

# Consent and Concepts

by

Jerome Hodges

M.S., North Carolina State University (2010)

Submitted to the Department of Linguistics and Philosophy  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

© Massachusetts Institute of Technology 2020. All rights reserved.

Author .....  
Department of Linguistics and Philosophy  
September 1, 2020

Certified by.....  
Sally Haslanger  
Ford Professor of Philosophy and Women's and Gender Studies  
Thesis Supervisor

Certified by.....  
Stephen Yablo  
David W. Skinner Professor of Philosophy  
Thesis Supervisor

Certified by.....  
Bradford Skow  
Laurance S. Rockefeller Professor of Philosophy  
Thesis Supervisor

Accepted by .....  
Bradford Skow  
Chair of the Committee on Graduate Students



# Consent and Concepts

by

Jerome Hodges

Submitted to the Department of Linguistics and Philosophy  
on September 1, 2020, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## Abstract

This dissertation lays out the groundwork for building a theory of radical consent and autonomy.

Chapter 1, “Framing Consent,” argues for a context-sensitive account of the semantic content of consent claims, and presents an introductory model of consenting as a speech act. In particular, I argue for a *contrastive* model of consent claims, in which consent is given against a backdrop of relevant alternatives. More generally, I argue that the context-sensitivity of the sort guaranteed by this model — *viz.* context-sensitivity of *what is consented to* in a consent claim — is an ineluctable feature of such claims. This has far-reaching consequences, I claim, for the use of consent in both normative ethics and political philosophy.

Chapter 2, “Conceptual Amelioration and Epistemic Responsibility,” co-authored with Ekaterina Botchkina, looks at the question of conceptual amelioration more generally: when thinking about concepts,<sup>1</sup> what is the proper role of value considerations? There we argue that, under a remarkably theory-neutral constraint, at least some conceptual interventions motivated by concerns of justice are acceptable.

In particular, we suggest that particular accounts of amelioration offered or suggested by Mark Richard and Sally Haslanger are too restrictive in their metaphysical and semantical commitments to provide a general account of how amelioration is possible. Instead, we suggest that the core aim of amelioration can be understood as preserving a sort of *conceptual possibility*, and that this preservation is precisely what is aimed at in scientific theories’ development of concepts — in the decision, for instance, to abandon the concept of [*ether*] while retaining (but refining) the concept of [*atom*]. As such, amelioration isn’t as unusual — or as troubling — as it might at first blush appear. We use a tool suggested by Steve Yablo — what we call the “Turning-Out Test” — as a way to test for conceptual possibility, and thus epistemic responsibility.

Chapter 3, “Performative Consent and Autonomy,” returns to consent, and specifically to the role of consent in legal and quasi-legal contexts of sexual assault and battery. I argue there that the speech-act account suggested in Chapter 1 is justi-

---

<sup>1</sup>Or, as some would put it, *doing conceptual analysis*.

fied in such contexts on the ameliorationist grounds articulated in Chapter 2. I then extend these considerations to extra-legal contexts, and show that this account is compatible with radical feminist claims around sexual assault. In particular, I defend the possibility of an account of assault in which, at least sometimes, *whether a sexual assault has occurred* may depend upon a survivor's posterior assessment of the event.

Thesis Supervisor: Sally Haslanger

Title: Ford Professor of Philosophy and Women's and Gender Studies

Thesis Supervisor: Stephen Yablo

Title: David W. Skinner Professor of Philosophy

Thesis Supervisor: Bradford Skow

Title: Laurance S. Rockefeller Professor of Philosophy

## Acknowledgments

Thanks are due to more people than I can possibly remember. Thanks especially to the MIT Philosophy community for a uniquely engaging and supportive environment; to John Carroll, Sandy Paur, and David Auerbach for teaching me how to think; to Lyndal Grant, Arden Ali, Ari Koslow, Matt Mandelkern, Milo Philips-Brown, Rose Lenehan, Marion Boulicault, Cosmo Grant, Brendan Dill, Kate Vredenburgh, Zeynep Soysal, Celine Leboeuf, Kevin Dorst, Bernhard Salow, Brendan de Kenessey, Nilanjan Das, Melissa Schumacher, Diana Acosta Navas, Nina Emery, Ginger Schultheis, Saray Ayala, and Nadiya Vasilyeva for countless fruitful conversations; to Dylan Bianchi, Rebecca Millsop, Arden Ali, and Jack Marley-Payne for co-constituting the best cohort I could have hoped for; to Arden Ali, Sally Haslanger, Damien Rochford, and Patricia Marechal for support in the roughest times; to the amazing group of scholars I work with now in algorithmic ethics at the Jain Family Institute, including David Grant, Lily Hu, Milo Philips-Brown, Arden Ali, Greg Keenan, and Francis Tseng; to Michael Stynes, for helping me find time to complete this; to Jack Spencer, Japa Pallikkathayil, Justin Khoo, Vann McGee, Julia Markovits, Sally Haslanger, Mark Richard, and Kieran Setiya for feedback on early versions; to Tommie Shelby, Sally Haslanger, Bob Stalnaker, and Steve Yablo for critical feedback, and, generally, for making my simplest confusions seem profound; to Barry Cynamon and Sidhya Balakrishnan for their patience; to Katya Botchkina for inspiration, collaboration, and general brilliance; to Brad Skow and Sally Haslanger for their mentorship and guidance throughout my time at MIT; to Sally Haslanger, Mary Kate McGowan, Nancy Bauer, and everyone else in WOGAP for creating a place to be; to Elijah Fortes for his endlessly loving patience and caregiving; and to Nikita Shepard for, well, everything.



# Contents

<b>1</b>	<b>Framing Consent</b>	<b>13</b>
1.1	Introduction . . . . .	13
1.2	The View . . . . .	14
1.3	Consenting and Consent Claims . . . . .	18
1.4	Considerations . . . . .	23
1.4.1	Surface Evidence . . . . .	24
1.4.2	Paradoxes . . . . .	25
1.4.3	Logical Questions . . . . .	30
1.5	Reclaiming Consent . . . . .	35
1.5.1	The Bog . . . . .	36
1.5.2	A Way Out . . . . .	39
1.6	Conclusion: Radical Theory, Radical Consent . . . . .	40
<b>2</b>	<b>Conceptual Amelioration and Epistemic Responsibility</b>	<b>49</b>
2.1	What is amelioration? . . . . .	50
2.2	Existing approaches to amelioration . . . . .	57
2.2.1	Haslanger: amelioration as metasemantic . . . . .	57
2.2.2	Mark Richard: concept identity across alterations . . . . .	62
2.3	Epistemic Responsibility . . . . .	67
2.4	Conceptual possibility and the Turning-Out Test . . . . .	74
2.5	Is amelioration enough? . . . . .	78

<b>3</b>	<b>Performative Consent and Autonomy</b>	<b>85</b>
3.1	The attitudinal model and its discontents . . . . .	85
3.2	The performative model . . . . .	88
3.3	Challenges to the performative model . . . . .	89
3.4	The disjunctive option . . . . .	94
3.5	Deflationism . . . . .	98
3.6	Radical consent . . . . .	99
3.7	Conclusion . . . . .	104



# List of Figures

2-1	The universe of conceptual interventions . . . . .	56
-----	--	----



# List of Tables



# Chapter 1

## Framing Consent

### 1.1 Introduction

Consent is a tricky subject for radical theory. On the one hand, some of the deepest and most compelling social and political critiques have consent at their heart: the state’s authority is unjust, claims the classical anarchist, precisely because its subjects do not consent to being governed. On the other hand, engaging in revolutionary practice requires, at least sometimes, ignoring the dissent of some of those affected. Radical theory, it seems, needs to be able to distinguish between consent that matters and consent that doesn’t.

Consent is similarly troublesome for ethical theory more generally. The *prima facie* plausible thought that we’re morally required to treat others only in ways to which they consent — even once “treating others” is suitably fleshed out — quickly falls to counterexamples: addicts who need treatment and patients who retract prior consent in the midst of painful surgery come immediately to mind. And further reflection uncovers a range of cases in which obtaining even explicit consent seems insufficient for meeting one’s obligations to take others into account; at minimum, as Onora O’Neill (1985, 258) puts it, “we need an account of genuine, morally significant consent, and to distinguish this from spurious or morally trivial consent.”

In what follows, I’ll sketch a view about the semantics of consent that, I hope, shows where there’s theoretical space for making these sorts of distinctions and resolv-

ing some of the more recalcitrant philosophical puzzles; taking a page from work in the philosophy of explanation and deontic logic, I'll argue that consent claims are best understood as *contrastive* claims. After presenting a preliminary account of what consenting might be, I'll first offer some linguistic considerations supporting this view, by showing that consent statements have a natural affinity with other contrastive statements. Next, I'll show how the contrastivity thesis offers up the possibility of productively reframing thorny questions around consent's ethical role.

In the last section, I'll turn to the application of this idea to the uses of consent in radical theorizing, showing that the distinctions the contrastive view allows make room for distinguishing between those times when radicals should, and those times when we shouldn't, prioritize obtaining consent from those affected by our actions. Finally, I'll offer some reasons to think that the conception of radical duties around consent that naturally emerges from this view aligns with a project radicals should take seriously anyway — namely, that of subverting hegemonic ideology.

## 1.2 The View

On the view I'm defending, consent is a fundamentally *contrastive* relation, between an agent, an option, and a set of alternative options. This is to be contrasted with what I'll call the *standard* view, on which consent is a dyadic relation between an agent and an option. According to the contrastive thesis, the semantic content of the claim includes more than a single state of affairs; it includes, as well, a class of alternatives.<sup>1</sup>

A typical example of a contrastive relation is given by *preferring* (see, *e.g.*, Schaffer, 2004). As a first pass, we might say that there's something missing in the sentence (standing alone) "John prefers Thai food." On hearing it, it's appropriate to ask, "John prefers Thai food *to what?* To American fast food? To all other sorts of food? To 1950s game shows?" Without an answer to this question, the sentence doesn't (to ape

---

<sup>1</sup>It's tempting to think that if *consenting* is contrastive, such directed affective states as *hoping*, *wishing*, *desiring*, and *fearing* are, too. Just to be clear, although I'm sympathetic to it, I'm not arguing for that thesis here.

Frege) express a complete thought. An utterance of it only communicates, indirectly, that there's something to which John prefers Thai food.

This is not quite to say that we can't *use* the sentence "John prefers Thai food" to express a complete thought. In ordinary uses, it will often be the case that the answer to the completion question "to what?" is supplied by conversational context. When discussing what to order for dinner, for instance, the natural contrast is with the other kinds of food available. To put the point in semantical terms, "prefers" denotes or expresses a relation that is ternary ( $S$  prefers  $X$  to  $Y$ ) rather than binary ( $S$  prefers  $X$ ); the final slot may be filled in explicitly ("John prefers Thai to New American") or implicitly through context ("Should we order pizza, curry, or what?" "John prefers Thai"), but a foil must be filled in *somehow* in order for a sentence using "prefers" to have determinate content.

The crux of my position is that the relation of consenting is like that of preferring: it's irreducibly contrastive. *Consenting* is a fundamentally triadic relation, between an agent, a situation, and a set of (alternative) situations; absent a set of alternatives, a consent statement is incomplete.<sup>2</sup>

To fill in this sketch a little more, I'll take a page from contrastive approaches in the philosophy of explanation, where contrastivists hold that explanations, and

---

<sup>2</sup>I'm assuming here and in what follows that *situations*, or *states of affairs*, are the right sorts of things to be the objects of consent claims. One alternative to this is to take *actions* to be the objects. This approach allows for a second way that contrastivity might appear. If actions are *themselves* contrastive — if, for instance, a complete characterization of an action requires a characterization of some set of alternatives — then contrastivity might appear *indirectly*. By having a fundamentally contrastive *object*, a binary relation might well have contrastive features. Crucially, though, on either way of cashing this issue out, we end up with a contrast class in the semantics — either in the structure of the relation itself or in that of the object. This feature is enough to guarantee the sort of contrastivity I'm after here.

I think there's good reason to prefer the states-of-affairs view over the action view. In particular, there seem to be a number of cases in which what we consent to isn't (obviously) an action. We can consent to a system of government, to a change in status, to a treatment strategy, to a new obligation (Raz (1981, 120) makes a similar point). None of these is naturally characterized as an action, but each has a transparent situational analogue: consenting to a system of government's being in place, a change in status's occurring, a treatment strategy's being followed, an obligation's binding. Of course, this consideration isn't decisive; there may be ways to massage each of these into similar-enough actions. But, without compelling reason to go this route, it seems preferable to do minimal violence to the surface form, and allow that there are non-actions which can be objects of consent.

the “why?” questions they answer, are always (if often covertly) contrastive.<sup>3</sup> Alan Garfinkel (1981, 30f) elaborates this notion using the technical apparatus of a *contrast space* or *frame*: a set of mutually exclusive relevant alternatives relative to which an explanation (or purported explanation) does its work. Such a space determines both which situations count as essentially equivalent for the purposes of the explanation, and which situations don’t count as relevant alternatives at all.

What determines (at least roughly) the frame at work in a given explanation? One set of parameters appears in the form the explanation takes: what the *explanantia*, *explananda*, and explicitly mentioned contrasts are. Additionally, there are pragmatic factors: overall context, conversational presuppositions, interests, and the concepts involved in the discourse all play a role. There’s a sense in which the question of precisely *how* this machinery works is orthogonal to that of contrastivity; unless we’re in doubt as to whether pragmatic mechanisms are capable of serving up contrast spaces, the absence of a story as to how they *do* deliver them is no reason to doubt the contrastive picture. And nobody should doubt that pragmatic mechanisms provide contrast classes, for they must do so with uncontroversially contrastive relations. To crib from Jonathan Schaffer (2004, 80): “whatever story one prefers for ‘prefers’, tell that story”.

Having duly fastened that parachute, there are some gestures I can make as to what I suspect the mechanisms are that govern contrast classes in a wide range of contrastive statements. Garfinkel, in his theory of explanation, insists on the importance of keeping in mind the why-question an explanation is addressing:<sup>4</sup>

Perhaps the most interesting cases of changes in explanatory frames are ones in which there is a shift in the nature of the *question* being asked. Explanations are sometimes answers to explicit questions . . . But often there is no explicit question at hand, and in those cases it can be very interesting to ask what question the explanation is really answering. (Garfinkel, 1981, 8)

---

<sup>3</sup>See, *e.g.*, (van Fraassen, 1980; Garfinkel, 1981; Sober, 1986; Schaffer, 2005).

<sup>4</sup>The idea of analyzing explanation in terms of questions traces back, as far as I can tell, to Sylvain Bromberger (see, *e.g.*, Bromberger, 1966).



More recently, Michael Glanzberg (2003) has proposed a contrastive semantics of focus-sensitive statements built up from the semantics of questions. Otherwise identical statements that differ in focus — say, “John kissed *Jamal*” and “*John* kissed Jamal” — differ in the questions they answer — respectively, “Whom did John kiss?” and “Who kissed Jamal?”. According to the prevailing view in the semantics of questions, a question takes a contrast class — a set of possible answers — as its semantic value. If we attend to the place of a focus-sensitive statement within a dialogue, this gives a quite natural way to fill in its contrast class: in the simple case, the contrast class of a focus-sensitive statement *just is* the semantic value of the question to which it responds. Of course, real-life conversations rarely take the form of question-answer interrogations, and the complications needed to address them would take us very far afield. Nevertheless, this suggests that there may well be fairly straightforward ways in which contrast spaces are determined by the structure and content of a conversation.

It’s worth making explicit one consequence of this picture of the pragmatics as applied to consent: the contrast set — and thus, on my view, the content of a consent claim — is *not* determined (solely) by the speaker’s intentions. This is stronger than the simple observation that a speaker need not have in mind explicit alternatives when she consents. It allows the possibility that, even if the speaker explicitly has a contrast class in mind, the frame that governs her consent claim may differ from it. Put another way: a speaker may consent to something, and yet (1) not know the content of her consent, (2) not believe she’s consenting to what she’s consenting to, (3) believe she’s consenting to something else, and (4) believe that she’s not consenting to what she is in fact consenting to. These are possibilities that some may find insufferable.

I said the picture allows this possibility; it doesn’t, however, necessitate it. There are at least two ways out.<sup>5</sup> The first is to accept that there is a univocal context that determines a contrast class that may differ from a speaker’s intended frame, but insist

---

<sup>5</sup>That is, two *plausible* ways. There’s also the strategy of insisting that the details of the mechanism (which may involve speaker intention indirectly) guarantee that such mismatches never arise. This is particularly implausible given that contrasts may never be voiced.

that the felicity conditions governing consent preclude its being given in such cases; in Austin’s (1975) terms, consent statements in such contexts *misfire*. This is what happens on the model sketched in §1.3: agent intent is built into what it is to consent. The second is to allow intentions to enter the mechanism directly, either by a system of rules that governs what happens in cases of mismatches between participants, or by allowing a multitude of participant-specific contexts.<sup>6</sup>

### 1.3 Consenting and Consent Claims

Although my primary purpose in this paper is to present a thesis about the semantic and pragmatic features of consent statements, and not to give a substantive analysis of consent and the ethical principles concerning it, some preliminary observations about the nature of consent, and, in particular, the relationship between *acts* of consenting and what consent claim *express*, may help to clarify both the scope of the proposal and the work to which I put it in the second half of this paper. To that end, in this section I’ll outline, without too much defense, a model of what it is to consent, before returning to the main project.

In the sense in which I’ll be using it, consenting is a speech act — in Austin’s (1975) terms, an illocution — performed when an agent engages in some custom or other to indicate a particular stance.<sup>7</sup> Just as promising is the sort of act performed by a person who, in ordinary contexts, says “I promise”, consenting is the sort of speech act performed by a person who, in ordinary contexts, says “I consent”. So: consenting is an illocution, archetypically performed by saying “I consent”.

When all goes well, we take consent to have a certain *legitimizing* force: if I consent to your action, and I’m the only affected party, then (barring complications), your action is *ceteris paribus* legitimate. Although what follows will not depend strongly

---

<sup>6</sup>Sorting out the details of this move is a thorny issue, and space doesn’t permit delving into it here. See, for instance, (DeRose, 2004) and (Richard, 1995) for views related to knowledge and belief ascriptions.

<sup>7</sup>There’s also a sense of ‘consent’ in which consenting is the stance itself — an attitude (a cousin of resignation, perhaps) that’s conventionally *indicated* by certain speech acts. Throughout, I’ll refrain from using ‘consent’ in this sense.

on any particular precisification of this idea, it may be helpful to have one on the table. To that end, one way of spelling this out is to characterize legitimacy in terms of *rights to object*: as a first pass, *S* has consented to *X*, in this sense, only if *S* has surrendered the right to object to *X*.<sup>8</sup> As I'll be using the term, if the outward form of consent-giving was followed, but (for whatever reason) the normative conclusion doesn't obtain, then consent was not really given.<sup>9</sup>

This much invites three points of clarification. First, I'm here relying on a distinction between the question of what consenting consists in, and that of what the semantic content of a consent statement is: what's *done* in uttering "I consent" is distinct from what that utterance *means* (just as with promising). We might put the relationship between the two by positing: the semantic content of a consent claim partly determines which acts of consenting are done in expressing the claim (with the appropriate force). Conversely, which acts of consenting are done — and, in particular, what their *objects* are — partly determines which consent claims are true.

