type of wrong—something like the wrong of violating the consenter's autonomy-based right to decide for themselves. But there are, of

course, many other ways of wronging someone: we can fail to care for them or give due regard to their wellbeing, because we are callous, selfish, or well-meaning but negligent; we can fail to treat them respectfully, although we do not interfere with their decisions; we can break promises to them; we can fail in duties of reciprocity, as when we fail to help a friend in need who has aided us in the past.

If this is right, then we can explain our intuitions that there is something morally troubling about cases in which agents fail to exercise their autonomous capacities in terms other than a failure to give valid consent. And, on reflection, such explanations seem highly plausible. Consider the last case in which a patient unreflectively acts against her true desires in consenting to a medical procedure. Insofar as the doctor wrongs the patient in acting on this consent, it does not seem that she does so because she commits the wrong of violating the patient's right to decide for themselves. Instead, it seems likely that the doctor has violated some other duty that is germane to the case: perhaps they have negligently violated a duty of care (for example, if she did not realize that the agent decided unreflectively, but she should have known, or at least should have taken more adequate steps to ensure that the procedure is something that the agent in fact wants and is good for them), or violated a duty of beneficence (for example, if she knew that the agent decided unreflectively and would likely come to regret the decision, but did not take the time to encourage them to reflect—because she doesn't care, because she is lazy, or because doing the procedure means a check in her pocket).

Furthermore, separating these different types of wrongs has the potential to illuminate the ethics of such interactions in a way that interpreting them in terms of invalid consent would not. We can see this by considering a slightly altered version of the case in which the patient decides the way she does because she lacks the *capacity* to decide reflectively at the time of

consenting (perhaps she is high on narcotics). Let's call this version of the case Incapacitated, and the former version Unreflective. And let's assume that, in both versions of the case, the doctor knows that the patient decided unreflectively, but doesn't care. Let's also set the stakes: though the patient does not really want to have this procedure, and having it will cause them regret, frustration and stress, it won't cause them very serious physical harm, and with some time and further procedures its effects can be largely reversed.

In both Unreflective and Incapacitated, the patient faces being wronged by a comparable violation of the duty of beneficence, since both cases involve the same potential act that will cause the same negative consequences of regret and so forth. But the Exercised Capacity and Mere Capacity accounts give different verdicts as to whether the patient faces an additional and qualitatively different wrong in Unreflective—an autonomy-based violation of their right to decide. According to an Exercised Capacity analysis, the doctor acts without the patient's valid consent in both Unreflective and Incapacitated; thus, on this account, both cases involve two wrongs, one beneficence-based, one autonomy-based. By contrast, according to a Mere Capacity analysis, the Unreflective patient gives valid consent, although the Incapacitated patient does not. Thus, according to the Mere Capacity view, there is an important ethical difference between the cases, since Unreflective does *not* involve a violation of the patient's autonomy-based rights.

This ethical difference posited by the Mere Capacity view seems to correspond to several other ethical differences between the cases. First and foremost, it provides a natural way of accounting for why the action in Incapacitated seems worse than the action in Unreflective. But there are additional ethical differences between the cases that the Mere Capacity analysis also seems well-positioned to explain. For instance, it seems that a

benevolent bystander is morally permitted to intrusively interfere to prevent the doctor acting on a patient who is incapacitated; by contrast, such an intrusion would be impermissible (indeed, unacceptably paternalistic) if the patient were only unreflective, even if the consequences for the patient's welfare of either the procedure or the intrusion would be similar in both cases. (Of course, we said earlier that an important move on behalf of the Mere Capacity view was to claim that some other duty is violated in Unreflective when acting on the agent's consent—perhaps a duty of beneficence. A fuller defense of this view would have to explain whether this duty demands that the doctor refrain from acting on the consent; and, if so, why such a refusal to *act on* the patient's consent is not also an unacceptably paternalistic intrusion into the decision of an autonomous agent.)

Another difference between the cases that the Mere Capacity view is able to capture is that the kind of moral attitude that seems warranted towards the wrongdoer seems different in each case: a doctor who knowingly acts on utterly unreflective consent may, in an egregious case, seem uncaring and callous at worst, but a doctor who knowingly acts on incapacitated consent seems altogether more sinister (assuming, of course, that there are no further facts about the case that would provide adequate justification for this act). A Mere Capacity analysis has a natural way of accounting for such moral differences, since it claims that the Unreflective agent gives valid, morally authoritative consent, while the Incapacitated agent faces a violation of their autonomy-based right to choose.

A proponent of the Exercised Capacity view has various ways of responding to these challenges. For example, they could deny that there really are these ethical differences between the cases. Or they could find a different way of accounting for them—for instance, by saying that the validity (and, correspondingly the violation) of consent is a

matter of degree, and that lacking capacity increases the degree to which one's consent is invalid, and thus the degree to which one suffers autonomy-based wrongs. Alternatively, another possible strategy would be to agree with the Mere Capacity view that the possession of autonomous decision-making capacities *does* have a qualitatively distinctive ethical function that accounts for the difference between these cases, but to argue that this difference is not one that pertains to *the validity of consent*; the proponent of the Exercised Capacity would then want to provide some explanation of what the additional, distinctive role of valid consent is supposed to be, over and above this putatively independent importance of the possession of autonomous decision-making capacities.

Another potential advantage of the Mere Capacity view that merits further investigation is that it has a natural way of capturing the intuition that the moral function of valid consent is to exercise the right to autonomously decide for ourselves what will happen to our own bodies and property from unconsented-to interference, and that the right to decide for oneself *includes* the right, and the corresponding normative power, to decide poorly—or, indeed, to not engage in a proper decision-making process at all. A proponent of the Exercised Capacity view, in response, might attempt to explain this intuition by appealing to higher-order decision-making, claiming that such decisions are *only* morally authoritative in cases where the agent has *autonomously decided* not to engage in a deliberative, autonomous decision-making process at a first-order level (for example, if an autonomous agent decides through a process of rational reflection that it will make them happier if they make some decisions spontaneously and without thought); the argument would be that whenever such first-order decisions are morally transformative, this power can be traced back to some exercise of autonomy at a higher-order level.

So these considerations are far from decisive. Future work will have to do much more to develop the details of a Mere Capacity theory, and to investigate whether or not convincing arguments can be given for favoring it over Exercised Capacity views.

CONCLUSION

Moral philosophers studying consent aim, thereby, to illuminate an important feature of ordinary ethical practice. Yet much philosophical discussion of consent is done independently of careful examination of empirical work detailing what the nature of these ordinary practices are—including how decisions to consent are actually made, and the factors that underlie ordinary ethical judgments about whether consent has been validly given. This is a mistake. On the one hand, it can lead to theories of consent that presuppose overly idealized pictures of autonomous decision-making. Not only do such theories fail to capture the decisions of ordinary, competent consenters, despite claiming to do justice to ordinary practice, they are developed without an awareness of the extent to which the theory departs from ordinary, intuitive concepts of consent, and thus without the provision of justifications for such a departure. On the other hand, it has tempted other philosophers to accept sweeping skepticism or pessimism about ordinary consent practices on the basis of psychological evidence that consent decisions are variable, disposed to be influenced by factors like framing effects—with the conviction that further investigation of how such factors come to influence decisions is not needed for such conclusions to be justly drawn.

In this dissertation, I have tried to forge a path that avoids these errors. In the first part of the dissertation, I outlined a picture of sufficiently autonomous, yet non-ideal, decision-making. I argued that this allows us to reconcile the commonsense view that ordinary decision-makers often give valid consent with the observation that consent is influenced by framing effects, because carefully examining empirical evidence illustrates that such factors can influence consent without bypassing, or subverting, autonomous decisions based on the agent's values. At the same time, in the second part of the dissertation, I showed that the commonsense view of valid consent is, in fact, less demanding than

philosophers have often taken it to be, requiring only that consenters have the capacity to decide autonomously, but not that they actually do so. While philosophical development of such a view must be left to future work, I hope, at least, to have shown that there is a promise for an ethical theory of valid consent—a Mere Capacity theory—that coheres in an important way with ordinary judgments, is more resistant to revisionism or skepticism while allowing for a psychologically realistic picture of ordinary decision-making, and that has the potential to provide an illuminating and nuanced picture of the complex ethical interactions that occur when a person gives consent.

BIBLIOGRAPHY

Ainiwaer, A., Zhang, S., Ainiwaer, X., & Ma, F. (2021). Effects of Message Framing on Cancer Prevention and Detection Behaviors, Intentions, and Attitudes: Systematic Review and Meta-analysis. *J Med Internet Res, 23*(9), e27634. URL: https://www.jmir.org/2021/9/e27634 DOI: 10.2196/27634

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin 126*, 556-574.

Allen, C. H., Vold, K., Felsen, G., Blumenthal-Barby, J. S., & Aharoni, E. (2019). Reconciling the opposing effects of neurobiological evidence on criminal sentencing judgments. *PLoS ONE 14*(1): e0210584. https://doi.org/10.1371/journal. pone.0210584

Almashat, S., Ayotte, B., Edelstein, B., & Margrett, J. (2008). Framing effect debiasing in medical decision making. *Patient Education and Counseling, 71,* (1), 102-107.

Altay, S., & Mercier, H. (2020). Framing Messages for Vaccination Supporters. *Journal of Experimental Psychology: Applied. 26,* 4, 567-578.

Appelbaum, P. S., & Roth, L. H. (1982). Competency to consent to research. A psychiatric overview. *Archives of General Psychiatry, 39*, 951–958.

Armstrong, K., Schwartz, J. S., Fitzgerald, G., Putt, M., & Ubel, P. A. (2002). Effect of Framing as Gain versus Loss on Understanding and Hypothetical Treatment Choices: Survival and Mortality Curves. *Medical Decision Making, 22*(1):76-83. doi:10.1177/0272989X0202200108

Banks, S. M., Salovey, P., Greener, S., Rothman, A. J., Moyer, A., Beauvais, J., & Epel, E. (1995). The Effects of Message Framing on Mammography Utilization. *Health Psychology, 14 (2),* 178-184.

Batchelder, E., Straight, C., Butt, M., & Kirby, J. S. (2020). Information framing effects on patients' decisions about dysplastic nevus management. *Journal of the American Academy of Dermatology, 82* (4), 1011-1013. DOI: https://doi.org/10.1016/j.jaad.2019.10.056

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4:Linear mixed-effects models using eigen and s4. *R package version 1*(7), 1–23.

Baumeister, R. F., & Monroe, A. E. (2014). Recent Research on Free Will: Conceptualizations, Beliefs, and Processes. *Advances in Experimental Social Psychology, 50*, 1–52.