Second, there is the possibility that a single utterance of a consent statement comprises multiple consent illocutions; so, roughly, we can say that consenting is *explicit* when it is done through the utterance of a consent statement whose content straightforwardly mirrors the object of consent, and *implicit* when it is done through uttering a consent statement whose content does not mirror its object. One can thus, in making a consent claim, perform both an explicit act of consenting and several distinct acts of implicitly consenting, all in the same breath.

Third, there are, as with any illocutionary act, issues of felicity: these will constrain both which utterances of consent statements constitute attempts at consenting, and which constitute successful acts of consenting. An account of such conditions might be further developed guided by a picture of consent that follows the contours

---

<sup>8</sup>One might compare this to promising: a promise made is, in part, a surrender of a right to later reconsider. If we understand promises as commitments, this is nicely consonant with Michael Bratman's view of the latter. See (Holton, 2003).

<sup>9</sup>My usage thus diverges from Joseph Raz's, discussed below. Raz separates the question of when consent is given from when it is *valid*. The question of validity, for Raz, is the question of whether a given act of consenting *in fact* changes a normative situation. This way of drawing a division, I think, risks obscuring more than it illuminates; in particular, it threatens to elide the possibility, to which I'll return, of *attempting but failing* to consent — the possibility of *misfires*.

of Grice's (1957) account of meaning: to consent is to perform a certain locutionary act with a certain attitude, with the expectation that hearers will believe that the speaker bears that attitude, and will believe this on the basis of their recognizing the utterer's intention that they come to believe it. The attitude here will involve a normative intention; Joseph Raz takes such a line:

Consent is given by any behavior (action or omission) undertaken in the belief (1) that it will change the normative situation of another; (2) that it will do so because it is undertaken with such a belief; and (3) that it will be understood by its observers to be of this character. (Raz, 1981, 119f)

Seana Shiffrin (2008, 500) takes a similar approach: "When an agent gives consent to another ... she transforms, in the way intended, the moral situation between the parties simply through the transmission of her will to do so." Although Raz's account is inadequate as it stands,<sup>10</sup> and Shiffrin's approach needs more elaboration, these characterizations are useful in clarifying just how consent utterances might fail to enact consent illocutions: how they might *misfire*. Attending to felicity conditions brings out two ways this might happen.

The first involves a class of felicity conditions for consenting that's primarily linguistic. For utterances of "I consent" to make acts of consenting, the speaker must be speaking English, not play-acting, and the like; a variety of ethically neutral conditions must be satisfied. In less explicit (or less utterance-like) cases, when the agent undertakes an action that, under the relevant governing conventions, is taken to express or entail consent, these conditions may be highly context-dependent. Indeed,

---

<sup>10</sup>Despite his insistence, for instance, that consenting and promising are different sorts of actions, there seems to be no room in his definition for making this difference out. In particular, his claim that there are promises that are not cases of consent — promises may, he claims, (putatively) confer "recipient rights", whereas consent may only (putatively) confer "rights of action" (Raz, 1981, 122) — seems straightforwardly false on his characterization. If I promise to pay you \$10, I, presumably, do so believing that it will change your normative situation, and that it will do so in virtue of that belief, and that you understand all this. Further differentiating consenting from promising, or from other similar speech acts intended to change normative situations (such as commandings, appointments, and marryings), requires, at minimum, further delimiting the sorts of normative effect involved.

in some contexts, participants may agree to drastically shift the felicity conditions, even to the point of making it the case that “I don’t consent” fails to count as an expression of dissent.<sup>11</sup> So here’s one way to attempt but fail to consent: to have all the beliefs (1–3), but be misguided about what the relevant communicative standards are.

The second involves the *normative* felicity conditions that bear on consent. In order to change a normative situation, I must be in a position to enact such a change. This has two pieces: there must be a *change*, and I must have the right *standing*. In order to make it permissible for you to enter my home, for instance, it must not have already been permissible for you to do so, and, moreover, I must have the power to grant such permission. When and whether such felicity conditions are satisfied is a question for ethical theory.

That said, we should *not* expect that every normative question around consent reduces to the question of whether consent is actually given. *Whether an agent consents* is (at least at first blush) a distinct issue from that of *whether the agent freely and knowingly consents*, and, for that matter, from that of *whether the agent’s consent matters for other moral questions*. Satisfaction of felicity conditions of the sorts just discussed determines the first question. Other facts — satisfaction of what might be called *auxiliary conditions* — determine the others.

One way of getting clearer on the distinction between normative felicity conditions and auxiliary conditions (so long as we’re willing to set aside scruples around metaphysical profligacy) is to ask just how consenting modifies the normative landscape. We might, following Shiffrin, think of it as a sort of dual of promising: just as promising is a way of generating obligations by communicating an intention to do so, consenting is a way of generating permissions by communicating an intention to do so.

Now this needs tweaking, for two reasons. First, it fails to respect the difference

---

<sup>11</sup>BDSM activists argue that “safe and sane” BDSM scenarios provide examples of this. Fowles (2008, 120) writes: “The safe, sane, and consensual BDSM landscape is made up of stringent rules and safe practices designed to protect the feelings of *everyone* involved, and to ensure *constant, enthusiastic consent*. The culture could not exist if this were not the case. . . . Before any ‘scene’ begins, the rules are made clear and the limitations agreed upon.”

between *consenting* and *giving permission*; the former, but not the latter, seems to be apt only when the object involves something over which the permitter has a normative claim. I may give you permission to punch yourself, but, unless I have some special stake in the state of your body, it seems strained to my ears to call this consenting (in contrast with a case in which I *do* have such a stake — if, for instance, I’m a photographer and you’re my model). Second, it fails to deliver the right results in cases of preclusion and preemption.<sup>12</sup> Suppose, for instance, that my roommate and I have a standing arrangement that either of us may permit a third party to enter our apartment. If, earlier today, my roommate consented for such a person, Jill, to enter, a permission fact has been generated: it’s permissible for Jill to enter. Since it’s already permissible for her to enter, I cannot now make it the case that it’s permissible. Nonetheless, I can consent to her entering; moreover, I can do so even if I’m aware that my roommate has already consented. Similarly, if my roommate and I have an arrangement that no one may enter without permission from *both* of us, and my roommate hasn’t yet given consent, I can consent to Jill’s entering without making anything permissible for her.

We might try to get around these problems by instead requiring that consenting consists in *guaranteeing* that certain permission facts obtain. Such a strategy, however, is unpromising, in light of the possibility of conditionally consenting (“I consent to Jill’s entering, but only if my roommate does”). Perhaps more importantly, this remedy fails to capture the special relationship my consent has to *me*: what I do when I consent to Jill’s entry seems to be less intimately tied to what’s permissible *tout court* than to what’s permissible relative to *my* claims.

This suggests another way of modifying the proposal. As a tentative hypothesis, we might say: *S* consents when she modifies the moral situation to generate relevant acceptability-to-*S* facts by communicating her intention to so modify the situation (and believing that this communication will so modify it). What might these *S*-relative facts be? The initial observations in this section provide one way of answering: they’re the sorts of facts that govern whether *S* has retained a right to object to a

---

<sup>12</sup>I’m sure this point has been made by someone else, but I can’t locate a source.

state of affairs.<sup>13</sup> So, as a toy model:

- (C) *S* consents to *X* when she forfeits a right to object in some situations in which *X* obtains, by communicating an intent to do so, on the basis of that communication.

With this in hand, we can restate our taxonomy of conditions. First, there are the linguistic felicity conditions whose satisfaction is required for the communicative element to hold. Second, there are the normative felicity conditions whose satisfaction is required in order that such communication constitutes a surrender of a right to object. Finally, there are the auxiliary conditions that determine how a subject's surrender of a right to object matters for broader ethical questions. It's worth emphasizing here that the boundary between normative felicity conditions and auxiliary conditions will depend upon the details of the larger ethical framework; as we'll see, there are views on which having access to information is required for a right to object to be abandoned — in which case the question of information access will count as a normative felicity condition — and views on which it's not — in which case, such a condition will count as auxiliary.

## 1.4 Considerations

To return to the main project, why should we think consent *is* contrastive in the sense I'm proposing? After all, unlike 'prefers', 'consents' is more often than not used in a superficially dyadic form; it's relatively rare to explicitly mention the alternatives when ascribing consent. In this respect, 'consents' is more like 'explains', 'knows', 'is a reason to', and 'ought to'. Yet philosophers have (albeit contentiously) argued that these are contrastive, too.

There's a sense in which this picture of consent is quite natural; after all, consenting seems a close cousin to preferring. Consent is often the product of a *deliberative*

---

<sup>13</sup>Alternatively, we might avoid the language of *rights* by instead focusing upon whether *S* has undertaken a commitment to refrain from objecting. This would obviously need further development, which I won't pursue here.

*process*. When we deliberate, however, we don't do so panoptically, surveying all of logical space. Rather, we consider a range of *options*. It's no secret that how we frame our options can make all the difference in how we deliberate.<sup>14</sup> The contrastive picture of consent builds this dependence in at the ground level.

There are two more sorts of consideration that speak in favor of contrastivity. First are facts about linguistic use; consent statements display some of the aspects of more overtly contrastive relations, along with some apparent paradoxes resolvable with the aid of the contrastive hypothesis. Second — and, to my mind, more important — is the new range of possibilities that machinery of contrastivity offers for articulating critical analyses of consent-based ethics and political philosophy. I'll address the first of these in the remainder of this section, and the second in the next.

### 1.4.1 Surface Evidence

Jonathan Schaffer (2004, 77ff) points out that contrastivity can sometimes be hinted at by surface-level linguistic phenomena. The most obvious of these is the availability of explicit contrastive formulations. “One reason,” Schaffer (2004, 78) writes, “for thinking that ‘prefers’ denotes a ternary relation, despite the existence of binary preference ascriptions such as ‘Ann prefers chocolate’, is the existence of overtly contrastive preference ascriptions such as ‘Ann prefers chocolate rather than vanilla.’” And we can notice a similar (though admittedly much less commonplace) phenomenon at work in consent ascriptions: “Jane consented to paying the fine rather than going to jail.”

Now this observation isn't conclusive. As Jonathan Snedegar (2012, §1) points out, there is the obvious possibility of ambiguity between contrastive and non-contrastive relations. But other considerations weigh against this possibility. Snedegar, considering the contrastivity of ‘reason’, considers the sentence “The fact that you sprained your ankle is a reason to wear your brace, and to lift weights rather than run.” Following Schaffer, he notes

---

<sup>14</sup>See, *e.g.*, the classic (Tversky and Kahneman, 1981).



There is just one occurrence of ‘reason’ here; this suggests that ‘reason’ latches on to both conjuncts, ‘to wear your brace’ and ‘to lift weights rather than run’ . . . If the single occurrence of ‘reason’ latches on to both conjuncts, that is evidence that ‘reason’ is univocal, whether it takes a contrastive or a non-contrastive complement.

Just so for ‘consent’: “Jane consented to paying the fine rather than going to jail, and to being searched” is perfectly appropriate. So there’s *prima facie* evidence that ‘consent’ is not ambiguous between a contrastive and a non-contrastive relation, and *prima facie* evidence that it is contrastive.

A second feature noted by Schaffer is focus sensitivity. As Fred Dretske (1972) observed, differences between statements due to shifts in tonal focus (or emphasis) can often be explicitly drawn out through contrast clauses. For example, we might articulate the difference between “Why did Fred sell his typewriter to *Clyde*” and “Why did Fred *sell* his typewriter to Clyde?” by paraphrasing the sentences, respectively, as “Why did Fred sell his typewriter to Clyde, rather than someone else?” and “Why did Fred sell, rather than give, his typewriter to Clyde?”. When such focus shifts lead to semantic differences, underlying contrastivity provides an explanation of this. And consent statements *do* display what seem to be semantic focus sensitivities: I can consent to be driven *home* without consenting to be *driven* home — when, for instance, the destination, but not the means of transportation, is under discussion.

### 1.4.2 Paradoxes

Over and above these superficial observations, the contrastive strategy is appealing for the work it can do in explaining (or explaining away) some puzzles that naturally arise for consent statements. Onora O’Neill describes some of them:

the boundaries of explicit consent are unclear. Like other propositional attitudes, consent is opaque. Consent may not extend to the logical implications, likely results, or the indispensable presuppositions of that which is explicitly consented to. (O’Neill, 1985, 255)

There are, I take it, two distinct sorts of puzzle here. One set of puzzles consists of what Schaffer calls “surface paradoxes”: how can we explain the fact that apparently contradictory consent claims sometimes seem true? The other consists of failures of seemingly plausible inference rules, particularly principles of closure under entailment and predictable results. I’ll address the first of these here, turning to the second in §1.4.3.

Contrastivity provides a general strategy for attacking such problems: the contrastivist can defuse apparent contradictions by locating a difference in the (positive) *content* of the claims, hidden in a covert contrast class. To see how this goes, it’s worth considering contrastive approaches to other concepts. Frank Jackson (1985), for instance, suggests that “ought” sentences should be understood as contrastive.<sup>15</sup> His apparatus allows Jackson to explain what’s going on in cases in which we seem to endorse contradictory ought-claims. He gives an example in dialogue:

I say ‘It ought to be that Lucretia used less painful poisons’. You retort ‘Oh no, it ought to be that Lucretia used painless poisons’. I then retort ‘Oh no, it ought to be that Lucretia used political means rather than poison to obtain her ends’. You then retort ‘Oh no, it ought to be that Lucretia never existed at all’. I then retort ‘Oh no, it ought to be that Lucretia existed but made people happy’. And so on and so forth. Each retort seems a fair one, how so? (Jackson, 1985, 181)

On the face of it, it looks as though we have, at each step in the dialogue, endorsed claims that flatly deny those in the preceding stages. What’s happening here, according to Jackson, is that the set of alternatives is *shifting*: with each new claim, new alternatives are introduced, relative to which the prior claims come out false.

Jonathan Schaffer adopts a similar strategy to diagnose disagreements around knowledge. In cases of skeptical challenge, for instance, we might understand our inclination to first accept, and then reject, knowledge ascriptions by reference to a change in contrast class:

---

<sup>15</sup>Also see (Snedegar, 2013, 124ff) for an indirect argument from this to the conclusion that the deontic permissive “may” is contrastive.

Contrastivism thus provides a straightforward semantical explanation for why ‘I know that it is 3:15’ is ordinarily acceptable on the basis of a glance at the clock, but not when the accuracy of the clock is in question. The explanation is that in the ordinary case the contrast does not include the clock-error possibility, while the introduction of the question of accuracy shifts the implicit contrast to include such. (Schaffer, 2004, 81)

Again, noticing that contrast classes change between utterances allows the contrastivist to explain both why we’re apt to accept each utterance when it’s uttered — because it’s *true* relative to the contrast class in play at that stage in the conversation — and why an earlier utterance strikes us as unacceptable once its sequel is uttered — because it’s *false* relative to the new contrast class.<sup>16</sup>

Parallel cases for consent claims are not hard to come by. Here are two:

*Constraint.* Lana chooses to enter into a prolonged sexual relationship with a man to protect herself against sexual and other violence from other men; she doesn’t enjoy the sex or the man’s company, but it’s a tradeoff she makes willingly.<sup>17</sup> There are conflicting intuitions about whether she consents: one is tempted to say both that she consents (she knowingly chooses) and that she doesn’t (because the constraints undermine her consent). Does she?

*Nondisclosure.* John is a cancer patient, whose doctor gives him the options of radiation therapy and doing nothing. Given those options, John agrees to painful radiation therapy. Ask John whether he consents to radiation treatment, and he’ll say yes. John later learns that there was a third, surgical, treatment available, as effective as radiation and far less

---

<sup>16</sup>There is, of course, an issue here with explaining just how disagreement or denial could arise; if the earlier accepted statement differs in content from what’s later denied, then the two aren’t contradictories. John MacFarlane’s (2005) approach to relative truth makes use of both the original context of utterance and the later context of evaluation to answer this challenge. As Sally Haslanger (2007) notes, this approach can be fruitfully applied to cases of (actual or apparent) disagreement, particularly around social facts. I suspect that such an approach might also be useful in analyzing higher-order questions around consent (“Did John believe that Jack consented?”), but won’t pursue that here.

<sup>17</sup>This example is due to Robin West (2007).

painful. Ask him now whether he consented to radiation, and it seems he'd be justified in responding, "No: I didn't know all the options." Which of his answers is correct?

Note that in these cases, the standard approach demands a single answer. It must treat as error the intuition that there's a sense in which John and Lana both *consent* and *don't consent*; this is a cost.

The contrastive strategy offered by Jackson and Schaffer, on the other hand, gives a way of respecting all the relevant consent claims: each of them is (arguably) true, relative to the appropriate contrasts. The first consent claim in each case arises in a context presupposing certain background constraints, treating other situations (in which avoiding violence isn't reason to take on a sex partner, or better treatments are available) as irrelevant; the contrast class doesn't include these possibilities. The second claims bear expanded contrast classes. Contrastively: Lana consented to being in a sexual relationship with the man, rather than being subject to violence from other men; she didn't consent to being in a sexual relationship with him, rather than either being subject to violence or living in a society in which unpartnered women aren't at risk. Contrastively: John consented to undergoing radiation, rather than not being treated; he didn't consent to undergoing radiation, rather than not being treated or getting surgery.

There's a closely related way that apparent contradictions arise, in the realm of reports: there may be one contrast set that's at play in an original utterance, and another at play in a report of that utterance. Suppose  $S$  utters "I prefer  $X$ " at some time  $t$ . Further suppose that the context of utterance at the event determines, for  $X$ , a contrast set  $\{X_i\}$ . Then the content of  $S$ 's utterance is:  $S$  (at  $t$ )<sup>18</sup> prefers  $X$  rather than  $\{X_i\}$ . Now suppose we're later called upon to evaluate  $S$ 's utterance, at some time  $t' > t$ , and that our context determines some contrast set  $\{X'_i\}$  distinct from  $\{X_i\}$ , and that  $S$  does *not* prefer  $X$  rather than  $\{X'_i\}$ . A witness reports: " $S$  said

---

<sup>18</sup>I'll suppress the time-dependence in what follows. Clearly, it's necessary to allow for *shifts* in  $S$ 's preferences (even relative to the same alternative) between the original utterance and the time of a report. This is, however, an orthogonal complication. Likewise, I'll ignore complications arising from the oddities of *de se* ascriptions.

that she preferred  $X$ ". What's the content of this assertion? On the most natural reading, it's that  $S$  said that  $S$  prefers  $X$  rather than  $\{X'_i\}$  — *not* rather than  $\{X_i\}$ . The imbedded claim in the witness's report, then, has a different content from  $S$ 's original statement. It's true that  $S$  said "I prefer  $X$ ", and it's true that  $S$  preferred  $X$  rather than  $\{X_i\}$ . It's false, however, that  $S$  preferred  $X$  rather than  $\{X'_i\}$ , and so, just as the sentence " $S$  preferred  $X$ " uttered at  $t'$  is false, so is the sentence " $S$  said that she preferred  $X$ " uttered at  $t'$ .<sup>19</sup> This suggests a strategy for resolving another apparent paradox:

*Reports.* Tom, a police agent, suspects that Alice, a student, is selling illegal drugs out of her apartment. Because his evidence is scant, Tom can't obtain a warrant. So he visits Alice's apartment, knocks on the door, and, when Alice answers, tells her that he has reason to believe a bomb has been planted in her apartment, and that he needs to sweep for it. Alice explicitly consents to a search, and Tom finds drugs. Later, when Alice challenges the search, Tom claims: "Alice said that she consented to a search. In fact, she explicitly said, 'I consent to a search'". Alice claims, apparently correctly, that the latter statement is true, and the former false. How can this be?<sup>20</sup>

The contrastive strategy gives Alice the tools to characterize what went wrong: although she uttered "I consent" — and thereby consented to *something* — it doesn't follow that *what she consented to* is what's relevant in the context of evaluation. It would seem perfectly natural for her to say: "Yes, I consented to his searching my home for bombs; but I didn't consent to his searching my home for drugs." Plausibly,

---

<sup>19</sup>This sort of phenomenon has a familiar analogue in quotational contexts involving indexicals. Imagine Ursula, who, standing at the gates of the Forbidden City, exclaims, "This is the most beautiful place I've ever been!". If she and I later discuss this comment in downtown Boston, I could accurately say "You said, 'this is the most beautiful place I've ever been!'". I could not, however, accurately say "You said that this was the most beautiful place you've ever been" (or, for that matter, "You said that this was the most beautiful place I've ever been."). The indexicals "this" and "I" in the indirect quotation have different referents (*viz.*, downtown Boston and me) in the context in which I'm reporting Ursula's speech than they had (*viz.*, the Forbidden City and Ursula) in the context in which Ursula uttered them.