Beach, L.R., Puto, C.P., Heckler, S.E., Naylor, G., and Marble, T.A. (1996), Differential versus unit weighting of violations, framing, and the role of probability in image theory's compatibility test. *Organizational Behavior and Human Decision Processes, 65*(2): 77-82.

Beauchamp, T. L. (2010). Autonomy and consent. In F. Miller & A. Wertheimer (Eds.), *The ethics of consent: Theory and practice* (pp. 55–78) New York, NY: Oxford University Press.

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of Biomedical Ethics* (7th Edition). Oxford: Oxford University Press.

Beres, M. A. (2014). Rethinking the concept of consent for anti-sexual violence activism and education. *Feminism & Psychology, 24*(3), 373–389.

Berg, J. W., Appelbaum, P. S., Lidz, C. W., & Meisel, A. (2001). *Informed Consent: Legal Theory and Clinical Practice (2nd edition)*, New York: Oxford University Press.

Bertoni, M., Corazzini, L., & Robone, S. (2020). The Good Outcome of Bad News: A Field Experiment on Formatting Breast Cancer Screening Invitation Letters. *American Journal of Health Economics*, *6* (3), pp. 372–409.

Best, R., & Charness, N. (2015). Age differences in the effect of framing on risky choice: A meta-analysis. *Psychol Aging. 30*(3):688-98. DOI: 10.1037/a0039447

Blakey, R., & Kremsmayer, T. P. (2018). Unable or Unwilling to Exercise Self-control? The Impact of Neuroscience on Perceptions of Impulsive Offenders. *Frontiers in Psychology, https://doi.org/10.3389/fpsyg.2017.02189*.

Blum, G., Bourdeau, J., Eclavea, R. P., Holben, J., Oakes, K., & Surette, E. C. (2021). *Fraud and Deceit*, American Jurisprudence (2nd ed).

Blumenthal-Barby, J. S. (2016). Biases and Heuristics in Decision Making and their Impact on Autonomy. *The American Journal of Bioethics, 16,* 5-15.

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive Biases and Heuristics in Medical Decision Making: A Critical Review Using a Systematic Search Strategy. *Medical Decision Making, 35,* (4), 539-557.

Bohns, V. K. (in press). Toward a psychology of consent. *Psychological Science.*

Bonnefon, J.F., Shariff, A., & Rahwan, I. (2020). The moral psychology of AI and the ethical opt-out problem. In M. Liao (Ed.), *The Ethics of Artificial Intelligence*. Oxford: Oxford University Press. (pp.109-126).

Broniatowski, D. A., & Reyna, V. F. (2018). A formal model of fuzzy-trace theory: Variations on framing effects and the Allais paradox. *Decision, 5,* 205–252. http://dx.doi.org/10.1037/dec0000083

Bui, T. C., Krieger, H. A., & Blumenthal-Barby, J. S. (2015). Framing Effects on Physicians' Judgment and Decision Making. *Psychological Reports: Mental & Physical Health, 117,* (2), 508-522.

Bullock, E. C. (2018). Valid Consent. In A. Müller & P. Schaber (Eds.), *The Routledge Handbook of the Ethics of Consent* (pp.85-94). New York, NY: Routledge.

Carling, C. L. L., Kristoffersen, D. T., Oxman, A. D., Flottorp, S., Fretheim, A., et al. (2010). The Effect of How Outcomes Are Framed on Decisions about Whether to

Take Antihypertensive Medication: A Randomized Trial. *PLOS ONE 5*(3): e9469. https://doi.org/10.1371/journal.pone.0009469

Carmody, M. (2005). Ethical erotics: Reconceptualizing anti-rape education. *Sexualities, 8*(4), 465-480.

Carmody, M., & Ovenden, G. (2013). Putting ethical sex into practice: Sexual negotiation, gender and citizenship in the lives of young women and men. *Journal of Youth Studies, 16*(6), 792-807.

Chakroff, A., & Young, L. (2015). How the mind matters for morality. *AJOB Neuroscience, 6*(3),41-46.

Chang, R. (1997). *Incommensurability, Incomparability, and Practical Reason*, Cambridge, MA: Harvard University Press.

Christensen, C., Heckerling, P., Mackesy-Amiti, M. E., Bernstein, L. M., & Elstein, A. S. (1995). Pervasiveness of Framing Effects among Physicians and Medical Students. *Journal of Behavioral Decision Making, 8,* 169-180.

Christman, J. (2020). Autonomy in Moral and Political Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), forthcoming URL = <http://plato.stanford.edu/archives/fall2020/entries/autonomy-moral/>.

Chwang, E. (2016). Consent's Been Framed: When Framing Effects Invalidate Consent and How to Validate It Again. *Journal of Applied Philosophy, 33,* 270-285.

Corbin, J. C., Reyna, V. F., Weldon, R. B., & Brainerd, C. J. (2015). How reasoning, judgment, and decision making are colored by gist-based intuition: A fuzzy-trace theory approach. *Journal of Applied Research in Memory and Cognition*, *4,* (4), 344-355.

Covey, J. (2014). The Role of Dispositional Factors in Moderating Message Framing Effects. *Health Psychology, 33,* (1), 52-65.

Crockett, M. (2013). Models of morality. *Trends in Cognitive Sciences, 17*(8), 363-366.

Cushman, F. A., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience, 7*(3), 269-279.

Daigle, J. L., & Demaree-Cotton, J. (2021). Blame mitigation: A less tidy take and its philosophical implications. *Philosophical Psychology*, DOI: 10.1080/09515089.2021.2000594

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 11*(4), 227-268.

Deci, E. L., & Ryan, R. M. (2009). Self-determination theory: A consideration of human motivational universals. In P. J. Corr & G. Matthews (Eds.), The Cambridge handbook of personality psychology (pp. 441–456). New York, NY: Cambridge University Press.

Demaree-Cotton, J. (2016). Do Framing Effects Make Moral Intuitions Unreliable? *Philosophical Psychology, 29*, 1-22.

Donovan, R. J., & Jalleh, G. (2000). Positive versus Negative Framing of a Hypothetical Infant Immunization: The Influence of Involvement. *Health Education & Behavior, 27,* (1) 82-95.

Dougherty, T. (2013). Sex, lies, and consent. *Ethics, 123*, 717–744.

Dougherty, T. (2019). Why does duress undermine consent? *NOUS*, 1–17. https://doi.org/10.1111/nous.12313

Douglas, C., & Proudfoot, E. (2013). Nudging and the Complicated Real Life of "Informed Consent". *The American Journal of Bioethics, 13,* 16-17.

Druckman, J. N. 2001. Using credible advice to overcome framing effects. *Journal of Law, Economics, & Organization, 17*, 62–82.

Dunegan, K. (2010). GPA and Attribute Framing Effects: Are Better Students More Sensitive or More Susceptible? *Journal of Education for Business, 85,* (4), 239-247.

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. New York, NY: Cambridge University Press.

Espinosa, J., & Starmans, C. (2020). Control it and it is yours: Children's reasoning about the ownership of living things. *Cognition, 202*, 104319.

Faden, R. R., and Beauchamp, T. L. (1986). *A History and Theory of Informed Consent*. New York: Oxford University Press.

Feinberg, J. (1986). Victims' Excuses: The Case Of Fraudulently Procured Consent. *Ethics, 96* (2), 330-345.

Feltz, A., & Cova, F. (2014). Moral responsibility and free will: A meta-analysis. *Consciousness and Cognition, 30,* 234-246.

Finkelstein, E. A., Cheung, Y. B., Schweitzer, M. E., Lee, L. H., Kanesvaran, R., & Baid, D. (2021). Accuracy incentives and framing effects to minimize the influence of cognitive bias among advanced cancer patients. *Journal of Health Psychology*, advance online publication. https://doi.org/10.1177/13591053211025601

Freling, T. H., Vincent, L. H., & Henard, D. H. (2014). When *not* to accentuate the positive: Re-examining valence effects in attribute framing. *Organizational Behavior and Human Decision Processes, 124,* 95-109.

Fridman, I., Fagerlin, A., Scherr, K. A., Scherer, L. D., Huffstetler, H., & Ubel, P. A. (2021). Gain-loss framing and patients' decisions: a linguistic examination of information framing in physician-patient conversations. *Journal of Behavioral Medicine, 44,* 38-52.

Friedman, O., Neary, K. R., Defeyter, M. A., & Malcolm, S. L. (2011). Ownership and object history. In H. Ross & O. Friedman (Eds.), *Origins of ownership of property* (pp. 79-89). Jossey-Bass.

Gallagher, K. M., & Updegraff, J. A. (2012). Health Message Framing Effects on Attitudes, Intentions and Behavior: A Meta-analytic Review. *Annals of Behavioral Medicine, 43,* 101-116.

Gallagher, K. M., Updegraff, J. A., Rothman, A. J., & Sims, L. (2011). Perceived susceptibility to breast cancer moderates the effect of gain- and loss-framed messages on use of screening mammography. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association*, *30*(2), 145–152. https://doi.org/10.1037/a0022264

Gamliel, E., & Kreiner, H. (2017) Outcome proportions, numeracy, and attribute-framing bias, *Australian Journal of Psychology, 69* (4), 283-292. DOI: 10.1111/ajpy.12151

Gamliel, E., & Kreiner, H. (2020). Applying fuzzy-trace theory to attribute-framing bias: Gist and Verbatim Representations of Quantitative Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(3), 497–506.

Gesser-Edelsburg, A., Walter, N., Shir-Raz, Y., & Green, M. S. (2015) Voluntary or Mandatory? The Valence Framing Effect of Attitudes Regarding HPV Vaccination. *Journal of Health Communication, 20* (11) 1287-1293, DOI: 10.1080/10810730.2015.1018642

Gong, J., Xiao, W., Gao, H., Wei, W., Zhang, W., Lv, J., Xiao, L., Duan, L., Zhang, Y., Liu, H., & Huang, Y. (2018). How to Best Convey Information About Intensive/Comfort care to the Family Members of Premature Infants to Enable Unbiased Perinatal Decisions. *Frontiers in Pediatrics, 6,* Article 348.

Gong, J., Zhang, Y., Feng, J., Zhang, W., Yin, W., Wu, X., Hou, Y., Huang, Y., Liu, H., & Miao, D. (2016). How best to obtain consent for thrombolysis. *Neurology, 86,* 1045-1052.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology, 47*, 55-130.

Gravelin, C. R, Biernat, M., & Bucher, C. E. (2019). Blaming the victim of acquaintance rape: Individual, situational, and sociocultural factors. *Frontiers in Psychology, 9,* 2422. https://doi.org/10.3389/fpsyg.2018.02422

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*, 101-124.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong, & C. B. Miller (Eds.), *Moral psychology: The neuroscience of morality: Emotion, brain disorders, and development*. Cambridge, MA: MIT Press (pp. 35-79).

Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. *Lund Psychological Reports, 2*, 1–23.

Hammond, E. M., Berry, M. A., & Rodriguez, D. N. (2011). The influence of rape myth acceptance, sexual attitudes, and belief in a just world on attributions of responsibility in a date rape scenario. *Legal and Criminological Psychology, 16*(2), 242-252.

Hanna, J. (2011). Consent and the Problem of Framing Effects. *Ethical Theory and Moral Practice, 14,* 517-531.

Hardisty, D. J., Johnson, E. J., & Weber, E.U. (2010). A dirty word or a dirty world? Attribute framing, political affiliation, and query theory. *Psychological Science, 21*(1), 86–92. https://doi.org/10.1177/0956797609355572.

Haward, M. F., Murphy, R. O., & Lorenz., J. M. (2008). Message Framing and Perinatal Decisions. *Pediatrics, 122,* 109-118.

Hayes, A. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from http://www.afhayes.com/public/process2012.pdf

Hayes, A., & Preacher, K. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology, 67*, 451–470.

Heilman, R. M., & Miclea, M. (2016). Risk seeking preferences: An investigation of framing effects across decisional domains. *Cognition, Brain, Behavior: An Interdisciplinary Journal, 20* (1), 1–17.

Henne, P., Niemi, L., Pinillos, Á., De Brigard, F., & Knobe, J. (2019). A counterfactual explanation for the action effect in causal judgment. *Cognition, 190*, 157-164.

Holleman, B. C., & Pander Maat, H. L. W. (2009). The pragmatics of profiling: Framing effects in text interpretation and text production. *Journal of Pragmatics, 41,* 2204-2221.

Hosmer, L. T. (1995). Trust: The Connecting Link Between Organizational Theory and Philosophical Ethics. *Academy of Management Review, 20* (2), 379-403.

Humphreys, T.P. and Herold, E. (2003). Should universities and colleges mandate sexual behavior? Student perceptions of Antioch College's consent policy. *Journal of Psychology & Human Sexuality, 15*(1), 35–51.

Hurd, H. (1996). The Moral Magic of Consent. *Legal Theory, 2*(2), 121-146. doi:10.1017/S1352325200000434

Jacoby, A., Baker, G., Chadwick, D., & Johnson, A. (1993). The impact of counselling with a practical statistical model on patients' decision-making about treatment for epilepsy: findings from a pilot study. *Epilepsy Research, 16,* 207-214.

Joffe, S., & Truog, R. D. (2010). Consent to Medical Care: The Importance of the Fiduciary Context. In F. Miller & A. Wertheimer (Eds.), *The Ethics of Consent: Theory and Practice* (pp.347-374). New York, NY: Oxford University.

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: a query theory of value construction. J*ournal of Experimental Psychology. Learning, Memory, and Cognition, 33*(3), 461–474. https://doi.org/10.1037/0278-7393.33.3.461.

Jozkowski, K. N., Peterson, Z. D., Sanders, S. A., Dennis, B., & Reece, M. (2014). Gender differences in heterosexual college students' conceptualizations and indicators of sexual consent: Implications for contemporary sexual assault prevention education. *Journal of Sex Research, 51,* 904–916. doi:10.1080/00224499.2013.792326

Kahan, D. M. (2010). Cultural, cognition, and consent: Who perceives what, and why, in acquaintance-rape cases. *University of Pennsylvania Law Review, 158*(3), 729-813

Kahneman D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kim, S., Goldstein, D., Hasher, L., & Zacks, R. T. (2005). Framing Effects in Younger and Older Adults. *The Journals of Gerontology: Series B, 60* (4), 215–218. DOI: https://doi.org/10.1093/geronb/60.4.P215

Kleinig, J. (2010). The Nature of Consent. In F. Miller & A. Wertheimer (Eds.), *The Ethics of Consent: Theory and Practice,* (pp.3-24). New York, NY: Oxford University Press.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*(3), 190-194.

Kreiner, H., & Gamliel, E. (2016). Looking at Both Sides of the Coin: Mixed Representation Moderates Attribute-framing Bias in Written and Auditory Messages. *Applied Cognitive Psychology, 30* (3) 332-340.

Kreiner, H., & Gamliel, E. (2017). Are highly numerate individuals invulnerable to attribute framing bias? Comparing numerically and graphically represented attribute framing. *European Journal of Social Psychology, 47,* 775-782.

Kreiner, H., & Gamliel, E. (2018). The role of attention in attribute framing. *Journal of Behavioral Decision Making, 31*(3), 392–401. https://doi.org/10.1002/bdm.2067

Kühberger, A. (1998). The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational Behavior and Human Decision Processes, 75* (1), 23-55.

Kuvaas, B., & Selart, M. (2004). Effects of attribute framing on cognitive processing and evaluation. *Organizational Behavior and Human Decision Processes, 95,* 198-207.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. doi: 10.18637/jss.v082.i13.

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: on the susceptibility to framing effects. *Journal of Behavioral Decision Making, 16* (2), 77-92.

Leong, L. M. (2020). The Role of Implicit Information in Choice Architecture. Doctoral dissertation, UC San Diego, USA. Retrieved from https://escholarship.org/uc/item/9xs2r56s

Leong, L.M., McKenzie, C.R.M., Sher, S., & Müller-Trede, J. (2017). The role of inference in attribute framing effects. *Journal of Behavioral Decision Making, 30* (5), 1147–1156. https://doi.org/10.1002/bdm.2030.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15,* 374-378.