<sup>20</sup>For a similar real-life case, see (*United States v. Harrison*, 2011).

these two consent statements would involve quite different contrast classes. And her consenting to a search from among {the police search (for a bomb), risking death} isn't *prima facie* equivalent to her consenting to a search from among {the police search (for drugs), the police don't search}. Alice's characterization, understood this way, seems entirely plausible.

To be clear, the important claim here is not that the contrastive strategy *resolves* the questions in these ways. Though the suggested resolution is plausible, I think, it's not *mandated* by the contrastive thesis alone; whether John, Lana, and Alice consented, relative to the appropriate contrast classes, is a substantive question that turns on the details of the analysis of consent. Instead, the crucial point is that the contrastive approach doesn't *preclude* the sort of answer just given; it offers a way of explaining how our ordinary intuitions could be accurate, despite their seeming contradictory. In contrast, the best the standard theory can do is offer an error theory that relies upon theses external to the analysis of consent proper. This benefit is primarily one of philosophy of language: the contrastive strategy gives a more straightforward way of resolving apparent paradoxes than does the alternative.

### 1.4.3 Logical Questions

A summary overview of examples gives a grim view of the prospects for uncovering interesting logical relations between consent claims. Martha can consent to being driven home by her girlfriend (who's drunk), but not to being driven home by a drunk driver. She can consent to paying a fine and (as a result) not going to jail, but not consent to paying a fine. Conversely, she might consent to having her hair dyed and consent to having her face painted but not consent to having her hair dyed and her face painted (which would be overkill). She may consent to being paid, without consenting to living in a money economy. The obvious inference rules — closure under entailment, closure under conjunction, extension to presupposition — all seem to admit of clear counterexamples.

This is particularly vexing because there seem to be times when the most natural way of explaining why a consent statement is true would be in terms of such a prin-

principle. When I consent to be caressed, for instance, it seems that I thereby consent to be touched. Why? It'd be nice to answer: because being caressed *entails* being touched.<sup>21</sup>

Contrastivity offers the possibility of addressing some of these issues. On the one hand, it can explain why failures of inference rules might appear. On the other hand, it can explain why this shouldn't distress someone looking for substantive connections between consent claims.

To see how, it's worth returning to Jackson's (1985, 184f) logic of obligation. On his semantics, "It ought to be that  $A$ " — or  $O(A)$  — is to be evaluated relative to a set of alternatives (propositions)  $\{A_i\}$ , and is true just in case, for each  $i$  with  $A \neq A_i$ , the world that would be the case if  $A$  were true is better than the world that would be the case were  $A_i$  true. This is in contradistinction to the standard modal semantics for "ought", which takes "ought" to be a sentential operator such that (in the simplest version) "It ought to be that  $A$ " is true just in case  $A$  is true in all the best possible worlds accessible from the actual world.

One puzzle he notes centers around *entailment*: on the standard deontic semantics, if  $O(A)$  is true, and  $A$  entails  $B$ , then  $O(B)$  is true: if  $A$  is true in the best possible world, and  $B$  is true whenever  $A$  is, then  $B$  is true in the best possible world. Yet, he notes, there seem to be counterexamples to this, in the form of "Good Samaritan" cases. Imagine that, on my way to work, I notice a wounded man on the side of the road; it ought to be the case that I help the wounded man. Yet I cannot help a wounded man without there *being* a wounded man. So there ought to be a wounded man on the side of the road.

Clearly something has gone awry. Jackson's alternative semantics allows him to say what that is: the claim that I ought to help the wounded man is made relative to a set of alternatives  $\{\text{There's a wounded man and I help him, There's a wounded man and I don't help}\}$ , and the first is the best option. The claim that there ought to be a wounded man, on the other hand, is made relative to a *different* contrast class:

---

<sup>21</sup>Or, perhaps, because being caressed is a *way* of being touched. Thanks to Sally Haslanger for this point.

{There's a wounded man, There's no wounded man}.

Similar remarks apply to consent cases. Martha consents to being driven home by Jane (who's drunk), but not by a drunk driver. This can be so because different descriptions bring with them different contrast classes. Ask Martha, "Did you consent to Jane's driving you home?" and you invoke a contrast class in which the relevant alternatives are being driven home by somebody else, or not being driven home — all the (relevant) ways of Jane's driving Martha home count as essentially the same. Ask her, "Did you consent to Jane's driving you home while she was drunk?" and you invoke a different space; now the relevant alternatives include being driven home by Jane while she was sober. The two questions involve distinct (covert) contrast classes — and so their corresponding consent claims are distinct.

Shifts in contrast class may thus explain why the entailment rule suffers from a kind of invalidity. But this analysis also shows that it's only a *kind* of invalidity. That is to say, what we have in this case is a failure of *the proposition expressed by an utterance* of  $O(A)$  to entail *the proposition expressed by an utterance* of  $O(B)$ , despite the *the proposition expressed by an utterance* of  $A$  entailing *the proposition expressed by an utterance* of  $B$ . There is what we might call invalidity of an "utterance rule". What this *doesn't* demonstrate immediately is invalidity of a *propositional* inference rule; in fact, the inference rule as stated doesn't even make sense as a propositional rule, because  $O(A)$  and  $O(B)$  aren't propositions (they're sentences), and don't unambiguously correspond to propositions (they need context to fill their contrast sets before they do).

Put another way: by asking about entailment relations between the sentences  $O(A)$  and  $O(B)$ , we're asking the wrong question; we should instead be asking about relations between the corresponding propositions  $O(A; \{A_i\})$  and  $O(B; \{B_i\})$ . And we should, on Jackson's semantics, expect *some* logical relations to hold. As an easy example, if each  $A_i$  is logically equivalent to each  $B_i$  (and  $A$  to  $B$ ), then  $O(A; \{A_i\})$  is true if and only if  $O(B; \{B_i\})$ .

To apply this to consent, it's true that there are counterexamples to naïve consent inferences. But that *doesn't* mean that there are no systematic logical relations gov-



erning the concepts at hand. It just means we've been looking in the wrong place. We shouldn't ask, "when does consenting to  $X$  entail consenting to  $Y$ ?" but rather "when does consenting to  $X$  rather than  $\{X_i\}$  entail consenting to  $Y$  rather than  $\{Y_i\}$ ?"

The answers to such questions will, of course, depend upon a substantive analysis of consent. But we can get a grip on how they might go by returning to the toy model (C) presented in §1.3. On (C), consenting to  $X$  is forfeiting a right to object in some situations. But in which situations? First pass: in all situations in which  $X$  obtains; if I consent to being driven to New York from Cambridge (in an ordinary context), I forfeit a right to object in situations in which I'm driven to New York from Cambridge. But this can't be quite right. On the one hand, there are certainly *ways* of being driven to New York from Cambridge — *via* Seattle, for example — that I can still appropriately object to. On the other hand, I might be thought to have forfeited a right to object in some situations in which I'm *not* driven to New York from Cambridge — when, for instance, the car breaks down in New Haven. This is, applied to consent claims, the "underdetermination problem" discussed by Delia Graff Fara (2012). Put crudely, the problem is that it seems impossible to spell out what's done when someone consents using only a relationship between the agent and state of affairs named in the consent claim.

What's needed, it seems, to connect a consent claim with an act of consenting is an extra parameter: some additional information to help draw the line between situations in which a right to object is retained and those in which it's forfeited. But on the contrastive thesis, we get just such a parameter: the content of a consent claim includes both the state of affairs named and the set of contrasts. So we might modify (C):

(C\*)  $S$  consents to  $X$  rather than  $\{X_i\}$  when she forfeits a right to object in some situations  $F(X; \{X_i\})$ , by communicating an intent to do so, on the basis of such communication.

What needs clarification here is the question of *which* situations  $S$ 's forfeiture applies to: what situations are in  $F(X; \{X_i\})$ ? The availability of the contrast space gives

resources for answering this. In particular, it seems natural to think: what the contrasting alternatives in a consent claim do is to provide exemplars of situations in which a right to object *isn't* forfeited in making the claim. To extend this to the more general question, we might try following Garfinkel's strategy of taking the contrasts to specify both which situations are excluded from consideration, and how the remainder of situations are partitioned.

Such an approach might propose that the line between the situations  $F(X; \{X_i\})$  implicated in the consent claim and those not can be drawn based upon comparative similarity between paradigm situations (named by 'X' and 'X<sub>i</sub>') along some salient lines:<sup>22</sup> a situation is in  $F(X; \{X_i\})$  just in case it's *sufficiently*<sup>23</sup> similar to X, and *more* similar to X than it is to any  $X_i \neq X$ .

On this model, we can give an answer to the entailment question: S's consenting to X rather than  $\{X_i\}$  entails S's consenting to Y rather than  $\{Y_i\}$  if (a)  $F(Y; \{Y_i\}) \subset F(X; \{X_i\})$  — so that S's having forfeited a right to object in any situation in  $F(X; \{X_i\})$  guarantees her forfeiture of such a right in any situation in  $F(Y; \{Y_i\})$  — and (b) her communication of an intent to forfeit such a right for the  $F(X; \{X_i\})$  situations is a communication of an intent to forfeit such a right for the  $F(Y; \{Y_i\})$  situations. If we may suppose that (b) is satisfied whenever (a) is, and apply the similarity analysis, then this sufficiency criterion reduces to: S's consenting to X rather than  $\{X_i\}$  entails S's consenting to Y rather than  $\{Y_i\}$  if every situation that is both sufficiently similar to Y and more similar to Y than to any  $Y_i \neq Y$  is sufficiently similar to X and more similar to X than to any  $X_i \neq X$ . *Et voilà: a bona fide* entailment relation!

How might this play out in a sample case? Consider my consenting to being caressed. Does consenting to being caressed, rather than left alone, by saying “I consent to being caressed, rather than left alone”, entail my consenting to being touched? Or, put another way, by explicitly consenting to being caressed rather than

---

<sup>22</sup>Of course, to defend this account fully, I'd need to say a great deal more about what determines which situations are paradigm, what such similarity relations amount to, and how these determined by contextual factors. But, since my purpose here is illustrative, I won't explore this further.

<sup>23</sup>Ideally, sufficiency could be spelled out in terms of the relative degrees of similarity between the alternatives.

left alone, am I thereby implicitly consenting to being touched rather than left alone? Answer: *yes* for some ways of being touched, but not for every way of being touched. For those ways of being touched that are relevantly similar to being caressed (say, an attempted caress that turns into a scratch), it seems right to say my consent applies; for those ways of being touched that aren't at all similar to being caressed (say, a punch), it seems right to say my consent doesn't. And, given the right similarity relation, this is exactly what the analysis delivers.

But we might worry about this case: won't consent cover too much? What about, say, a massage? It seems that being massaged is more similar to being caressed than to being left alone — but it also seems that I don't consent to a massage by consenting to being caressed. A defender of this approach has two ways out. First, she might insist on the distinction between our intuitive judgments about situational similarity and the technical notion of similarity at play in the analysis; indeed, she might follow a Lewisian strategy of using such test cases as a way of building constraints on such a notion. Or instead, she might note that a stronger consent claim — consenting to being caressed rather than being massaged or left alone — won't, even on an intuitive notion of similarity, entail my consent to being massaged rather than left alone (on the basis of this principle); she might then argue that pragmatic factors determine that my utterance of “I consent to being caressed rather than being left alone” leaves tacit some of the contrasts. Pursuing this in more depth would take us far too far afield; nonetheless, the availability of such responses on behalf of this (toy, preliminary, and merely illustrative!) analysis should be enough to convince us that it's altogether plausible that there *are* logical relations to be found among consent claims, discoverable once a (serious, developed, and well-defended!) analysis of consent is at hand.

## 1.5 Reclaiming Consent

Over and above these linguistic and logical virtues of the contrastive approach, the strategies outlined offer a boon to ethical and political theorizing. Because the stan-

dard approach, as I've noted, is committed to an unambiguous, context-insensitive, answer to the question of whether an agent consented, theorists around consent have often found themselves facing a tangle of conflicting intuitions when analyzing the relationship between consent and the permissibility or legitimacy of situations that follow. The trouble arises in a myriad of cases of consent-gone-wrong: situations involving coercion, constraint, ignorance, and unforeseen consequences. How can we explain how badness or injustice arises in these cases, in the face of apparent consent?

### 1.5.1 The Bog

The dominant way of answering this is to claim that consent, by itself, counts for very little. According to this view, the agents in such situations *do* consent, but their consent fails to legitimize what follows. This can be so, the story goes, because although, strictly speaking, the subjects consent, their consent isn't *full* — it's normatively defective — due to the conditions in which it's given. To put this in terms of the taxonomy in §1.3, this approach locates the defect in the question of whether *auxiliary conditions* for consent's moral force are in place, rather than in the question of whether *felicity conditions* for consent are satisfied.

In medical ethics, for instance, the requirement that patients give informed consent to treatment is usually bifurcated into the demand that patients be *informed* both of the range of treatment options available to them and of the relative risks of the options, and the demand that they *consent* to the treatments they receive. This factorization has, as Jay Katz (1977) argues, generated something of a legal morass. For the purposes of legal remedies, it matters whether a failure on the part of a treating physician to fulfil the informational component counts as failing to receive consent for treatment. Treating without consent falls under the regime of *battery*, while failure to fulfil other standards falls under *negligence*. As Katz points out, the first is understood as an affront to dignity, and so counts as a harm in itself; the latter, however, only applies when there is some (other) harm. By taking an easy-consent approach, then, the law tends to preclude malpractice claims against physicians who intentionally misinform (but do not otherwise harm) patients. This

result is questionable, at best; there is, I think, a strong intuition that extreme cases of misinformation (imagine a doctor who tells a perfectly patient she'll die without surgery, simply because the doctor wants a little practice) *do* seem to undermine consent.

Analogous questions arise in other realms. In the (United States) law governing warrantless search, for instance, the question of whether consent was given *voluntarily* controls that of whether evidence obtained through such a search is admissible. In cases like Alice's, a court must decide whether the level of "deception or trickery" used by police agents is enough to render the consent coerced (*United States v. Harrison*, 2011, 8). Cases of trickery, deception, constraint, and coercion, then, must be resolved by appeal to general principles about *how much* trickery, deception, constraint, and coercion it takes to render consent involuntary. The coherence of such a question is, to my mind, dubious; and, unsurprisingly, it's proven exceedingly difficult for courts and legal scholars to agree on any such general principles, or to articulate the relationships between these undermining factors.

In political theory, similar worries crop up in relation to Lockean and Hobbesian views that root political obligation in consent of the governed. This is a point that's been recognized since at least the time of Hume:

It is in vain to say, that all governments are, or should be, at first, founded on popular consent, as much as the necessity of human affairs will admit . . . I maintain, that human affairs will never admit of this consent, seldom of the appearance of it; but that conquest or usurpation, that is, in plain terms, force, by dissolving the ancient governments, is the origin of almost all the new ones which were ever established in the world. And that in the few cases where consent may seem to have taken place, it was commonly so irregular, so confined, or so much intermixed either with fraud or violence, that it cannot have any great authority. (Hume, 1994, 191f)

Those committed to the conclusion that government can be just, then, must either

abandon consent as a source of justification, or turn to a highly altered notion.<sup>24</sup> Thus we arrive at theories such as Rawls's (1996), making use the "hypothetical consent" of a carefully circumscribed class of people.

In addition to the conceptual chaos the auxiliary-conditions approach fosters, it brings with it a more serious cost when it comes to explaining why consent matters. We might have hoped that consent should pull substantial explanatory weight; that, for instance, a relationship between consent and autonomy would explain why things done to us to which we consent should be thought of as (*pro tanto*) permissible, and those to which we do not should not. On the approach we've been considering, however, the bulk of the work is done by auxiliary factors; consent itself *doesn't* legitimize. Consent becomes such a lightweight concept that there's little hope that considerations around consent *simpliciter* could ground any substantial moral theses.

The alternative for the classical theorist is to claim that, in these cases, consent isn't given at all: that for an agent to consent, she must be — to some minimal degree — informed, undeceived, free from coercion or constraint, aware of potential consequences, and so on. This move *builds in* such substantive requirements into the notion of consent itself, by identifying a broad array of normative felicity conditions that must be satisfied for a speech act to count as an act of consenting. But the theorist taking this approach faces a serious challenge when it comes to deciding just how strong the felicity conditions on consent are. Too weak, and consent becomes too easy to come by, leaving us to lean again on auxiliary conditions to buffer against unpalatable conclusions. Too strong, and consenting becomes an ideal far removed from actual practice; if, for instance, consent requires full awareness of the downstream consequences, it will virtually never be the case that anyone consents. Needless to say, finding a middle ground between these poles has proven nigh-impossible.

---

<sup>24</sup>Of course, those who welcome the conclusion that coercive government cannot be justified may take such observations, together with a commitment to the moral necessity of consent, to demonstrate their conclusion. Thus, *e.g.*, Bakunin (2005): "a State without slavery, open or concealed, is inconceivable: that is why we are enemies of the State."

## 1.5.2 A Way Out

Of course, this is not a conclusive reason for thinking that the classical approach is unworkable; thankfully, long histories of failure count for very little in philosophy. Nonetheless, a mess in classical theory does heighten the appeal of alternatives.

What I want to suggest is that the classical theorist of consent gets off on the wrong foot. As a result, whichever way she resolves the consent-gone-wrong cases — either by identifying substantive normative felicity conditions and claiming that these aren't met, or by identifying defects in auxiliary features of the situation and claiming that these undermine the legitimizing effect of the consent — her analysis must focus on a maze of external considerations. As Bas van Fraassen puts it regarding explanation:

The discussion of explanation went wrong at the very beginning when explanation was conceived of as a relationship like description: a relation between theory and fact. Really it is a three-term relation, between theory, fact, and context. No wonder that no single relation between theory and fact ever managed to fit more than a few examples! (van Fraassen, 1980, 156)

With the contrastive picture in mind, however, there's room for a way of addressing such cases that avoids this path. Once we recognize the frame-dependence of consent, we're in a position to consider a new possibility: that what's going on when consent doesn't legitimize is not, or at least not always, that felicity or auxiliary conditions fail to be met, but rather that consent that *is* given isn't consent *to the right thing*.

Consider again *Reports*: Alice uttered "I consent to a search"; in so doing she consented to a search for a bomb, rather than risking death — but, she claims, not to a search for drugs, rather than being left alone. Plausibly, it's the latter, and not the former, that matters for whether the search was legitimate, since it's the latter alternative set, and not the former, that's in play when, in the context of a court proceeding, the question of whether she consented to a search arises. So the right way to frame the question of whether the search was legitimate is *not* to ask, as courts have done, whether the trickery involved rose to the level of coercion, thus

undermining consent, but rather to ask whether the consent she evidently gave entails that she forfeited the right to object to the search that took place.