Levin, I. P., Gaeth, G. J., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes, 88*, 411–429.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes, 76,* 149-188.

Levin, I. P., Schnittjer, S. K., & Thee, S. L. (1988). Information framing effects in social and personal decisions. *Journal of Experimental Social Psychology, 24,* 520-529.

Lim, G. Y., & Roloff, M. E. (1999) Attributing sexual consent. *Journal of Applied Communication Research, 27*(1), 1-23, DOI: 10.1080/00909889909365521

Macchi, L., & Zulato, E. (2021). The numbers do not add up! The pragmatic approach to the framing of medical treatments. *Judgment and Decision Making, 16,* (3), 596-613.

Malle, B. F, Guglielmo, S., & Monroe, A. E. (2014). A Theory of Blame. *Psychological Inquiry, 25*(2) 147-186, DOI: 10.1080/1047840X.2014.877340

Mandel, D. R., & Kapler, I. V. (2018). Cognitive Style and Frame Susceptibility in Decision-Making. *Frontiers in Psychology, 9,* 1461. doi: 10.3389/fpsyg.2018.01461

Marg, L. Z. (2020). College Men's Conceptualization of Sexual Consent at a Large, Racially/Ethnically Diverse Southern California University. *American Journal of Sexuality Education, 5*(3), 371-408, DOI: 10.1080/15546128.2020.1737291

Marteau, T. M. (1989). Framing information: Its influence upon decisions of doctors and patients. *British Journal of Social Psychology, 28,* 89-94.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review, 10,* 596-602.

McNeil, B. J., Pauker, S. G., Sox, H. C. Jr., & Tversky, A. (1982). On the Elicitation of Preferences for Alternative Therapies. *New England Journal of Medicine, 306,* 1259-1262.

Mele, A. (1995). *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press.

Millar, J. C., Turri, J., & Friedman, O. (2014). For the greater goods? Ownership rights and utilitarian moral judgment. *Cognition, 133*, 79-84.

Mills, C. (2013). Why Nudges Matter: A Reply to Goodwin. *Politics, 33,* 28-36.

Minton, A. R., Young, N. A., Nievera, M. A., & Mikels, J. A. (2020). Positivity helps the medicine go down: Leveraging framing and affective contexts to enhance the likelihood to take medications. *Emotion*. Advance online publication. https://doi.org/10.1037/emo0000798

Monroe, A. E., Brady, G. L., & Malle, B. F. (2017). This isn't the free will worth looking for: General free will beliefs do not influence moral judgments, agent-specific choice ascriptions do. *Social Psychological and Personality Science, 8*(2), 191-199. doi:10.1177/1948550616667616

Moxey, A., O'Connell, D., McGettigan, P., & Henry, D. (2003). Describing Treatment Effects to Patients: How They Are Expressed Makes a Difference. *Journal of General Internal Medicine, 18,* 948-959.

Moye, J., Gurrera, R. J., Karel, M. J., Edelstein, B., & O'Connell, C. (2006). Empirical advances in the assessment of the capacity to consent to medical treatment: Clinical implications and research needs. *Clinical Psychology Review, 26*, 1054-1077.

Muehlenhard, C. L., Humphreys, T. P., Jozkowski, K. N., & Peterson, Z. D. (2016) The Complexities of Sexual Consent Among College Students: A Conceptual and Empirical Review. *The Journal of Sex Research, 53*:4-5, 457-487, DOI: 10.1080/00224499.2016.1146651

Neary, K. R., Friedman, O., & Burnstein, C. L. (2009). Preschoolers infer ownership from "control of permission". *Developmental Psychology, 45*(3), 873–876

Neff, K. D., (2001). Judgments of personal autonomy and interpersonal responsibility in the context of Indian spousal relationships: An examination of young people's reasoning in Mysore, India. *British Journal of Developmental Psychology, 19*(2),233-257.

Niemi, L., & Young, L. (2014). Blaming the victim in the case of rape. *Psychological Inquiry, 25,* 2, 230-233.

Niemi, L., & Young, L. (2016). When and why we see victims as responsible: The impact of ideology on attitudes toward victims. *Personality and Social Psychology Bulletin, 42*(9),1227-1242.

O'Neill, O. (2003). Some limits of informed consent. *Journal of Medical Ethics*, *29*, 4-7.

Ortiz, D. V., Martinez, R. O., & Espino, D. V. (2015) Framing Effects on End-of-Life Preferences Among Latino Elders. *Social Work in Health Care, 54* (8), 708-724. DOI: 10.1080/00981389.2015.1059398

Peace, K. & Valois, R. (2014). Trials and tribulations: Psychopathic traits, emotion, and decision-making in an ambiguous case of sexual assault. *Psychology, 5,* 1239-1253. doi: 10.4236/psych.2014.510136.

Peng, J., Jiang, Y., Miao, D., Li, R., & Xiao, W. (2013). Framing effects in medical situations: Distinctions of attribute, goal and risky choice frames. *Journal of International Medical Research, 41* (3) 771-776.

Perneger, T. V., & Agoritsas, T. (2011). Doctors and Patients' Susceptibility to Framing Bias: A Randomized Trial. *J Gen Intern Med, 26,* 1411-7.