Or consider *Nondisclosure*. If his physician obtains John's consent to treatment after failing to inform him of all the relevant options, we might say that she hasn't committed simple battery (since John consented to being treated over not being treated), but has nonetheless done something more serious than simply failing to follow good practice. The doctor has failed to obtain an ethically crucial *kind* of consent: consent to treatment from all the relevant options.

None of this is to say that considerations of coercion, deception, and constraint are irrelevant to inquiry around consent. Indeed, as we've seen, they may enter in at least three ways: they may play a role in determining the contrast class for a consent claim; they may defeat the felicity of a consent claim; or they may enter into the analysis of more complicated moral questions involving consent. The point is rather that the contrastive analysis both illuminates these different ways and provides a new starting point for addressing the questions. And, just as importantly, it gives a way to reclaim the centrality of the question "was consent given?" to the analyses. The theoretical space thus afforded is, I take it, another point in favor of the contrastive analysis.

## 1.6 Conclusion: Radical Theory, Radical Consent

In the preceding sections, I've tried to make the case that there are good philosophical reasons to think of consent as a contrastive relation, by considering linguistic evidence, the potential for substantive logical relations, and the possibility of more fruitfully tackling old problems in the ethics of consent. To conclude, I want to turn to the question of how this observation can prove useful for revolutionary and critical theory.

Historically, consent has been at the core of a tremendous amount of radical (and particularly anarchist) intuition. Going back to Godwin, Proudhon and Bakunin, one of the most pervasive critiques of statism relies upon consent: people who live



under the tyranny of the state never consented to it.<sup>25</sup> A related critique has it that *autonomy* is the key concept; Robert Paul Wolff, for instance, has argued that the state unavoidably undermines individual autonomy, and should thus be opposed on Kantian grounds (Wolff, 1998).

At the same time, however, consent seems to be deployable — and, increasingly, actually deployed — for counterrevolutionary purposes. People whose commutes are made longer by direct actions don't consent to being so affected. Similarly, I want the downfall of industrial capitalism; I should hope that my politics doesn't require that I first convince everyone who's in the grip of the capitalist mythos that they should consent to its overthrow before I pursue it. A recent zine from the CrimethInc. Ex-Workers' Collective voices a similar sentiment:

Observe how an anxious liberal from our local Occupy movement, dismayed by an illegal building occupation undertaken by autonomous occupiers, strives to distance the Occupy group from the occupation. . . . This liberal wishes to communicate that the building occupation felt like a violation of his consent. Why? Because it was related to a current in which he felt invested, yet he had not been invited to participate in decision-making, and it involved actions that he personally disdained. Of course, we undertook the occupation autonomously precisely for that reason: we knew we could never achieve consensus in the public general assemblies to do something that so dramatically challenged consensus reality. Whether or not the occupation hurt anyone was beside the point: its “violence” had less to do with its literal effects than its challenge to consensus reality. To him, such a challenge constituted a violation of collective consent. (CrimethInc. Ex-Workers' Collective, 2012)

More generally, misplaced concern over consent threatens to undermine radical tactics; if we count this sort of complaint as (always) grounds for concern, we run the risk of erring on the side of inefficacy.

---

<sup>25</sup>For a whirlwind tour of such considerations, see (McLaughlin, 2007, ch. 4).

Feminists, too, have worried about the usefulness of consent as a critical tool. In particular, some argue that treating even explicit, actual consent as legitimizing runs the risk of condoning subjection, because explicit, actual consent is all too easy to come by in an oppressive world. If we take the possibility of internalized oppression seriously, we may doubt whether a woman who consents to being treated as inferior (because she believes she *is* inferior) gives any justification at all to her mistreatment.<sup>26</sup> Along similar lines, Jody Freeman writes that

The liberal individualist understanding of consent is an example of how ideology affects consciousness. Much of women’s misery could be thought of as voluntary. To the extent that women marry, seek employment, and engage in heterosexual sex, they expose themselves to abuse, sexual harassment, and male expectations of their sexuality, all of which are often pleasurable for men and painful for women. And yet, to some extent all of these social arrangements are consensual. Liberal theory does not reflect the fact that choice and consent mean different things for men and women and that women’s “choices” usually reinforce male power. (Freeman, 1989, 100)

It seems that basing a political theory on consent (or autonomy) doesn’t do enough to guarantee justice. Consent is preconditioned by structure; structure determines, at least in part, what things we desire, and, perhaps more importantly, what things it’s possible for us to desire. Given that our social world is racist, patriarchal, capitalist, ableist, heterosexist, transphobic, and exploitative — given, that is, that we are mired in structural injustice — why should we think that our situated consent counts for much of anything?

Trying to evade this worry by averting to counterfactual or hypothetical consent seems unpromising. On the one hand, feminist insight into the pervasiveness of structure and ideology casts doubt on the coherence of this move; if, in particular, we understand individuals as partly constituted by their social relations, abstracting

---

<sup>26</sup>See, *e.g.*, the discussion of Susan Babbitt’s work in (Brennan, 1999, 878f).

away all the core features of *our* social relations means theorizing about creatures that differ crucially from ourselves: it's deeply unclear what, if anything, consent of alien counterparts in far-flung worlds has to do with that of actual embodied subjects, much less why we should take it as a guide to theorizing justice. Nor is it clear that such ideal situations *are* particularly ideal, given their elision of social sources of value in favor of individual attributes and desires.<sup>27</sup>

More troubling still is the possibility that views that center around highly modified notions of consent may covertly offer justification for oppression. Even if sense can be made of consent in a state of nature, for instance, some theorists have argued that oppressive social structures can arise from the iteration of the development of structure with free consent (see, *e.g.* Pateman, 1988, ch. 1). Similarly, Marilyn Friedman (2003, 173) worries that one “strand of thought in the liberal tradition simply dismisses the poorer classes as incapable of the attitudes required by liberal citizenship. On this view, the poor seek only self-serving and unfair terms of social cooperation and are, therefore, by Rawls’s criteria, unreasonable.”

It should by now be clear how the contrastive strategy offers hope to the radical trying to balance these threats with the intuition that respect for consent is a vital part of any positive revolutionary project. The contrastivist can acknowledge these sorts of concerns without giving up on consent. Yes, those who’ve internalized their oppression may consent, in one sense, to mistreatment — but there’s another sense in which they *don’t* consent to being so treated. The contrastivist can take Freeman’s lesson quite literally: the failing of liberal theory is that it *doesn’t* reflect the range of meanings that consent claims may take. The contrastive picture shows where we can draw the line: between consent claims made relative to *liberal* contrast spaces, which carve up the world in a ways that reflects concepts defined by, and in the interests of, the state and the ruling classes, and those that don’t.

That is, there’s new room for a critical strategy: the critical theorist can argue that the consent that’s given is consent *internal* to a structure, and so the justification

---

<sup>27</sup>See, *e.g.*, Iris Marion Young’s critique of individualism within distributive justice paradigms in the first chapters of (Young, 1990).

it confers is conditional on that structure. Most consent claims involved in attempts at justifying the liberal state, or at undermining radical projects, don't cut it; this is not because consent, broadly speaking, doesn't matter, but because the consent that does matter for these questions cannot, as liberal consent does, exclude from consideration radical alternatives (including, for instance, just social structures) or presuppose liberal structures (authority, ownership, family, money, career). For, in an important sense, liberal frames presuppose the existence of unjust structures in ways that obscure real possibilities of challenging them. This mirrors an insight of Garfinkel's:

The individualistic question takes the structural conditions as given. In particular it requires that we not question why the structural conditions are what they are but that we limit our questioning to states of affairs consistent with the structure. The consequence of this is that the individualistic question does not in any sense challenge the structure; rather, it chooses to accept the structure and sees its own problematic as navigating within it.

And so the individualistic framework ends up, in practice, supporting a proinstitutional bias. (Garfinkel, 1981, 152)

In other words, radicals shouldn't take themselves to be (too) constrained by concerns around consent relative to liberal frames; those very frames are an integral part of the social structure we're trying to undermine — and, arguably, part of the ideological defense mechanism of that structure. This thought is close to another of Garfinkel's. When we frame questions of unemployment in individualistic terms, he points out, we take for granted the structural facts that govern employment relations. As he puts it,

Someone operating within this problematic is seeking, in effect, to make sure that someone else gets unemployed. . . .

Moreover, while people are running around trying to improve their individual predicates, the structural condition remains unaddressed. The effect

of this is that the structure has received a silent blessing, accomplished by presupposing it and thus painting it out of the picture. (Garfinkel, 1981, 153)

By engaging with the question of unemployment on the individual level, we fall into a trap that renders non-cognizable the possibility of rejecting the structure altogether.

Indeed, we might find an even stronger view appealing: radicals should refuse, as much as possible, to engage with hegemonic frames; to the extent that this makes consent claims made relative to those frames unintelligible, those consent claims should be ignored.<sup>28</sup> In contrast, in the context of radical critique — and, in particular, critique which brings into the arena of inquiry the structural presuppositions themselves — the relevant contrasts shift; consent relative to these new frames should be taken seriously.

Importantly, this conception, which makes a distinction between revolutionary and non-revolutionary ways of carving up the world, points to where *practical* work lies, and to where it doesn't. Convincing people to consent to radical proposals relative to a liberal contrast space isn't just difficult, it's ultimately unimportant; it's only once consent is given relative to frameworks that matter that the consent itself matters.

This is *not* to say that those who are engaged in revolutionary practice shouldn't concern themselves with the consent of those who aren't. For a radical theory grounded in consent, a crucial end suggests itself, *viz.*, bringing about the preconditions for radical consent — which is to say, getting people to see the world in a (literally) radically new way. Onora O'Neill writes:

The morally significant aspect of treating others as persons may lie in making their consent or dissent *possible*, rather than in what they actually consent to or would hypothetically consent to if fully rational. (O'Neill, 1985, 259)

---

<sup>28</sup>Graeber: "There is a technical term for all this: 'prefigurative politics'. Direct action is a form of resistance which, in its structure, is meant to prefigure the genuinely free society one wishes to create. Revolutionary action is not a form of self-sacrifice, a grim dedication to doing whatever it takes to achieve a future world of freedom. It is the defiant insistence on acting as if one is already free" (Graeber, 2007).

To coöpt this insight: one’s duty (as radical) is to promote *radical* autonomy by making *radical* consent possible; and making radical consent possible means making radical framings of the world available. And, when we make radical frames available, we simultaneously *subvert* hegemonic frames.

Of course, radicals have independently good reasons to engage in this sort of subversion. Anarchist ethnographer David Graeber (2007) has offered a number of more specific examples of ways in which hegemonic framings can undermine radical projects. He argues, for instance, that the way journalists understand newsworthiness, violence, and journalistic integrity virtually guarantees that media coverage of radical actions will be either nonexistent or negative:

When [journalists] cover activist events, they are very self-conscious about the dangers that they might be manipulated — particularly [with] protests they see as “violent”. For journalists, there is an inherent dilemma here, because violence in itself is inherently newsworthy. A “violent” protest is far more likely to be covered; but for that reason, the last thing journalists would wish to think of themselves as doing is allowing violent protesters to “hijack” the media to convey a message. The matter is further complicated by the fact that journalists have a fairly idiosyncratic definition of “violence”: something like ‘damage to persons or property not authorized by properly constituted authorities’. This has the effect that if even one protester damages a Starbucks window, one can speak of “violent protests”, but if police then proceed to attack everyone present with tazers, sticks and plastic bullets, this cannot be described as violent. (Graeber, 2007)

Indeed, for such reasons, the subversive project — that of undermining hegemonic frames, and offering radical alternatives — isn’t primarily one of rational argumentation within the constraints of liberal discourse; as CrimethInc. put it, “ we don’t believe that we can *persuade* everyone to consent to our dreams of anarchist revolution; not only is the deck stacked against us, but the dealer, the table, and the whole house” (emphasis mine, CrimethInc. Ex-Workers’ Collective, 2012). Instead,

it's part of the project they call *seduction*, viz., that part that “creat[es] space for new possibilities and thus new desires to flourish.”

Put another way, effecting *conceptual revolution* is a crucial part of the revolutionary project so conceived. Of course, it would be a mistake to understand this sort of project as purely ideological. As Haslanger (2007) has emphasized, the material and ideological elements of social structure mutually constitute and reinforce one another; and, as Guy Debord put it:

Ideological entities have never been mere fictions — rather, they are a distorted consciousness of reality, and, as such, real factors retroactively producing real distorting effects; which is all the more reason why that *materialization* of ideology . . . results in the virtual identification with social reality itself of an ideology that manages to remold the whole of the real to its own specifications. (Debord, 1995, 150)

What's called for, then, is intervention along dimensions both material and ideological, engaging in projects that center around what Graeber (2007) calls the “political ontology of the imagination”, by articulating and enacting new structures capable of undermining dominant ones, and by drawing others into those frames. Be it through delivering rational argumentation, providing transformative aesthetic experience, or sabotaging the material bases of hegemonic structures, to succeed in these projects is to make radical consent possible. And, when we do succeed in such projects, we respect and value consent in a way that's not only compatible with, but central to, liberatory struggle.





## Chapter 2

# Conceptual Amelioration and Epistemic Responsibility<sup>†</sup>

There has recently been increasing interest in the practical aspects of the relation we have to our concepts, and in the possibility of motivating conceptual change with considerations of these aspects.<sup>1</sup> In particular, an emerging trend in social and political philosophy involves proposals to intervene in some of our concepts based upon moral consideration of what is involved in, and what are the effects of, having them in our repertoire. An especially interesting kind of intervention is one that purports to deliberately improve certain of our existing concepts, as opposed to discarding them and adopting wholly new ones — what’s been called *conceptual amelioration*.<sup>2</sup>

The very possibility of conceptual amelioration, however, has met with some resistance. There are two broad worries: first, that the idea of conceptual change is a non-starter; second, that conceptual improvement cannot be driven by practical or moral considerations. In this paper, we address the first concern briefly, by showing

---

<sup>†</sup>With Ekaterina Botchkina.

<sup>1</sup>See, for instance, (Barnes, 2015), (Burgess and Plunkett, 2013), (Plunkett and Sundell, 2013), (Haslanger, 2006).

<sup>2</sup>A prominent example of the ameliorative approach to conceptual intervention is Sally Haslanger’s work on the concepts of race and gender: while Haslanger agrees with numerous theorists that these concepts are deeply flawed, her interventions are distinguished from those of other approaches in virtue of their claim to preserve the very concepts we already have, rather than associating entirely different concepts with the relevant terms, or eliminating them altogether. See (Haslanger, 2006, 2012a). For conceptual intervention that seeks to eliminate the concepts in question, see (Appiah, 1985).

that, when the kind of conceptual change in question is appropriately understood, the first worry is misguided. The bulk of the paper addresses the second, by tackling the general question of what conceptual improvement amounts to, and what it is that differentiates conceptual amelioration, in particular, from other kinds of conceptual intervention.

We consider existing proposals for how to make such a distinction, issuing from work by Mark Richard and Sally Haslanger, but conclude that these proposals are too intertwined with the authors' other philosophical commitments to yield the kind of broadly palatable characterization we seek here. Moreover, from the ameliorationist's perspective, the views have the unwelcome consequence of imposing arbitrary constraints on the range of legitimate proposals. We propose instead a more doctrinally neutral view: that what distinguishes conceptual amelioration from other kinds of conceptual intervention is a commitment to a norm of *epistemic responsibility*, and we show that this view makes room for a wide variety of ameliorative proposals.

## 2.1 What is amelioration?

While conceptual amelioration has been receiving increased attention, there has not been much systematic work done in spelling out what, exactly, a successful ameliorative project would require. The presupposition behind conceptual amelioration is that parts of a community's conceptual repertoire are, at least in principle, subject to deliberate and reason-guided change; moreover, that certain of these changes are best construed as an improvement, rather than wholesale replacement or elimination, of the concepts in question. But projects that propose such changes face a predictable set of charges from opponents: that they're changing the subject, that they're radically breaking with past usage, that they're demanding that we be irrational or endorse nonsense, or that their proposal amounts to propaganda.

Of course not every proposal for conceptual change can defend against these charges; some will indeed involve mistakes, propaganda, misdirections, and confusions. Nor does every proposal aim at doing so: some conceptual interventions, for

instance, may quite *intentionally* be in the business of breaking radically with past usage, or changing the subject. It’s helpful, then, to stipulate some terminology to help demarcate the particular kind of change we’re interested in exploring.<sup>3</sup>

First, we’re concerned only with *deliberate* changes (or attempts at changes) to our conceptual-linguistic practices: we’ll call these *conceptual interventions*. This is a broad category that includes the intentional introduction of new terms for both new concepts (coining “quantum superposition” to pick out [*quantum superposition*]) and existing concepts (coining “enhanced interrogation” to pick out [*torture*]). At the other end of the spectrum, it includes the targeted elimination of terms in order to expunge them from our conceptual repertoire (discouraging the use of “spinster”).

Second, we’re not interested here in coining or erasing; rather, we’re interested in interventions directed at the concepts that particular existing terms are associated with: we’ll call this *conceptual engineering*. Engineering includes the intentional redefinition or reconceptualization of terms; including examples like the sex-positive use of “slut” to pick out [*sexually liberated*]), or the use of “disabled” to pick out [*people whose rights are at stake in existing disability activism movements*].<sup>4</sup>

Third, we’re focusing particularly on engineering that is, very roughly put, substantial but anodyne. We’re not interested in changes of subject, stipulative redefinitions, propaganda, nonsense, and doublethink. Rather, we’re hoping to understand *legitimate* conceptual changes; it’s these we’re calling *ameliorations*.

We might think that much of the best theoretical and philosophical work, illuminating obscure concepts and clearly delineating their interrelations, clarifying the murky, refining the ambiguous, sharpening the vague, unifying the disparate, enacts

---

<sup>3</sup>It is by now a familiar complaint that there are as many understandings of the word ‘concept’ as there are philosophers using it; the equally familiar recourse is to settle on one of these as a matter of stipulation. In this paper, we want to remain as neutral as possible about as many features of concepts as possible, adopting the broadest account that would accommodate the commitments of actual ameliorationist projects. We believe that our eventual proposal succeeds in this respect: our account will not depend upon any detailed account of what concepts are, so long as concepts are expressed by, or in some weaker way associated with, terms in natural language. For the remainder of this paper we will be formatting names of concepts, like [*dog*] and [*Africa*], italicized and in brackets.

<sup>4</sup>See Hardy and Easton’s *The Ethical Slut* (2017) for the former, Barnes’s *The Minority Body* (2016) for the latter.

just this sort of change. When humans began to be conceived of as animals, functions as sets of ordered pairs, childcare as labor, gravity as a geometric effect — these are all cases where there was a conceptual alteration of an *unproblematic* sort. There’s nothing untoward or unacceptable about them; at worst, they point to a need for unforeseen distinctions (between, e.g., the category-theoretic and set-theoretic notion of [*function*]). And, importantly, many of these changes were the result of *deliberate* intervention, by philosophers, mathematicians, activists, and scientists. So, in our terminology, they are (and so *there* are) ameliorations.

Now some readers may object here: these aren’t conceptual changes at all! It’s our *language* that’s changed; these are cases in which we started using words to refer to different concepts than they had previously denoted. The concepts themselves were untouched by the transformation.

Other readers may object along different lines: these aren’t conceptual changes at all! It’s our *understanding* of the concepts that’s changed; these are cases in which we revised core beliefs about the concept. The concepts themselves were untouched by the transformation.

These sorts of objections demand we be more explicit about what we mean by “conceptual changes.” In general, there are three kinds of ingredient under consideration: concepts, terms, and speakers. Correspondingly, we might interpret “conceptual changes” in three different ways: as changes in *the concepts themselves*, as changes with respect to *which concept adheres to a term*, or as changes in *how users think about a concept*. Which of these, if any, are we talking about here?

Our answer is to demur: by using the phrase “conceptual change”, we don’t take ourselves to be assuming one or another of these views. The changes are conceptual, in the expansive sense we’re using, in that they reach to the core of the corresponding terms and concepts: rather than being run-of-the-mill changes in empirical beliefs, they constitute changes in what might be thought of as *a priori*, or *essential*, or *necessary*, or *definitional*. The change from conceiving of childcare as a wholly private, familial duty to conceiving of it as a form of contribution to the economy, deserving of consideration as a kind of work and apt for negotiation, isn’t best described as a

new empirical discovery about childcare; instead, it is a more fundamental shift in our conception of what is essential to, or definitive of, that practice.

The bulk of the work in understanding amelioration, in our view, lies in understanding what it takes for a change to be anodyne in the sense above. This is perhaps surprising: at first glance, we might think that it's the very possibility of conceptual change, rather than the conditions under which a change is appropriate, that is most in need of defense. However, there is a sense in which conceptual changes are commonplace and unparadoxical: in cases of intervention (generally) or amelioration (specifically), some term T can be described as corresponding to a concept [X] before the intervention and to a concept [Y] afterwards. While the *basic nature* of the change that takes place here needs some explication, this is largely a matter of spelling out metaphysical and semantic details within one's preferred metaphysical and semantic accounts. Perhaps the change consists in the two descriptors X and Y picking out different concepts; or perhaps they latch on to different time-slices of the same concept; or perhaps they refer to the same, immutable concept but (in a Davidsonian spirit) induce different ways of thinking about it — concepts under descriptions, as it were. Different pictures in metaphysics, the philosophy of language, and the philosophy of mind will have different stories to tell about what's going on here. But any view worth its salt had better acknowledge that there's *something* going on: it will have to tell some story or other.

However, spelling out the basic mechanisms of an intervention of this sort using one's preferred theoretical tools is just the start. There is more work to be done: that of delineating the licit from the illicit, of saying what it takes to be an *amelioration*, rather than a more problematic sort of change.

We can start by considering the sorts of objections that problematic changes are apt to face; a *bona fide* amelioration should be able to withstand these. As we've seen, a number of such objections come to mind as soon as we imagine someone proposing a conceptual change. Take, for instance, a proposed change in the definition of 'poor', to include all and only those people with no income at all. It would seem right to object that the proponents of such a change are:

1. *Changing the subject.* When, with the new usage, we talked about “the poor,” we’d in fact be talking about only a small fraction of poor people. The change wouldn’t improve our understanding of what it is to be poor, but instead shift the subject to something quite different.
2. *Radically breaking with past usage.* The proposal doesn’t emerge from proper considerations or research about what it is to be poor; nor does it emerge from a process of rational deliberation around the nature of poverty and its relationship to other socio-theoretic concepts to reach a new reflective equilibrium. Rather, it countenances a brute redefinition of the term.
3. *Leading to absurd or nonsensical conclusions.* Under this proposal, it’d be entirely possible to be unable to afford to minimally house, clothe, or feed oneself, and to still fail to be poor. But this seems absurd: it’s part of *what it means to be poor* that anyone who can’t afford minimally decent living conditions is poor, and it’s part of *what it means to be poor* that receiving welfare subsidies and being poor are not mutually exclusive.
4. *Endorsing propaganda.* Any legislation of such a change — and the consequent headlines announcing that poverty has turned out to be far less prevalent than once thought — would have a distinctly Orwellian cast; using the term in this way would be misleading and deceptive.

Examples of such challenges in real-world contexts aren’t hard to come by. Anti-abortion declarations that life begins at conception, and attendant conclusions that abortion is murder, face the challenge that the notion of “life” on which life begins at conception is not the notion that’s relevant for murder; it’s a tacit, and potentially propagandistic, change of subject. Pro-same-sex marriage activists face(d) the claim that extending “marriage” to include same-sex unions amounts to “changing the definition” of marriage, and that it is conceptually impossible for, say, a woman to wed a woman. Suits brought by slaves in federal court were jurisdictionally barred when the Supreme Court held that “citizen” excluded slaves and their descendants; holding

that the Constitution counted them citizens would break with historical usage of the term, while holding that state law could make them United States citizens would be a conceptual confusion.<sup>5</sup> Disputes over what sort of things can be owned — people? land? water? nonhuman animals? ideas? — in fields ranging from anthropology to law to postcolonial theory often involve charges of conceptual confusion and changes of subject. Recent proposals on campus to declare that sexual assault occurs whenever there's sexual contact absent explicit, verbal declarations of consent face all of the above objections: that the new rules aren't really talking about sexual assault at all, that they're radically redefining the term, that they're clearly misunderstanding the nature of sexual assault, that they're a kind of feminist propaganda.

As should be clear from these examples, these concerns very often overlap, providing some reason to think that they help to pick out a unified kind: conceptual engineering that doesn't pass muster. But they aren't quite coextensive: there may, for instance, be propagandistic changes that aren't subject-matter changes, when a genuine ambiguity is resolved by law in one direction for political reasons.<sup>6</sup> Without getting into the details of exactly how these problems relate to one another, we can use them as a rough guide: we can take meeting these challenges to be desiderata for ameliorations. An amelioration, but not a less anodyne sort of engineering, should be safe from these worries. See Fig. 2.1.

Part of our project here, then, is to identify what it takes to guarantee that a proposal can meet these challenges: as a shorthand, we'll say we're looking for the *ameliorative constraint*: a property, or set of properties, that distinguishes genuine ameliorations from the more inclusive categories of conceptual interventions. While plenty of existing proposals for conceptual change seem to tacitly depend on such a constraint to defend the acceptability of their programs, no explicit account, to our knowledge, has been offered so far. An account of this constraint would be a tool for the ameliorationist, giving him a way to frame the relevant considerations. At the same time, such an account could offer to those not engaged in ameliorative projects

---

<sup>5</sup>(*Dredd Scott v. Sandford*, 1856)

<sup>6</sup>Thanks to Justin Khoo for this point.

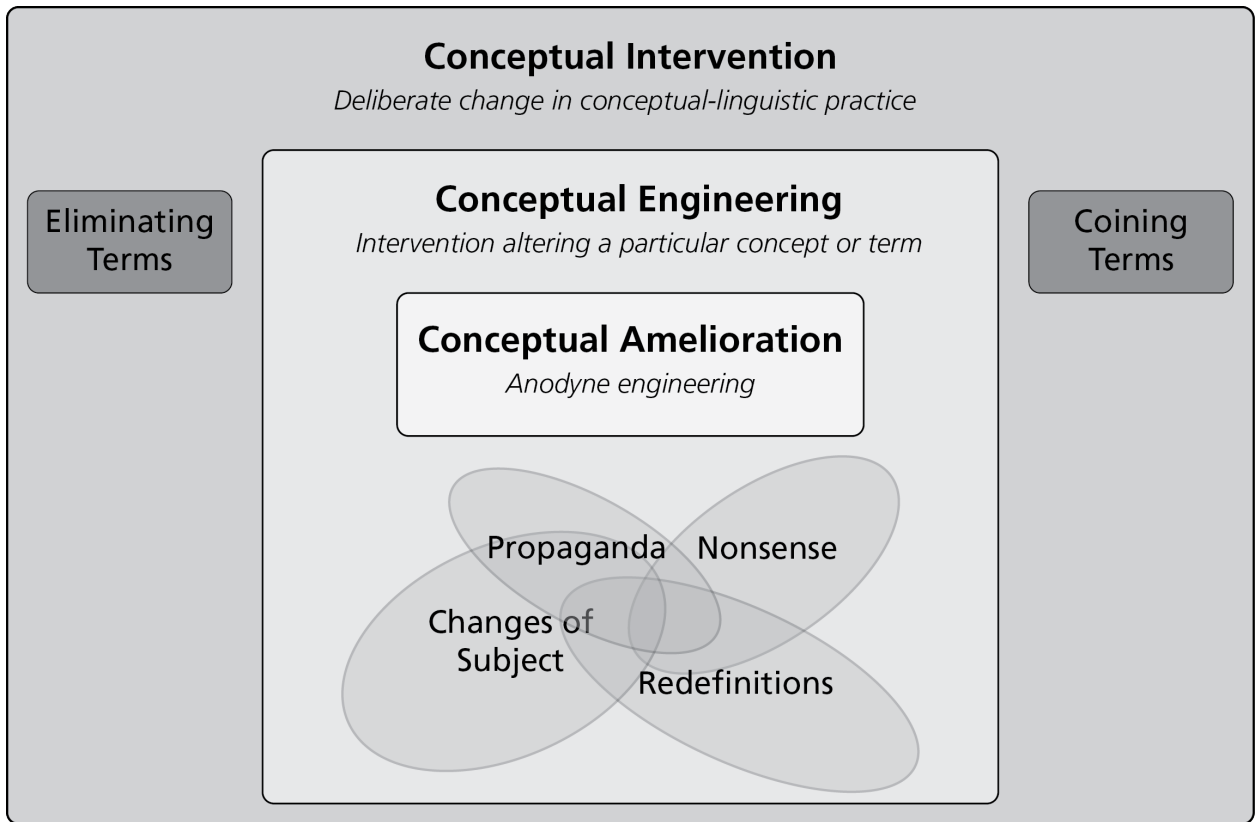


Figure 2-1: The universe of conceptual interventions

a way of understanding what such projects are up to, and why they're not as hard to swallow as they might at first appear.

In order to best serve both these purposes — providing both a tool for the engaged ameliorationist and an explanation for the interested outsider — we set our sights on a criterion that is acceptable to the widest range of theorists; to this end, we hope for an account of the ameliorative constraint that can be stated neutrally with respect to most substantive issues in semantics, philosophy of mind, and metaphysics. We've already started down this road with our characterization of conceptual change, and indeed, similar considerations apply to the broad project: it seems that the issues that arise in the debates between the ameliorationist and her opponents aren't the sort that are resolved by highly technical considerations in linguistics or metaphysics, but are instead questions that any serious-contender theory of metaphysics, language, or mind should render coherent, debatable, and live. Thus, in terms of the external philosophical commitments required to accept the reliability of the constraint, we aim



for ecumenicism.

In addition, we want our constraint to be general enough that it adequately characterizes the full range of ameliorative projects. It should not be applicable only to a particular range of conceptual interventions — applying to concepts in art, say, but not in politics. Nor should it undergenerate: the constraint should, ideally, provide not only a sufficient condition for a project to be ameliorative, but also characterize all — or at least all of the core — ameliorations.

In what follows we will provide an account of this ameliorative constraint inspired by considerations from philosophy of science. Our account is wide-ranging, quite neutral with respect to metaphysical and semantic theory, and yields a constraint that can guarantee that a conceptual engineering proposal is a genuine amelioration meeting the objections enumerated above. Before laying out our theory, however, we will look at two prominent alternatives and outline ways in which they are too narrow to apply to what might antecedently seem (and, given our eventual analysis, turn out) to be legitimate proposals

## 2.2 Existing approaches to amelioration

There are a number of proposals in social theory and philosophy of language that may be mined for the sort of resources that are useful for spelling out what is distinctive about ameliorative projects. Two such accounts, one due to Sally Haslanger and one to Mark Richard, are particularly promising; we will consider whether they yield the sort of constraint we're after. We will argue that while both proposals help illuminate a good deal about what amelioration entails, they nevertheless fall short of offering the fully general, neutral constraint we're after.

### 2.2.1 Haslanger: amelioration as metasemantic

Haslanger has argued for social constructionist understandings of [*race*] and [*gender*], and their associated families of concepts. For example, she takes the concept [*woman*] to be best analyzed as, roughly, “an individual marked for oppressive treatment on the

basis of presumed female reproductive capacities ”; racial concepts like [*Latinx*] and [*White*] are given analogous explanations.<sup>7</sup> In both cases, the central idea is that the concepts in question pick out social properties and relations, rather than biological ones.

What Haslanger takes to ground her claim that the social constructionist project is ameliorative, rather than revisionary or eliminativist,<sup>8</sup> is her insistence that the proposals, surprising though they undoubtedly are, are nevertheless “not changing the subject” — that they can legitimately claim to be revealing what race and gender “really are,” (Haslanger, 2012b) rather than proposing entirely new and merely stipulative uses for these terms.

Importantly, Haslanger understands the social constructionist as being in the business of correcting our *understandings* of the relevant existing concepts, either by drawing our attention to pervasive misconceptions or by pointing out that the term has evolved in unnoticed ways, rather than changing those concepts themselves. Since the constructionist is only addressing our own confusions about how things already stand, she can’t be accused of changing the subject: indeed, she’s helping us understand what the subject is (and was), in the first place.<sup>9</sup> Haslanger expresses this thought in various ways:

The constructionist is not changing the subject, or changing our language. [A]lthough the constructionist suggests that we come to a new understanding of our concept, this does not require replacing our old concept with a new one, but understanding our original concept better. (Haslanger, 2006)

or

Social constructionists aim to reveal that the concepts we employ are

---

<sup>7</sup>For simplicity we’ll focus here on the gender terms, but what we say is applicable, *mutatis mutandis*, to Haslanger’s other foci.

<sup>8</sup>*cf.* (Appiah, 1985)

<sup>9</sup>Her approach here is reminiscent of Rawls’s use of the distinction, due to Gallie (1956), between a concept and a person’s conception of it: on Rawls’s view, different people disagree over particular conceptions of justice, but share a common concept of justice (Rawls, 1996).

not exactly what we think they are. ... The proposed analysis may be surprising; and it may even be that the term has come, through practice, to express a different concept than it used to. ... *But the constructionist is not causing this, or even promoting such a change, but is rather revealing it.*” (Haslanger, 2012b, emphasis ours)

On what grounds can the constructivist claim to be offering a more accurate understanding of, rather than a change in, what our concepts already are? Here, Haslanger appeals to a combination of semantic externalism and a sort of realism about social kinds. Close examination, her story goes, of our actual practice of deploying the concept [*woman*] reveals that we are tracking a real, unified type, albeit one based in social practices. Then, by appealing to an account of reference on which “concepts pick out an objective type, whether or not we can state conditions for membership in the type, by virtue of the fact that their meaning is determined by ostension as paradigms,” she arrives at the conclusion that the constructionist account captures what [*race*] has been picking out all along, even though this fact was unapparent to naïve introspection.<sup>10</sup>

Haslanger’s social constructionist thus justifies her claim to amelioration by seeing the intervention as *metasemantic*: what we talk about when we talk about gender stays the same, though our understanding of it shifts. This approach suggests a contender for our general ameliorative constraint:

***Amelioration as Metasemantic:*** A conceptual intervention is an amelioration just in case it improves our understanding of the very concept that is already in use. This concept is individuated via direct reference to a real, unified kind. The concept, and its adhesion to the corresponding term, remains unaltered by the amelioration.

---

<sup>10</sup>This is on analogy with the pictures offered by Putnam in (Putnam, 1973) and (Putnam, 1975), and by Kripke in (Kripke, 1980), on which, roughly, the meaning of “gold” is determined by ostension of a paradigm example of  ${}_{79}\text{Au}$ , and, as a result, we can learn about our own concept [*gold*] by learning about the natural kind. By allowing that certain features of our concepts are fixed by external relations to items in the world – meanings ain’t wholly in the head – we can admit that the ‘real’ meaning of our concepts are significantly different from what we took them to be.

Does this proposal (which we'll shorthand to *AM*) provide us with an acceptably general and neutral constraint? Not quite, since, as we'll argue in the remainder of this section, *AM* imposes strong restrictions on the ameliorationist project; if there's a weaker constraint that an ameliorationist could use, ecumenicism suggests we explore whether that constraint could do equally well.

More explicitly, *AM* imposes overly strong ontological and justificatory requirements on viable ameliorations. Consider a proposal<sup>[9]</sup> according to which the concept [*national security*] should be ameliorated to mean something like [*the welfare of a nation's inhabitants*]; it follows from this that such matters as healthcare, access to food, and environmental protection should all fall under the purview of [*national security*], whereas on our current understanding, they don't. Let's grant for a moment the semantic view that Haslanger favors, in which reference is fixed by some external relation to a real, appropriately unified kind. In order to establish that the proposal above meets *AM*, one would have to independently make the case that the (not-yet-ameliorated) concept [*national security*] really *does* track some such unified kind.

Now, perhaps that case can be made. But defending the legitimacy of this particular ameliorative proposal should not require making it. After all, the case that [*national security*] is unsatisfactory in its current form is made, by and large, by reflecting on such matters as the prevalence of child malnutrition, the portion of the federal budget allocated to the military, and so on. The question of whether we should understand [*national security*] differently does *not* seem to depend, by contrast, on the correctness of externalism and realism about social kinds. The ameliorationist who encounters, in response to an argument that we should reframe our understanding of [*national security*], a reply addressing this latter issue — “I grant that that'd be nice, but it's too bad; the real kinds necessary to do so just aren't out there in the world” — would rightly feel that her argument had been misunderstood, that this response was point-missing. But if *AM* were correct as a general ameliorative constraint, this criticism would be perfectly apt. *AM* thus seems to unduly entangle ameliorative proposals with subtle, and intuitively unrelated, metaphysical questions.

If, on the other hand, *AM* is a constraint with a restricted domain of application – to, say, only those proposals targeting a certain *sort* of concept – then it fails to offer the generality we’re looking for.

For another way of seeing why *AM* is undesirable as a *general* description of amelioration, note that it requires that the current concept be, in a sense, *already* what the amelioration is urging it to be: “the suggestion here is not that the concept itself changes ... but rather that our understanding of it does” (Haslanger, 2006). But consider an ameliorationist arguing, in 1960, that [*marriage*] ought to encompass same-sex relationships. To make this proposal acceptable by the lights of *AM*, the ameliorationist must produce evidence for a certain concrete hypothesis about the social world: namely, that the 1960s concept [*marriage*] antecedently has the feature she is urging us to acknowledge — being applicable to same-sex relationships — even if we fail to recognize this introspectively. But might it not turn out that that hypothesis is false? That is to say, Haslanger’s constructionist has the resources to deal with one kind of failure: if the concept [*marriage*] that 1960s users *take themselves* to be applying fails to yield the feature the ameliorationist wants, the constructionist can reply: “that’s no problem; the concept that they’re *actually* applying does yield that feature.” But what about a more troubling situation, in which the concept won’t cooperate? What if, when we examine the totality of determinative facts about the concept — facts about its historical origin and development, its use within communities of expert speakers and at large, its significance to actual social practice, the purposes with respect to which this concept is currently useful, and so on — we fail to find any evidence that the 1960s [*marriage*] accords with the feature that the amelioration is calling for? Perhaps the concept as used in the 1960s simply *does* exclude the possibility of same-sex marriage. In that case, if *AM* is generally correct, the concept can’t be ameliorated: the best we can do is to agree to use the term “marriage” in a new, non-subject-preserving way.<sup>[11]</sup>

But in accepting this consequence, the ameliorationist would concede too much: there is, it seems, a plausible case to be made that even if it’s true that the concept [*marriage*] doesn’t, as used in 1960, countenance same-gender unions, it *ought* to.