Peter-Hagene, L. C., & Ratliff, C. L. (2020). When jurors' moral judgments result in jury nullification: moral outrage at the law as a mediator of euthanasia attitudes on verdicts. *Psychiatry, Psychology and Law*, DOI: 10.1080/13218719.2020.1751741

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgment and decisions. *Current Directions in Psychological Science, 21*, 31–35. DOI: https://doi.org/10.1177/0963721411429960

Peters, E., Sol Hart, P., & Fraenkel, L. (2011). Informing Patients: The Influence of Numeracy, Framing, and Format of Side Effect Information on Risk Perceptions. *Medical Decision Making, 31* (2), 432-436.

Ploug, T., & Holm, S. (2015). Doctors, Patients, and Nudging in the Clinical Context—Four Views on Nudging and Informed Consent. *The American Journal of Bioethics, 15,* 28-38.

Pugh, J. (2020). *Autonomy, Rationality and Contemporary Bioethics*. Oxford: Oxford University Press.

Putrevu, S. (2014). Effects of Mood and Elaboration on Processing and Evaluation of Goal-Framed Appeals. *31,* (2), 134-146.

Rerick P.O., Livingston T.N., Davis D. (2019) Rape and the Jury. In: W. O'Donohue W., & P. Schewe (Eds.), *Handbook of sexual assault and sexual assault prevention.* Springer, Cham. https://doi.org/10.1007/978-3-030-23645-8_33

Reyna, V. (2018). When Irrational Biases Are Smart: A Fuzzy-Trace Theory of Complex Decision Making. *Journal of Intelligence, 6* (29), 1-16.

Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice. Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making, 4,* 249–262. http://dx.doi.org/10.1002/bdm.3960040403

Reyna, V., Nelson, W. L., Han, P. K., & Pignone, M. P. (2015). Decision Making and Cancer. *American Psychologist, 70* (2), 105–118.

Richardson, J. L., Milam, J., McCutchan, A., Stoyanoff, S., Bolan, Robert., Weiss, J., Kemper, C., Larsen, R. A., Hollander, H., Weismuller, P., Chou, C., Marks, G. (2004). Effect of brief safer-sex counseling by medical providers to HIV-1 seropositive patients. *AIDS*, *18,* (8), 1179-1186.

Rise, J., & Halkjelsvik, T. (2019). Conceptualizations of addiction and moral responsibility. *Frontiers in Psychology, 10,* 1483. doi: 10.3389/fpsyg.2019.01483

Rodríguez-Arias, D., Rodríguez López, B., Monasterio-Astobiza, A., & Hannikainen, I. R. (2020). How do people use 'killing', 'letting die' and related bioethical concepts? Contrasting descriptive and normative hypotheses. *bioethics, 34*(5), 509-518.

Rönnlund, M., Karlsson, E., Laggnäs, E., Larsson, L., & Lindström, T. (2005) Risky Decision Making Across Three Arenas of Choice: Are Younger and Older Adults Differently Susceptible to Framing Effects? *The Journal of General Psychology, 132* (1), 81-93, DOI: 10.3200/GENP.132.1.81-93

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*, 574-586. doi:10.1037/0022-3514.76.4.574

Savulescu J, & Momeyer, R. W. (1997). Should informed consent be based on rational beliefs? *Journal of Medical Ethics, 23,* 282-288.

Savulescu, J. (1994). Rational desires and the limitation of life sustaining treatment. *Bioethics, 8,* 191-222.

Schlenker, B. R., Britt, T. W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological Review, 101*, 632–652. doi: 10.1037/0033-295X.101.4.632

Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2013). Young children understand and defend the entitlements of others. *Journal of Experimental Child Psychology, 116*, 930-944.

Schneider, T. R., Salovey, P., Apanovitch, A. M., Pizarro, J, McCarthy, D., Zullo, J., & Rothman, A. J. (2001). The effects of message framing and ethnic targeting on mammography use among low-income women. *Health Psychology, 20,* 256-266.

Schwab, A. P. (2006). Formal and effective autonomy in health care. *Journal Medical Ethics, 32,* 575-579.

Shaver, K. G. (1985). *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*. New York, NY: Springer.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101,* 467-494.

Sher, S., & McKenzie, C. R. M. (2008). Framing effects and rationality. In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian cognitive science,* Chapter 4, pp.79-96. Oxford: Oxford University Press.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). Taylor & Frances/Routledge.

Siminoff, L. A., & Fetting, J. H. (1989). Effects of Outcome Framing on Treatment Decisions in the Real World: Impact of Framing on Adjuvant Breast Cancer Decisions. *Medical Decision Making, 9*(4), 262–271. https://doi.org/10.1177/0272989X8900900406

Siminoff, L. A., & Fetting, J. H. (1991). Factors affecting treatment decisions for a life-threatening illness: The case of medical treatment of breast cancer. *Social Science & Medicine, 32*(7), 813–818. https://doi.org/10.1016/0277-9536(91)90307-X

Sinnott-Armstrong, W. (2008). Framing Moral Intuitions. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity (*pp. 47–76). Cambridge, MA: MIT Press.

Slothuus, R. (2008). More Than Weighting Cognitive Importance: A Dual-Process Model of Issue Framing Effects. *Political Psychology, 29,* (1), 1-28.