The force of this *ought* derives not from some already existing but unacknowledged feature of the 1960s concept, but from the fact that that's what the best description of marriage would *be*. What the ameliorationist needs, in this case, is not just a change in our understanding of the concept, but rather a real change in the concept itself — precisely what Haslanger's social constructionist is not in the business of providing.<sup>11</sup>

<sup>12</sup>

Now conditioning on the correctness of metaphysical and semantic theories is a tricky business, so at the risk of being repetitive, it may be worth spelling out a subtlety here. We are not claiming that there is no conceivable world in which *AM* is a correct and complete account of amelioration. *If Haslanger's metaphysical and semantic views are correct*, it may be the case that amelioration is entirely described by *AM*. We are, however, making two closely related claims. First, from the perspective of the theorist, we should be seeking a characterization that does not hinge on this antecedent, but rather points to a phenomenon that can be articulated, and understood, within alternative metaphysical and semantic views. Second, from the perspective of the practitioner, we should be seeking a characterization that clearly identifies the realms of disagreement that are relevant to an ameliorationist defending her proposals, and that clearly connects these realms with the desiderata above.

### 2.2.2 Mark Richard: concept identity across alterations

To contrast with *AM*'s picture, on which amelioration is not a change in the concept, but rather a change in our understandings of our usages, consider an account on which what the ameliorationist proposes is *altering* the relevant concepts: effecting real changes to our concepts themselves (and not just to our understandings of them) in a way that nevertheless preserves those concepts' *identities*. This would seem to

---

<sup>11</sup>See, again: “but the constructionist is not causing this, or even promoting such a change, but is rather revealing it” (Haslanger, 2006).

<sup>12</sup>Haslanger's work on amelioration, in general, is sometimes easy to conflate with her work on social constructionism, in particular. So far, in examining options for substantiating the Ameliorative Constraint, we've only considered the latter. But there are important insights in Haslanger's work that point in more promising directions for a general account of amelioration; we will return to these in due course.

do equally well, if not better, when it comes to the desiderata: if a case can be made that an ameliorative proposal leaves us with the very same — that is, numerically identical — concepts that we had before the intervention, that would seem to provide strong defense against charges of willfully changing the subject, engaging in a kind of propaganda, and other criticisms that an intervention’s status as an amelioration is meant to deflect. So let us consider a constraint that ties amelioration directly to concept identity:

*Amelioration as Concept Identity:* A conceptual intervention is an amelioration just in case the alterations it effects in the relevant concepts are identity-preserving.

Of course, it’s worth asking whether this constraint on amelioration can so much as get off the ground: does it make sense to suppose that concepts can undergo substantial alteration while remaining numerically identical? After all, many ameliorative proposals call for significant changes in usage, such as changes in the prototypical features of the concept and in the modal profile of the associated term, often in ways that are incompatible with judgments by current speakers.<sup>13</sup> If a concept cannot undergo such changes without ceasing to be the same concept that it was, then this conceptual identity-based constraint will fail to characterize such proposals as ameliorative.

But Mark Richard, in forthcoming work, suggests that the identity conditions of concepts are flexible enough to withstand the sorts of substantial changes mentioned above. He develops a view on which the concepts involved in natural language are not “static and unchanging . . . abstract denizens of a ‘third realm,’” (Richard, 2019) but are rather fluctuating, mutable, historical entities, capable of persisting across significant changes. For Richard, concepts can evolve much like biological species

---

<sup>13</sup>Haslanger’s proposal for the amelioration of race- and gender-related concepts exhibit all these features. So for instance, she allows that while actual possession of female reproductive organs may currently be a central feature associated with [*woman*], that would no longer be true if the concept were altered as Haslanger proposes. Similarly, if “women” comes to mean what Haslanger proposes, it will be false that women would exist in a perfectly just world, yielding a different modal profile than the term currently has. Finally, Haslanger accepts that the majority of existing English speakers may deny, on reflection, that her proposed ameliorations would yield legitimate applications of, say, [*woman*] or [*Latino*].

do, without thereby ceasing to be the very same concept.<sup>14</sup> That's because, on this account, concepts aren't fixed by any particular, rigid set of identity criteria of the kind that are frequently invoked in theorizing about concepts, such as specific role in inferential relations, association with a fixed set of prototypes, causal relation to the environment, pattern of application, and so on. Instead, he claims, all of these factors *contribute* to making a particular concept the concept that it is, but no one of them, nor any fixed combination thereof, does so in a way that is decisive and immune to change. As a concept evolves, it is possible to imagine any one of these factors shifting significantly — for example, a concept that used to have XYZ as a prototype comes to have UVW as a prototype — while the concept remains numerically the same.<sup>15</sup>

It is useful to think of Richard's view of concepts as involving two types of latitude, or flexibility, for how a concept may develop. The first type of flexibility has to do with existing vagueness and indeterminacy in current and historical use of a concept. Would the concept [*cat*] be accurately applied to creatures that we discover are actually robots from Mars, and have been all along? On Richard's view, it's plausible that the matter is simply underdetermined by the nature of the concept in question — even if we knew everything that there was to know about the concept's history, we would find nothing that would rationally demand that we judge one way or another. In that sense, there is a kind of freedom when it comes to our application of the concept [*cat*] to the robotic creatures from Mars; it's a matter of decision, not compulsion.<sup>16</sup> Call this dimension of flexibility *backward-looking*, since it is flexibility

---

<sup>14</sup>Truth be told, Richard remains deliberately vague on whether his account is an account of *meanings*, *concepts*, *words*, *languages* or some combination of the above. For the purposes of this paper, we will apply the conclusions of his arguments to concepts in particular, *modulo* the disclaimers at the beginning of this chapter.

<sup>15</sup>It's useful to return to Richard's analogy with biological species. Just as there are many factors that contribute to constituting the nature of a species — genotype, phenotype, lineage, environmental relations, and more — it is arguably implausible to suggest that any one of these, or any specific combination, serve to determine once and for all what the species *really* is. Instead, in the right circumstances, any of these particular factors may change without thereby entailing that speciation has occurred.

<sup>16</sup>It is worth stressing that, on this view, the ambiguity is not a matter of it being open which concept the term expresses, but rather is a matter of the concept itself being ambiguous between including and excluding the robotic specimens. So the proposed ambiguity should also not be understood, as on the Haslangerian model, of a feature of our understandings of the concept. The backwards-looking flexibility is not (merely) flexibility to change our understanding, but flexibility to change the concept itself.



with respect to facts about the concept as it currently is, issuing from its particular historical development.

But there's another kind of flexibility made available by Richard's view: a *forward-looking* kind. This kind of flexibility isn't due just to vagueness or underdetermination, but also to the possibility of evolution and change: on this view, even if a certain judgment is currently definitively licensed (or ruled out) by the present facts about a concept, this doesn't yet settle the question of whether that judgment will *continue* to be definitively licensed (or ruled out). *That* question can only be settled given the actual course that the concept takes.<sup>17</sup> Consider the concept [*rape*]: it is arguably true that, according to the stage of the concept's development in 1850, it was false to say that a woman could be raped by her husband. It was, in other words, a fully determinate feature of (that stage of) the concept of [*rape*] that it excludes intramarital sex. But Richard's view allows that that very same concept has undergone an evolution, such that it is now *no longer* a feature of the concept that it excludes intramarital sex, and that it is now true to say that a husband could rape his wife.

Richard's picture, then, displays several features that certainly seem important to the kind of considerations relevant to the ameliorationist's defense. For one thing, his view<sup>18</sup> brings out the limits of introspection in determining, broadly speaking, what it is that we're talking about. For another, it emphasizes our role in *shaping* our concepts, as opposed to being passively bound by the constraints that they independently impose. Both are crucial gains for a project that depends on the possibility of our successfully taking deliberate control of certain aspects of our conceptual practices, often in ways that deviate sharply from existing usage.

Given that on views of concept identity like Richard's, *ACI* doesn't immediately rule out any possibility of conceptual amelioration, we might ask whether it offers an acceptable reading of the ameliorative constraint. As we've seen, *ACI* does offer a

---

<sup>17</sup>Returning once again to the species analogy: nothing in the history of a species can definitively determine whether or not its members will evolve to exhibit feature Y (within certain limits, of course, for what kind of feature Y can be), since that is a matter that depends on a huge variety of factors, including environmental contingencies.

<sup>18</sup>The central themes of which are also present in arguments made by Yablo (2008) and Williamson (2007), and date, with some interpretative charity, back to Quine (2013).

*prima facie* advantage over *AM*, in that the former makes room for certain seemingly worthwhile proposals that the latter immediately rules out. One of the limitations of *AM*, recall, was that it prohibited the ameliorationist from enacting genuine *alterations*: that is, from enacting adjustments or extensions of the existing concept that would leave it with features that it previously lacked. But because *ACI*, given a picture of identity like Richard's, leaves room for both forward- and backward-looking flexibility in the shape of the concept, it allows real changes to the concept to still qualify as ameliorations, so long as conceptual identity is preserved.<sup>19</sup>

Unfortunately, *ACI* turns out to be ill-suited as a general constraint on amelioration for much the same reason as *AM*. For one thing it, like *AM*, requires a certain very specific stance on a seemingly independent question about the nature of concepts: under *ACI*, if amelioration is to be possible, then the question of concept identity cannot be fully determined by a concept's analyticities; it must, at least in part, be answered by reference to other facts. Indeed, the range of positions on concept identity required for a plausible *ACI* is even narrower than this: on any view on which the question of identity rests upon the *outcome*, independent of the *process*, of linguistic evolution, it seems that that question is entirely separable from the ameliorationist's concerns. The ameliorationist concerned with countering claims that she's changing the subject would be giving an unsatisfactory answer, and, more sharply, the wrong *sort* of answer, were she to reply by providing detailed sociolinguistic survey results, for example. So *ACI*, paired with a view that locates concept identity in statistical regularities in usage without reference to just how those regularities came to be, will fail to provide the defense the ameliorationist needs.

In fact, even a path-dependent view of identity won't clearly suffice; in order to offer an adequate defensive strategy for the ameliorationist, *ACI* requires a view of concept identity on which identity is not just path-dependent, but path-dependent in the right way. If the features of evolution that determine identity under Richard's

---

<sup>19</sup>Returning to our marriage example from the previous section, an ameliorationist following *ACI* now has the option to call for a real change to the concept of marriage without thereby conceding a shift in subject matter. According to *ACI*, the fact that the 1960s concept [*marriage*] definitively rules out same-sex unions no longer stands as a conclusive impediment to amelioration, as it did under *AM*.

view, for instance, fail to preclude unacceptable shifts, then the ameliorationist is still left with an inadequate response to the changing-the-subject challenge. In order to provide a satisfactory strategy, *ACI* must come together with a view on which identity depends upon quite specific features of the path of linguistic evolution. Yet at this point, it seems that admitting the possibility of amelioration given *ACI* imposes an unreasonably stringent requirement in demanding such a specific thesis in philosophy of language; if the nature of amelioration can be accounted for through a constraint that's neutral with respect to such questions, better to do so.

Such considerations also bring to the fore another question: given the ameliorationist's principal concern with maintaining connections with prior usage, why should she care about establishing *concept identity*, over and above the adequacy of the proposed shift in practice itself? That is, it may be — as on Richard's view — that the question of identity is fully answered by considerations about practice; to the extent that it's not, however, there seems to be no reason that the ameliorationist should be worried about it. The question of identity seems to be a technical question in semantics, one distinguishable from the concerns most pressing for the ameliorationist.

## 2.3 Epistemic Responsibility

Let's take stock of where we are. A substantive account of conceptual amelioration requires distinguishing it from both eliminativism and wholesale conceptual overhaul, while avoiding charges of mere stipulation, propaganda, or irrationality. But when we tried to locate a general constraint inspired by accounts given by Haslanger and Richard, we came up short: *AM* left the ameliorationist overly constrained, whereas *ACI* wasn't directly responsive to the sorts of issues that subject-matter preservation is intuitively supposed to address. We're left looking, once again, for a usable criterion.

Let's take a step back by looking, in more detail, at what exactly goes wrong with proposals that *fail* to meet the desiderata for a genuine amelioration. Consider, at the limit, conceptual interventions resulting in statements like the following:

WAR IS PEACE

FREEDOM IS SLAVERY  
IGNORANCE IS STRENGTH

The first thing to note is that it's not just the dystopian consequences of Newspeak that make these conceptual interventions unacceptable. For, setting those problems aside, there seems to be something purely *epistemically* objectionable about those sentences: war just is *not* peace, we want to say, and it would be crazy to believe otherwise.

Nor is it difficult to imagine cases in which moral and epistemic considerations come apart. Consider another proposal, perhaps closer to home, that intuitively fails as a conceptual amelioration, despite potentially positive moral consequences:

*Sexual Harassment*: Alter our practices so that “sexual harasser” is properly applied to anyone who is accused of sexual harassment, whether or not the accusation is true.

Now, on at least some views, the practical consequences of adopting this practice would be positive, overall: the cost in guiltless people being treated as sexual harassers would be outweighed by such benefits as there no longer being a question of the credibility of survivors. There's room to resist this moral claim, of course. But even if we granted it, conceding that such engineering would produce a net positive moral gain, the proposal would nonetheless strike most of us as troubling; it's worth exploring why, exactly, we are right to feel uncomfortable.

A first-pass response is to think that what's going wrong with the proposal is that it would commit us to many simple attributive uses of the term “sexual harasser” that would be false in the envisioned scenario; for instance, we'd be committed to stating, falsely, “John is a sexual harasser” when John was accused of sexually harassing Terrence, but has in fact never harassed anyone. But we need to be careful here, because there's a sense in which “John is a sexual harasser” *wouldn't* be false: someone saying the words “John is a harasser”, or “John sexually harassed Terrence”, is saying something *true* as uttered in a context where the proposal has taken effect — a context in which the meanings of “sexual harassment” and its kin have altered. So it can't be

the falsehood, *per se*, of these simple attributive uses that's the problem.

The worry comes into sharper focus when we consider the totality of the new linguistic practice that would result from adopting the proposal. If we look at sentences a little further downstream from the simple attributive uses — sentences such as “sexual harassers have done wrong” and “sexual harassers should be publicly condemned” — we see that *these* would come out false. And, with a little reflection, it will become clear that a vast number of the sentences involving “sexual harassment” that are true in our current usage will, on the revised usage, be false.

One result of this is loss of the ability to simply state important truths. For instance, we would no longer be able to express a moral fact by saying “sexual harassment is wrong,” since that sentence would now be false. Instead, we would need to articulate the differences between those cases of purported sexual harassment that are wrong and those that aren't, and find a way to express *this* with the revised language. And even if we grant that such an account could be produced, its new formulations of the relevant moral facts would undoubtedly be significantly more tortuous than the originals. If we think that it's an epistemic virtue to adopt a language that allows for perspicuous expression of simple truths, then the proposal brings about epistemic ills.

Of course, this result relies upon the proposed change being unaccompanied by concomitant changes in usage of other terms, such as “wrong”. And, indeed, we might think that the example is underspecified: the proposal hasn't said how, exactly, the change would be enacted. Would the scheme fare better if we imagined the proposed change *propagating*, such that much of our current usage could be recuperated in virtue of corresponding changes in concepts for other words?

It's not obvious that it would, for such shifts would likely be epistemically unpalatable in other ways. If, for instance, we think that it's an epistemic virtue to have non-disjunctive kind terms that capture salient moral features, then some ways of ensuring that we don't end up uttering falsehoods will fail to pass muster, by resulting in the blurring of important distinctions. In the case we're considering, for example, if we end up able to truthfully assert “sexual harassment is wrong”, this

can only be due to a shift in the concept expressed by “wrong;” such a shift might be expressed (using our terms) by “‘wrong’ will apply to a case  $C$  if and only if  $C$  is not a case in which someone’s accused of sexual harassment and  $S$  is wrong, or  $S$  is a case in which someone’s accused of sexual harassment.” This shift, needless to say, looks as though it has made a non-disjunctive expression into a disjunctive one. So this proposal, too, has epistemic defects.

Reflecting on these changes, it seems that we can point to a *prima facie* reason to be suspicious of proposals for conceptual intervention whose implementation fail to meet the desiderata for genuine ameliorations. Whatever their *moral* virtues, it seems, such proposals run serious *epistemic* risks: they’re likely to have bad results for our ability to describe, inquire about, and coordinate action within the world.

Let’s now consider a contrast case, in which those sorts of epistemic worries don’t seem to arise: the case of conceptual revision in the course of scientific theory development. For instance:

*Brain Tumors:* Traditional classification of IDH-mutant anaplastic gliomas (a type of brain tumor) has historically been based entirely upon morphology and histology, and has recognized three distinct categories AA, AOA, and AO. New discoveries indicate that AAs display a certain molecular marker  $M$ , and AOs do not. Moreover, AOA tumors display markedly different development patterns and call for differential treatment; those marked with  $M$  act like AAs, those not so marked act like AOs. Neurooncologists propose to revise the classification on these grounds, to eliminate the middle category of AOA, and understand classification into AA or AO to be based on molecular markers, rather than morphology and histology. (Wiestler et al., 2013)

We can imagine the scientists here declaring: “whereas our concept of [AA] has up until now been based upon morphology and histology, the concept we *should* be using is based rather on the presence or absence of  $M$ .” Under the old way of thinking, no tumor with a mixed glial cell origin could be an AA, under the new way of thinking, this is no longer the case.

It seems clear that *Brain Tumors* is an example of successful conceptual amelioration; whatever Orwellian misgivings were triggered by *Sexual Harassment* seem no longer to apply. In particular, despite the fact that the proposed change would result in significant revisions to which sentences we take to be true (*e.g.*, “there are three distinct types of IDH-mutant anaplastic gliomas” will be false under the proposal), it still seems that the revision leaves us talking about the same things - but in a better way . To put it in the terms of a familiar metaphor, we can see that the proposed alteration would make the right sorts of ripples in our web of belief. Of course, as with *Sexual Harassment*, the revised usage does lose expressive power in a sense — in particular, we lose the ability to classify entities as AOAs - but this is no real loss: what we imagine here is a new discovery about features of the world that the linguistic alteration is responsive to. So what, exactly, is the difference between the two cases, and is there a way to distinguish between them that would be systematic and reliable? In other words, can we identify more specifically the features in virtue of which *Brain Tumors* counts as a general amelioration, while *Sexual Harassment* does not?

It’s worth getting one difference between *Sexual Harassment* and *Brain Tumors* out of the way immediately: in the second, but not the first, the conceptual change comes about as a part of the ordinary process of going about our epistemic business, as opposed to aiming at conceptual change directly. But it would be a mistake to think that this difference reflects the features in virtue of which one is problematic where the other is legitimate. For there seem to be cases in the sciences in which conceptual change was developed and adopted for the purpose of attaining a better systematization; think, for instance, of the shift in late twentieth-century mathematics toward characterization of mathematical objects like groups and manifolds in terms of category theory rather than set theory. And, quite generally, it isn’t clear why the intentions of the theorists should be relevant to the viability of their conceptual intervention. The neurooncologists may not have been aiming explicitly at a conceptual intervention, but we can agree that their practice clearly did result in a perfectly acceptable such intervention. We can judge the character of an interven-

tion by looking at the character of the change envisioned, without reference to the intentions that motivated it.

A more relevant difference between the *Brain Tumors* case and the *Sexual Harassment* case is that the putative conceptual shift in the first, and not the latter, is *epistemically responsible*. The neurooncologists who propose to shift our conceptual practices are responsive to the right sorts of features in the world, and they're proposing a change for the right sorts of reasons. They take themselves to be engaged in the same sorts of inquiry, to be talking about the same subjects, as their predecessors. The conceptual change they offer, they argue, serves up a better way of theorizing about brain tumors, by recognizing what's really important about the differences in a way that will lead to better treatment practices, without leading to false beliefs.

So, to characterize it ostensively: an *epistemically responsible shift* in the concept associated with a term is a shift that's relevantly similar to the sort of shift that occurs in good examples of scientific theory change. Without offering a full-fledged theory of epistemic normativity, we can point to the sorts of considerations that enter into a determination of whether or not an engineering proposal is a responsible one. A recognition, for instance, of internal inconsistency or incoherence in a theory's concepts provides a very good reason to revisit those concepts. Likewise, discoveries about the world may lead us to recognize that our concepts are inadequate. Or we may come to see that changing the terms in which we theorize may lead to a more elegant or fruitful theory. And, of course, whether an altered scheme commits us to revising our judgments on what we take to be truths<sup>20</sup> is also a factor. An epistemically responsible revision, then, is one that balances these sorts of theoretical considerations.

These considerations lead us, at last, to a new proposal:

*Amelioration as Epistemically Responsible Intervention:* A proposed conceptual intervention is an amelioration just in case it is epistemically responsible.

This analysis will give the theorist a way to justify claims to subject matter preservation, freedom from absurdity, continuity with past usage, and so on, without

---

<sup>20</sup>Consider "Pluto is not a planet", for example.



relying on semantic externalism or any kind of reference magnetism; more broadly, it allows her to approach the question of whether a proposal is an amelioration without deferring to the details of semantic theory.

Note that we're not here proposing a 'science first' account. As we see it, there's no need to import a sophisticated metaphysics or epistemology of scientific practice in order to be able to evaluate ameliorationist proposals for epistemic responsibility. The idea is not that epistemic responsibility is something that's particular to science, nor that every ameliorationist proposal needs to imitate a scientific revision. Rather, we're pointing out that there's *something* to which we're responsive in scientific practice that constitutes a standard for subject matter preservation, a standard that governs just when it's appropriate, and when not, to continue to use a term with a new understanding. There's some standard, that is, that calls for "atom" to be retained, and "aether" abandoned. And this standard is not limited to formal or informal scientific theorizing: the reasons for abandoning "aether" and retaining "atom" are much the same as the reasons for abandoning "witch" and retaining "race".<sup>21</sup> This standard, *whatever it is*, is what we're calling "epistemic responsibility".

To put this another way, whether a shift in concept for a given term is an epistemically responsible one is a criterion that distinguishes an interesting kind of conceptual engineering. If an engineered shift in a term's usage is epistemically responsible, then the new term can be put to use for the same sorts of inquiry as the old one; just as in good scientific practice, when we continue to use a term under a revised understanding — as opposed to discarding the term altogether — this is an indication that there's something in common in the uses the theories have for the term. This, we suggest, is precisely the sense in which a genuine amelioration will preserve subject matter, in addition to — and often thereby — resisting the remaining criticisms that non-ameliorative interventions face.

---

<sup>21</sup>See (Haslanger, 2013) and (Díaz-León, 2018).

## 2.4 Conceptual possibility and the Turning-Out Test

To further shore up the claim that epistemic responsibility is the unifying feature that distinguishes genuine ameliorations, let's consider the outer limits of conceptual change. Think of the skeptic who, responding to a proposal to allow polygamy, objects that it's just not *possible* for the concept [*married*] to apply to three people; whatever other merits of the proposal, adopting it would involve accepting beliefs that are tantamount to nonsense. To be clear, this particular worry issues not from theoretical convictions about the nature of concepts, nor from doubts about whether the proposed shift is analogous to a scientific revision that allows us to better describe and understand the world, but comes instead from an intuitive judgment about what is, and what isn't, conceivable. Such an objector might think, for instance, that although our concept of [*atom*] has changed drastically over time — after all, we no longer take atoms to be indivisible — there's simply no way that the concept [*atom*] could come to apply to the planet Jupiter. Something similar, the objector worries, is happening in the proposed change to [*marriage*].

One neutral way to phrase the concern is in terms of claims about *conceptual possibility*: that it is *conceptually impossible* for Jupiter to be an atom, or for three women to be married. Spelling out what conceptual possibility amounts to is a notoriously difficult task, but we can begin with a helpful tool proposed by Yablo (2008):

*The Turning-out Test*: It is conceptually possible that  $S$  iff some world  $w$  is such that it would have turned out that  $S$ , had  $w$  turned out to be actual.

The *Turning-Out Test*, as we understand it, doesn't supply a definition or substantial analysis of what conceptual possibility is. Rather, it provides a device for determining whether any particular sentence  $S$  *expresses* a conceptual possibility: if we can describe a world  $w$  that, had it been actual,<sup>22</sup> would lead us to claim that things would have turned out to be  $S$ , then  $S$  is conceptually possible.

---

<sup>22</sup>Or perhaps better, *if it is actual*.

Let's take this test as a working hypothesis: that is, to test whether a particular proposal is conceptually possible, we can plug it into the turning-out test and see what we get. For example, the question of whether it's conceptually possible for three men to marry becomes the question: is there some way the world could have turned out such that, if the world had turned out that way, it would have turned out that three men were married?

To get intuitions going, it's useful to consider some neutral examples. Hesperus is the first starry object to appear in the evening sky; this, as it happens, is Venus. Phosphorus is the last starry object to disappear from the morning sky; this, as it happens, is also Venus. Yablo points out first that conditionalizing on *purely linguistic* changes will never result in the world having turned out differently: if "Hesperus" had been used to refer to Mars, it would *not* have turned out that Hesperus isn't Phosphorus; our linguistic practices have no effect on the planets. Does this mean that it's conceptually impossible for Hesperus to be distinct from Phosphorus? No, for there *is* a way the world could have turned out in which Hesperus should have turned out not to be Phosphorus: one in which the first starry appearance in the evening sky was caused by Mars. If it had turned out that the celestial appearances that occasioned the use of "Hesperus" were caused by Mars, while those giving rise to "Phosphorus" were caused by Venus, then it *would* have turned out that Hesperus was not Phosphorus.

Note what's happening in these two conditionals. In the first, a way the world might have turned out (in Yablo's terms, a *counterfactual* scenario) is specified by describing a *purely linguistic* change; the world, we imagine, is just as it is, except that we use the word "Hesperus" differently. This antecedent doesn't lead to *our* adopting the proposed change when we evaluate the consequent. By contrast, the second counterfactual scenario is characterized by describing a difference in empirical facts, *viz.*, those causal factors at work in the initial reference-fixing of the names. And this way of picking out a counterfactual world *does* trigger a modification in our own usage in the consequent: we grant that "Hesperus" counterfactually refers to Mars.

Yet it would be hasty to characterize the difference as one of whether the antecedent specifies a merely linguistic change as against more substantial worldly change. To see this, consider two more examples.

*Virus*: It's conceivable that gradually the scientific community comes to see reproduction and stimulus response as much more central to what it takes it to be alive than metabolism and cellular organization; as part of the ordinary regimentation of scientific language, "life" begins to be used in a way that decides some borderline cases, such as viruses. If it turned out that scientists applied "alive" to viruses, then viruses would turn out to be alive.

*Goldberg*: It's conceivable that gradually the scientific community comes to see possession of a intricate, evolving structure and capacity to elicit psychophysical response as defining features of life; as a result "life" begins to be used in such a way that it's applied, among other things, to Bach's *Goldberg Variations*. If scientists came to apply "alive" to the *Goldberg Variations*, then the *Goldberg Variations* would not turn out to be alive.

In both the *Virus* and the *Goldberg* cases, we imagine a change in classificatory practice among the scientists whose job it is to classify living things. Both cases follow the same formal structure: we are first given a backstory supplying some relevant context, followed by a conditional. Importantly, both cases culminate in a "turning out" statement couched in terms limited to what we've been calling merely linguistic considerations: the antecedents of the respective conditionals refer, in both cases, to a conceptual intervention. Nevertheless, the cases have very different outcomes. The *Virus* situation is like the second Hesperus scenario: given this change in classificatory practice, we're happy to adopt that practice in our evaluation of the consequent. The *Goldberg* case, on the other hand, is much closer to the first Hesperus conditional: given this alteration in classificatory practice, we're *not* tempted to adopt it when we evaluate the consequent. The *Virus* conditional is a witness to the conceptual possibility of viruses' being alive; the *Goldberg* conditional, in contrast, does not establish the conceptual possibility of the *Goldberg Variations*' being alive. The

difference in the plausibility of the two cases seems to lie precisely in the sort of information that's supplied as background context, in information about *just how* the linguistic change comes about.

What we would like to suggest here is that what governs, in these cases, our willingness to adopt such a change *just is* our judgment about epistemic responsibility. If we describe a counterfactual situation in which our concepts shift in an epistemically responsible way, we find ourselves willing to try on the revision; if we describe the situation in a way that makes it clear that the imagined change is epistemically irresponsible, we hold on to our current meanings.

It's worth explicitly acknowledging, at this point, a somewhat subtle feature of this picture. "Conceptual alteration" and "conceptual amelioration" do not, strictly speaking, categorize pairs of concepts, or even pairs of concepts with respect to a term; rather, they apply to *ways* of shifting what a term expresses: being an amelioration is a *path-dependent* property. The broad question of whether a term's concept can be ameliorated into another is a question of whether there is *some* way for an epistemically responsible shift in that term's usage to take place that results in the term expressing the new concept (or a new version of the old concept); to establish that an amelioration is possible, the engineer needs to describe such a path.<sup>23</sup>

This allows us to understand an important distinction between ameliorations that are possible in the broadest sense, and those that are *practically* possible: the first is answered by reflecting on *metaphysically* possible paths our usage could travel, while the second restricts its attention to paths that are attainable given what we take to be practically possible. To put this point in terms of conceptual possibility, there are actually distinct senses of "conceptually possible," relative to different restrictions on the space of ways the world could turn out.<sup>24</sup>

---

<sup>23</sup>It's worth noting also that this contrast establishes that the change envisioned is not "mere stipulation" by a privileged group of language users. Even if we are in fact disposed to defer to the judgments of biologists about what life is, in the Goldberg case, we'd be tempted to think the biologists had gone off the rails: we'd not take their judgments to be binding.

<sup>24</sup>Our proposal is also compatible with lines of thought found in Richard and Yablo, among others, according to which there are hardly any analyticities found in ordinary language, and thus hardly any conceptual impossibilities at all. If that position is right, then it will emerge that virtually any proposed change could be made to pass the turning-out test, as long as the background conditions

Thus, the conceptual engineer has a clear general strategy for responding to claims of conceptual impossibility. If she can describe a means to shift a term's usage to the target concept that would be epistemically responsible, then in so doing, she will be passing the *Turning-out Test*: she will be describing a way the world could turn out such that the concept would turn out to apply in this revised way. Yet this, by the Yablovian test, is all she needs to do to confirm conceptual possibility: if there is a way the world could turn out such that the concept would turn out to apply in the revised way, then it's conceptually possible for it to do so.

In this way, our account takes seriously the objector's worry: there may well be certain proposed conceptual interventions that in fact *would* violate the bounds of conceptual possibility, and would for that reason be unacceptable. At the same time, our account provides a means for testing where any particular proposal stands with respect to conceptual possibility: we are suggesting that if an adequately described account of the proposed change passes the turning-out test, the proposed change is one that falls within the bounds of conceptual possibility.

## 2.5 Is amelioration enough?

We have been defending conceptual engineering against a kind of Orwellian unease: the sense that attempts at deliberate conceptual change are tantamount to changing the subject at best, and irrationality at worst. We've responded by pointing to a requirement on conceptual intervention that would ensure that a proposed conceptual change is no more epistemically reckless than our best processes of ordinary empirical inquiry. But now we face discomfort from the other direction: the ameliorationist might worry that the account on offer is too restrictive, in effect ruling out radical ameliorative proposals. If we restrict the class of legitimate ameliorative projects to those that are analogous to scientific cases and pass our Yablovian test, we risk

---

are specified in the right way. But because we distinguish between conceptual impossibility *tout court* and conceptual impossibility relative to a practically possible direction our world could take, there is still robust sense to be made of conceptual changes that are epistemically responsible and those that are not.

rejecting proposals that we might have reason to accept.

There are two elements of our response. First, it's no part of our account that we ought never to adopt a proposal for conceptual intervention that doesn't rise to the level of a genuine amelioration; perhaps, on balance, a proposal that would fall short of epistemic responsibility would have good enough moral consequences as to outweigh the disadvantages.<sup>25</sup> Moreover, whether such a case will arise seems to be largely an empirical matter. Consider two proposals around a derogatory term like 'slatternly': one which advocates simply eliminating the term altogether, while the other advocates instead using it to express [*praiseworthy*]. If we think that there's no practical way to epistemically responsibly bring about the latter proposal, then we might think that the first proposal would have all the (non-epistemic) moral consequences of the second, with the added benefit of incurring fewer epistemic costs. But if the latter proposal is, as a matter of sociology, more likely to succeed, this fact may weigh in favor of the replacement over the elimination. Similarly, we might think that in some cases, there's a moral and political benefit to engineering a word's concepts rather than eliminating the word, in allowing the coöpting of inferential patterns.

Second, it is important to recognize that imposing a constraint on analogy to that applied to scientific practice may not be as restrictive a requirement as one might initially expect. We might at first suspect that our epistemic responsibility constraint would summarily rule out, for instance, proposals that depend upon judgments of moral value or of effectiveness at attaining specific ends. But this conclusion would be hasty. As feminist epistemologists such as Elizabeth Anderson (1995) and Helen Longino (1990) have pointed out, values and practical purposes are part and parcel of the practice of scientific inquiry itself. We can see one relatively weak way in which this is true by observing that whenever purely epistemic considerations fail to decide between competing best available theoretical revisions, and when we're called to adopt *some* revision, there's no purely epistemic bar to using other factors to choose which to adopt. Additionally, such theorists have argued, there's a deeper sense in which

---

<sup>25</sup>Teasing out this thought more precisely is a complicated matter best done elsewhere; how to characterize it will depend, for instance, on whether we should understand such an epistemic hindrance as a moral wrong in itself.

it's true. Since facts about which questions we count as worth investigating, which results we find interesting, which discoveries we're hoping to make, and which features of nature we wish to model all depend upon our purposes and our values, these shape the elements of our scientific practice: features such as the theoretical methods we use, the terminology we adopt, the evidence we expose ourselves to, the idealizations we utilize, and the distinctions we're apt to make between core and outlier cases. And of course the shape of our scientific practices directly impacts how our theories develop; far from being alien to scientific practice, considerations of value and purpose are integral to it, according to this line of thought.

We needn't even go this far, however, to address the worry that value has no place in ameliorative proposals. For there are clear examples of sciences that *explicitly* consider values and goals in evaluating revisions; medicine, meteorology, and community psychology are all examples. And, irrespective of whether these are "fundamental" or "pure" sciences, it's clear that the sort of constraint we're calling "epistemic responsibility" is active in these fields, when evaluating terminological revisions.<sup>26</sup>

The same is true in the case of cross-domain term usage, as well. Consider the closely-related enterprises of gardening and botany. Each has an interest in the length of plants' lifecycles, and classifies plants as annuals (completing the lifecycle in one year), biennials (taking two years to complete the lifecycle), or perennials (living longer than two years, often indefinitely). But the gardener is interested in the plant as a production mechanism; she is apt to count the lifecycle as complete when the plant has matured to the point of producing what she wants to harvest. The botanist, on the other hand, is interested in what the plant does on its own: when does it mature and decline, biologically? As a result, plants that count as biennials to the botanist — because they take two years before reaching biological maturity — may count as annuals to the gardener — because they grow from seed to mature harvest within one year: carrots and onions are common examples.

That an ameliorative proposal avails itself of value and purpose considerations, then, is no mark against its claim to preserve subject matter.

---

<sup>26</sup>See (Lange, 2000).



A related objection might be sketched: not every beneficial, desirable, or otherwise legitimate instance of conceptual engineering will necessarily be recognized as an amelioration from the perspective of most current concept users. Consider the following example. Had Democritus been asked to reflect on whether the world could ever turn out in such a way that the concept [*atom*] could be properly applied to divisible units of matter consisting of charged and uncharged particles, he would likely have resisted. A case may be made that his resistance needn't be due to empirical ignorance or failure of imagination: even if we were to describe to him the entirety of scientific development that took place from his time to ours, he might still deny that our concept [*atom*] is continuous with his own. We might imagine, for instance, that he'd go along with our story until the nineteenth century, but on coming to Cantor's Theorem, he'd put his foot down: "Something's gone awry," he might say. "I don't know what, but this idea that some infinities can be larger than others is just nonsensical." But this makes it impossible for him to accept functional analysis, differential geometry, and the mathematical foundations of twentieth-century physical theory, leaving him with no basis to accept the new understanding of [*atom*]. We might characterize his reluctance in terms of his epistemic standards: *our* standards, but not his, deem acceptable the shift in our understanding of infinities. Perhaps, then, certain conceptual changes just can't be conceived of in advance, or even rationally argued for; perhaps some shifts can only happen gradually, as a community of actual language users lives through actual change. For some shifts, that will be precisely because they change the standards by which we judge epistemic responsibility in the first place.

But, the argument continues, surely the case can be made that the shift to our concept [*atom*] from that of Democritus *is* a *bona fide* amelioration: there is a rational, continuous, and epistemically responsible shift from his understanding of [*atom*] to ours, despite prodigious and profound differences in what we take to be true of them. At the very least, it seems a mistake to rule out the possibility of such continuity based solely on his resistance to the proposal, issuing as it does from an epistemically normative objection that we recognize as mistaken. Yet if this is a case in which we

are in a position to recognize responsible conceptual development while Democritus is not, then surely we, too, may sometimes find ourselves in Democritus's shoes, unable to recognize legitimate amelioration when we see it. So, the worry concludes, a restriction on conceptual amelioration based on epistemic responsibility is much too limiting, liable to deliver false negatives for certain legitimate proposals.<sup>27</sup>

But reflecting on Democritus's situation shouldn't distress us too much. For the engineer's point in aiming at conceptual *amelioration*, over less constrained forms of intervention, is not to guarantee accordance with some highly specific metaphysical, mental, or linguistic structure, but instead to assuage a certain disquietude that the change would make for too drastic a shift in use. So we should ask: whose fears need the ameliorationist address? The answer: ours. And to this end, she needs to concern herself not with any possible set of epistemic standards whatsoever, nor with absolute standards, if there are such, but rather with *our* standards. Whether a change is, or is not, an amelioration, then, is a question that only makes sense in the context of a set of standards. Of course, relative to Democritus's standards, the path from his [*atom*] to ours is not an amelioration, whereas relative to our standards, it is; this is, perhaps, a surprising result, but it's not clear why this should be troubling at all.

There is, though, a more subtle lesson to be gleaned from this example. For we might further ask: if we don't all share the same standards, or if standards are subject to change, whose standards count? Expert standards now? Untutored standards after the change? Of course, it may well be that in the overwhelming majority of cases, this distinction will not be at issue, because epistemic standards among different audiences will be largely in agreement. But if they do diverge, then the answer, it seems, may be a strategic matter. If the conceptual engineer hopes to ameliorate a concept by encouraging widespread adoption of the new usage, then in most cases, it will be a severe *practical* disadvantage to the proposal if the mass of speakers can't be convinced that there's the right sort of connection between the new usage and the old; thus, she has good reason to attend to the folk epistemic standards at play

---

<sup>27</sup>This misgiving is equally strong from the other direction: certain proposals that we judge to be legitimate on the basis of the turning out test may later be discovered to have been mistaken.

initially.