Smith, I., Ancillotti, M., de Bekker-Grob, E., & Veldwijk, J. (2020). PDG77 The IMPACT of Framing an Attribute As Failure or Effectiveness on Preferences for Antibiotic Treatment in a Discrete Choice Experiment. *Value in Health, 23,* Suppl. 2., S531. DOI: https://doi.org/10.1016/j.jval.2020.08.760

Smith, S. M., & Levin, I. P. (1996). Need for Cognition and Choice Framing Effects. *Journal of Behavioral Decision Making, 9,* 283-290.

Sommers, R. (2020). Commonsense consent. *The Yale Law Journal, 129*(8), 2232-2324.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672–695. https://doi.org/10.1037/0022-3514.94.4.672

Starmans, C., & Friedman, O. (2016). If I am free, you can't own me: Autonomy makes entities less ownable. *Cognition, 148*, 145-153.

Steiger, A., & Kühberger, A. (2018). A Meta-Analytic Re-Appraisal of the Framing Effect. *Zeitschrift für Psychologie, 226,* (1), DOI: https://doi.org/10.1027/2151-2604/a000321

Steward, W. T., Schneider, T. R., Pizarro, J., & Salovey, P. (2003). Need for cognition moderates responses to framed smoking-cessation messages. *Journal of Applied Social Psychology*, *33*, 2439–2464.

Struchiner, N., Hannikainen, I., & de Almeida, G. d. FCF. (2020). An experimental guide to vehicles in the park. *Judgment and Decision Making, 15* (3), 312-329.

Tabesh, P., Tabesh, P., & Moghaddam, K. (2019). Individual and contextual influences on framing effect: Evidence from the Middle East. *Journal of General Management, 45*(1), 30–39. https://doi.org/10.1177/0306307019851337

Teigen, K. H. (2015). Framing of Numerical Quantities. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making, Vol. II*, (pp. 568-589), John Wiley & Sons.

Thaler, R, H., & Sunstein, C. R. (2003). Libertarian Paternalism. *American Economic Review, 93*(2), 175-179.

Tobia, K. P. (2022, Forthcoming). Experimental Jurisprudence, *89 University of Chicago Law Review*. Available at SSRN: https://ssrn.com/abstract=3680107 or http://dx.doi.org/10.2139/ssrn.3680107

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science, 211,* 453-458.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A Person-Centered Approach to Moral Judgment. *Perspectives on Psychological Science, 10*(1), 72-81.

Urs, M., Goodmon, L. B., & Martin, J. (2019). Too Much on My Mind: Cognitive Load, Working Memory Capacity, and Framing Effects. *North American Journal of Psychology, 21*, (4), 739-768.

Van 't Riet, J., Ruiter, R., & De Vries, H. (2011). Preaching to the choir? The influence of personal relevance on the effects of gain- and loss-framed health-promoting messages. *Journal of Health Psychology, 17,* (5), 712-723.

Van de Vondervoort, J. W., & Friedman, O. (2015). Parallels in preschoolers' and adults' judgments about ownership rights and bodily rights. *Cognitive Science, 39*, 184-198.

Vonasch, A., Baumeister, R., & Mele, A. (2018). Ordinary people think free will is a lack of constraint, not the presence of a soul. *Consciousness and Cognition, 60*, 133-151.

Wall, D., Crookes, R. D., Johnson, E. J., & Weber, E. U. (2020). Risky choice frames shift the structure and emotional valence of internal arguments: A query theory account of the unusual disease problem. *Judgment and Decision Making, 15* (5), 685–703.

Weber, E. U., Johnson, E. J., Milch, K. K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science, 18*(6), 516–523. https://doi.org/10.1111/j.1467-9280.2007.01932.x.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York, NY: Guilford Press.

Wertheimer, A. (2003). *Consent to sexual relations*. Cambridge: Cambridge University Press.

Wertheimer, A. (2011). *Rethinking the Ethics of Clinical Research*. Oxford: Oxford University Press.

Whatley, M. A. (1996). Victim characteristics influencing attributions of responsibility to rape victims: A meta-analysis. *Aggression and Violent Behavior, 1*(2), 81-95.

Whisnant, R. (2021). Feminist Perspectives on Rape. The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2021/entries/feminism-rape/>.

Wignall, L., Stirling, J., & Scoats, R. (2020). UK university students' perceptions and negotiations of sexual consent. *Psychology & Sexuality*, DOI: 10.1080/19419899.2020.1859601

Wilson, D. K., Kaplan, R. M., & Schneiderman, L. J. (1987). Framing of Decisions and Selections of Alternatives in Health Care. *Social Behaviour, 2,* 51-59.

Wirtz, J. G., Sar, S., & Ghuge, S. (2015). The Moderating Role of Mood and Personal Relevance on Persuasive Effects of Gain- and Loss-Framed Health Messages. *Health Marketing Quarterly, 32* (2), 180-196, DOI: 10.1080/07359683.2015.1033936.

Woodhead, E. L., Lynch, E. B., & Edelstein, B. A. (2011). Decisional Strategy Determines Whether Frame Influences Treatment Preferences for Medical Decisions. *Psychology and Aging, 26,* (2), 285-294.

Yang, D., Wang, C., & Chen, C. (2015). Promoting Functional Outcome of Stroke Patients: The Effect of Regulatory Focus, Therapy Frequency and Message Framing. *International Journal of Management, Economics and Social Sciences, 4,* (2), 42-57.

Yndo, M. C., & Zawacki, T. (2020). Factors influencing labeling nonconsensual sex as sexual assault. *Journal of Interpersonal Violence, 35,* 1803-1827.