If, on the other hand, she has an alternative vision for enacting the change — through policy or expert use, for instance — she may not be concerned with current folk standards at all. So, for instance, when Haslanger (2000, 35f.) says that the target audience for her proposed revision of gender and race concepts comprises feminist and anti-racist critical social theorists, she may only be interested in establishing that her proposal is, for *them*, an amelioration. And in certain special circumstances, such as those that arise with Haslanger’s proposed amelioration of [*knowledge*], we might think that adopting the new usage would impinge on our epistemic standards themselves. This fact alone does not give reason to think that the proposal is bound to fail; if its practical success requires only that, once we’ve carried it out, we’ll not judge it defective, then it’s amelioration *with respect to the post-engineering standards* that Haslanger should aim at.

What all of this shows is that our framework is not committed to devaluing truly revolutionary proposals. Rather, it offers a tool to the critical conceptual engineer: it shows her what she should consider, beyond the moral consequences, when she crafts her proposal.



## Chapter 3

# Performative Consent and Autonomy

In the first chapter of this work, I suggested a conception of consent that revolved around *speech acts*, or more generally, expressive acts. Something like this picture, the *performative model* of consent, has been defended within legal theory, as part of a long-running discussion pitting it against the *attitudinal model*, a conception of consent revolving around attitudes, or more generally, mental states.

In this chapter, I consider these two accounts, and argue that the kinds of ameliorative considerations discussed in the second chapter give good reason to prefer the performative model in legal contexts. I then show how the performative model can make sense of the most radical of radical feminist approaches to claims of sexual violence — that whether a sexual assault has occurred may, sometimes, depend upon a survivor’s after-the-fact assessment of what happened. Taken together, I believe these create good reason for liberals, and strong reason for radicals, to embrace the performative view.

### 3.1 The attitudinal model and its discontents

As Patricia Kazan (1998) puts it:

On the attitudinal account, consent consists in any one of a range of mental states, from desire to grudging acquiescence, held by the consenting agent.

It is a subjective matter whose determination involves the state of mind of the agent. The agent's behavior is regarded merely as an indication of consent — her actions are the objective evidence or tokens from which consent is inferred.

This attitudinal model, we might think, is the "default" model — it is the basic model that courts have used historically, for instance, when treating consent-related cases. It fits well our intuition that consent is, in core cases, about whether someone *wants* to be treated in a given way. We often speak of consent in a way that (at the very least) presupposes that the transformative normative power of consent emanates from, and is explained by, a particular sort of distinguished mental state. In particular, in both sexual and medical contexts, it seems natural to think that consent is morally transformative because it ensures that what's happening to our bodies is aligned with our will; if I am subjected to a surgery that I do *not consent*, my basic autonomy right to control over my own body is compromised. And it is clear, on the attitudinal model, where the room is for things like *informedness* or *agential capacity* to matter: if I am confused or, worse, misled about what the surgeon is going to make my body do, what I wish to be done to my body is likely not what will be done to my body; likewise, if the relationship between my (rational?) will and my body is already compromised because I am intoxicated or otherwise incapacitated, the alignment of will that consent consists in will not have the usual normative implications.

So the attitudinal account has a clear appeal. But as many feminist scholars and activists have noted,<sup>1</sup> the attitudinal model has had terrible consequences when used as a legal model. For if the crime of sexual assault consists in having sex with someone without their consent, and if whether someone consents is a matter of what they desire (or some variation thereof), then *whether a sexual assault occurred* depends upon *what the survivor wanted*, and so the survivor's mental state becomes a critical criminal element: as part of proving that a sexual assault occurred, the prosecution must establish that the survivor did not *want* sex. But this is then — rightfully so, if we demand consistency in the allocation of evidentiary burden across

---

<sup>1</sup>See Kazan (1998) for an excellent summary.

different crimes — subject to adversarial attack by the defense. The defense has only to establish reasonable doubt about the survivor’s (lack of) desire to merit an acquittal.<sup>2</sup> And so, predictably, sexual assault and rape trials end up “putting the victim on trial.” Survivors of sexual violence are hauled into court to be interrogated about their sexual history, their clothing styles, their decisions about where to walk, and an endless number of other trivial details, which interrogations are aimed at one thing: generating uncertainty about whether the survivor wanted the assault to occur. Survivors are forced to recount traumatic experiences while facing coordinated attacks on their integrity, their sexual purity, and their moral worthiness, all while their attackers look on from a front-row seat. Small wonder, the thought goes, that sexual assault is so vastly underreported and underprosecuted.

---

<sup>2</sup>This is obviously too quick; for one thing, different jurisdictions may differ in legal definitions of battery, sexual assault, and similar consent-related crimes. However, at common law criminal battery requires non-consent; *Black’s Law Dictionary* (2019) includes lack of consent explicitly in its definition of “battery” : “[t]he nonconsensual touching of, or use of force against, the body of another with the intent to cause harmful or offensive contact.” Even modern formulations of the tort of battery (consisting in harmful or offensive touching) understand consent to preclude a battery claim (see, e.g., *Restatement (3d) of Torts* (2010, Ch. 2)), as an extension of the common-law doctrine of *volenti non fit injuria* (see, e.g., (Warren, 1895)). The story about sexual assault is more complicated. Schulhofer (1992, pp. 61ff) notes that at common law, “rape” was defined in terms of consent. But Blackstone’s definition, “the carnal knowledge of a woman forcibly and against her will,” (4 *Blackstone Commentaries* \* 210 (2016)), includes both non-consent and force. Aside from cases of impersonation, these were virtually always understood by courts to be coextensive. Well into the mid-20th century virtually all statutory definitions explicitly required both bodily force and resistance, and the prevailing legal standard required that a survivor have resisted “to her utmost” (Tchen, 1983, p. 1518). Modern legislative reform of sexual assault laws was the Model Penal Code of 1962, which for the first time allowed for threat of force to be considered in addition to actual force, but at the same time dropped the consent language altogether (Tchen, 1983, p. 1529). The contemporary reversion to a consent-based framing, and more particularly a standard of “meaningful consent,” is the result of hard-fought battles by a tenuous alliance of feminists, law-and-order reformers, and victims’ rights advocates in the 1970s, which eventually resulted in the adoption of new sexual crime laws in all the states and in Congress (see Bachman and Paternoster (1993, 558ff)). Even after these reforms, however, courts were prone to revert to force-based tests centered around violence rather than consent. As Schulhofer puts it: Of course, legal protection of “meaningful consent” (as distinguished from protection against violence) was never more than superficial. No sooner read out of the definitional elements and organizing principles, force quickly resurfaced as the only generally reliable indicator of what nonconsent could mean. Even the [Model Penal Code] Commentaries, a relatively modern effort to criticize and reform the law of rape, jump quickly and with no sense of contradiction from the premise that rape centers on “the idea of meaningful consent” to a view that “necessitates the drawing of a line between forcible rape on the one hand and reluctant submission [sic] on the other.” (Schulhofer, 1992, p. 63, internal citations omitted, “*sic*” original)

## 3.2 The performative model

If willingness — a mental state — is what consent is all about, it's difficult to see how to remedy this without major rewrites of the law (such as removing *consent* from the elements of criminal battery, assault, and rape). However, feminist legal theorists have argued that this problem only emerges because it rests on a mistaken conception of consent. Consent is not, according to theorists such as Nathan Brett,<sup>3</sup> about attitudes or mental states at all, but rather about *expression*. Whether someone consents, according to this performative model, is not about their secret, unobservable desires, but rather about their public, observable actions. Consenting is a speech act, whose illocutionary force gives permission. As Brett (1998, p. 69) puts it, “To consent is to give permission . . . This means that the question of consent is *not* just a question about the state of mind or attitude of the complainant. Rather, the matter which should be central to a court's consideration of consent is the question of what was said or done that could be construed as granting permission to do the acts in question.”

Where the attitudinal account draws on an intuitive parallel between consent and desire, the performative account draws on an intuitive parallel between consent and permission. The “moral magic of consent”,<sup>4</sup> on this account, is grounded in the force of the speech act of giving permission. Giving permission is a speech act like promising; like promising, it shifts the moral landscape. Indeed, we can think of giving permission as a sort of logical complement to promising. Where promising creates new obligations, permission removes them; contrapositively, where promising makes the permissible impermissible, permission makes the impermissible permissible.<sup>5</sup> A promise from someone creates a claim against them; permission to someone removes one (see Thomson, 1990, ch. 14). Tacit permissions, like tacit promises, are tacit in the sense that they are expressed by means other than words, taking into account context, norms, and the usual Gricean bells and whistles; they are *unspoken*, but not

---

<sup>3</sup>Kazan (1998) recognizes Brett as the central defender of this view, in early work given as talks but unpublished. It has since been taken up by others, and is articulated in writing in Brett (1998).

<sup>4</sup>Stolen from the title of (Hurd, 1996).

<sup>5</sup>Though this is not quite the tautology it appears at first blush, it has an undeniable ring of plausibility.



*unexpressed*. And while the norms of promising may be more regular than those of permitting (Brett, 1998) there *is no such thing* as unexpressed permission, just as there *is no such thing* as an unexpressed promise.

The performative model of consent offers a solution to the otherwise daunting problem of compounded harm to the survivor in a sexual assault trial. Under this model, we don't need to revise legal definitions that take (non-)consent to be central to the crime. If consent is not about the survivor's mental state, but about what she expressed, we can avoid inquiry into what the survivor wanted. The anti-rape slogan “ ‘no’ means ‘no’, ” on this view, is not merely good advice about making inferences about someone's desires given his words; if he says ‘no’, he does not consent, and having sex with him is sexual assault.

### 3.3 Challenges to the performative model

Now of course, it's doubtful that the performative model would (or does) actually have this effect when adopted into law. In particular, “‘no’ means ‘no’” is, strictly speaking, false: people can consensually engage in sexual role-play that involves saying ‘no’ — and while some may believe that rape-fantasy role-play is morally suspect,<sup>6</sup> it is certainly not *a kind* of rape. This is not a problem for the performative account as such — since, as noted above, it is *the expression of permission (or its denial)*, rather than *a particular set of words*, that does the moral work. Contextual factors can alter what is said (and what is done) by uttering what would ordinarily constitute permission or refusal; it seems obvious that sexual role-play involves these contextual factors (and explicitly considers them through devices such as safe words). This is of course true of promising as well: an actor who says “I promise to take care of you” in a film is (ordinarily) not truly promising anything. But once we acknowledge this, it seems much less evident that legal adoption of the performative model would (or does) avoid “putting the victim on trial;” if what was said (and done) by an utterance

---

<sup>6</sup>To be clear, I do not endorse this view, and I think it runs the risk of shaming survivors of sexual trauma. See (Billock, 2018).

is in dispute, further inquiry into contemporaneous desires, prior encounters, and habitual modes of expression may be part of a diligent contextual inquiry.

Patricia Kazan makes a similar point when critiquing Brett's performative account. As she puts it,

[F]eigned or coerced consent may be behaviorally indistinguishable from genuine consent. Such cases emphasize the need to appeal to an agent's attitudes in order to determine the matter of consent. (p. 35)

However, I differ from Kazan in what I take to be the lesson of the examples. For Kazan, the fact that there are contextual factors that include facts about the mental states of the parties *undermines* the motivation for the performative account, because of Brett's insistence that consent questions be resolvable without recourse to "subjective mental states" (Kazan, 1998, p. 36). While this seems right as a critique of Brett's particular view, it does not seem to trouble the performative account as such.

In fact, the point generalizes far beyond the case of consent: in general, in order to know what is expressed by an expression, we need to know something about the expressor's state of mind: that she is not being sarcastic, that she is not play-acting, that she is not quoting someone, that she is speaking our language.<sup>7</sup> Quite generally, *any* performative account of *anything* is going to require some degree of inquiry into performers' mental states; this is an unavoidable part of an analysis of communicative acts.

While Kazan is right that there is a "subjectivity" involved in the performative account, this subjective inquiry is of a quite different sort from the subjective inquiry demanded by the attitudinal account. On the former, it is one factor that is sometimes worth considering as evidence about what was expressed; on the latter, it is a core part of establishing that a crime was committed. *Pace* Kazan, there seems to be little reason to believe that legal processes under the performative account would involve

---

<sup>7</sup>Bob Stalnaker raises related points in "Assertion Revisited" (2004, pp. 295ff): "the truth value of [an] utterance will depend on the facts in two different ways: first, the facts determine what is said; second, the facts determine whether what is said is true."

*harmful* inquiry into survivors' mental states; what's demanded by the performative account is exactly enough inquiry to establish *what was expressed*.

Moreover, it matters that here we are concerned with definition within a legal context. Legal standards for establishing consent or its denial may come apart from ordinary-language intuitions about consent; it may well be that the all-things-considered balance of reasons permits, or even requires, adopting a legal definition that is stronger than the true definition, in the sense that there are possible (genuine) expressions of consent that will be counted as illicit by the legal standard. In particular, we may anticipate that there will be systematic bias against women and survivors in the application of whatever legal standard is adopted. It seems to me that adopting a too-strong legal standard may be a permissible way of countering this bias. Thinking of a trial as a decision procedure, this would, in effect, be a kind of regularization, analogous to the widespread practice of adjusting for known bias in an algorithmic classification by intentionally modifying the objective function to compensate. If historically, legal standards have served to privilege and protect (overwhelmingly male) assailants to the detriment of (predominantly female) survivors, perhaps it's in the interest of justice to give the pendulum a push in the other direction.

Glibness aside, the point here is that the question of how *the law should define consent* is not self-evidently coextensive with that of *what consent is*. There is, of course, a separate question of whether adoption of an intentionally overbroad definition of consent-related crimes or causes of action would itself be lawful.<sup>8</sup> In particular, it would seem to be a perverse way of sneaking a redistribution of evidentiary onus through the back door. We might also worry about whether legally defining 'sexual assault' in a way we all acknowledge to be dubious is blocked by fundamental constitutional due process and justice considerations.

There are really two versions of this worry. The first is pragmatic:<sup>9</sup> because

---

<sup>8</sup>Here and elsewhere I restrict my attention to the United States. This should not be taken to imply that similar problems do not exist elsewhere.

<sup>9</sup>More explicitly: worries about the constitutionality under prevailing United States law of this sort of proposal are relevant neither because the United States Constitution has some special kind of normative authority nor because it has some special kind of conceptual authority. Rather, they are relevant because many of the putative benefits of courts or legislatures adopting the performative account depend precisely upon such an adoption's legal efficacy. It would be foolhardy indeed to

such a revision would be (held to be) unconstitutional, it would not be efficacious. Although there is a *prima facie* plausibility to this worry, it is unconvincing precisely because it seems overwhelmingly doubtful that courts would invalidate such a move. Courts have permitted far more radical revisions — as, for example, in statutory redefinitions of murder to allow prosecution of killing by drunk driving as murder. And generally, the overarching trend in contemporary United States jurisprudence gives legislatures tremendous leeway in defining crimes<sup>10</sup> and broad power to control courts’ jurisdiction, evidentiary standards, and sentencing practices of courts.<sup>11</sup>

The second is philosophical: we might worry that the intentional adoption of an incorrect definition of “consent” would undermine the law’s moral authority or moral legitimacy. A detailed discussion of this worry would take us too far afield. Nevertheless, I’ll offer my response without detailed argument: first, it’s contentious that the law *has* moral legitimacy or authority; second, it’s not at all obvious that whatever legitimacy or authority it does have is undermined by the sort of definitional move under consideration here; third, the balancing of enforceability and other consequential considerations against moral intuitions in crafting law and policy is a perfectly ordinary and widespread practice — so if this move is a problem on these grounds,

---

advocate adopting a proposal, knowing that it would be immediately vitiated.

<sup>10</sup>See, e.g., Justice Brennan’s discussion in *Taylor v. United States* (1990) of the staggering breadth of state definitions of ‘burglary’, as well as the more recent line of cases addressing vagueness challenges to the violent felony provision of the Armed Career Criminal Act: *Begay v. United States* (2008), holding that felony drunk driving is not a violent felony for the purposes of ACCA; *Chambers v. United States* (2009), holding that failure to report for jail is not a violent felony for these purposes; *Johnson v. United States* (2015), striking as unconstitutionally vague ACCA’s “residual clause” extending application to actions that present serious risk of injury; this trend is admittedly complicated by the most recent term’s ruling in *Stokeling v. United States* (2019), in which Justice Thomas’s majority opinion draws on both legislative history and common-law understandings of burglary to determine that burglary is a violent felony.

<sup>11</sup>See especially *Felker v. Turpin* (1997), upholding Congress’s power to jurisdictionally bar *habeas corpus* petitions traditionally allowed by the courts, and *United States v. Booker* (2005), in which Justice Breyer’s controlling portion of the fractured opinion holds in part that the Federal Sentencing Guidelines are not facially invalidated by the Sixth Amendment, and in which Justice Stevens’s controlling portion acknowledges that the respondent’s Sixth Amendment challenge to the Federal Sentencing Guidelines as applied would not invalidate the overwhelming majority of sentences dispensed under the Guidelines. See also the Court’s rulings on the Religious Freedom Restoration Act, which was explicitly adopted in order to nullify a prior Supreme Court Ruling (*Employment Division v. Smith* (1990)). In *Gonzales v. O Centro Espírita Beneficente União do Vegetal* (2006), the Court declined to strike down RFRA as unconstitutional on separation of powers grounds; the Court reiterated this holding in *Burwell v. Hobby Lobby* (2014).

we have much bigger legitimacy and authority worries on our hands.

All this is to say: while the initial claims for the revolutionary potential of a performative conception of consent may be a tad overblown, some version of the consequential claim made by its advocates seems entirely plausible: adopting this definition would (and in some cases has) result in fewer cases of victim-shaming, more prosecutions for sexual assault, and a net win for gender equity.

It seems, then, that we have a prime candidate for an ameliorative justification on our hands. We have two definitions of “consent,” the attitudinal and the performative. Both have intuitive appeal: the former because of consent’s similarity to *desire*, the latter because of its similarity to permission. And there is good reason to believe that adopting the latter over the former will have better consequences for justice.

This last, it’s worth noting, does *not* establish that, in Alex Byrne’s words, “there is a danger that one’s ‘intuitive’ assessment . . . will be clouded by non-epistemic factors” (Byrne, 2020). My point here is that our intuitions — or, more precisely, *my* intuitions — pull in both directions. Despite my enmeshment in a milieu of radical feminists, angry queer and trans\* activists, and Black Lives Matter demonstrators, I’m happy to report I feel no great threat of being shunned as a “wrongthinker” if I defend the attitudinal account.

At the same time, it is important to recognize that the justice considerations here are *appropriate* considerations, for precisely the reasons I discuss in the previous chapter. I see no reason to accept that we shouldn’t care about the justice consequences of a definitional choice.<sup>12</sup> There is tremendous, real harm done to survivors of sexual violence, and to potential victims of sexual violence, by a legal system that adopts an attitudinal account; *of course* we should care about this harm — not caring is not virtuously detached, it’s sociopathic. Indeed, this sort of case is the easiest case for amelioration: if we have good reason to believe that adopting the performative model would make the world better, and *in addition* there is no clear winner between the accounts as far as purely epistemic considerations<sup>13</sup> are concerned, using the

---

<sup>12</sup>To be clear, I do not read Byrne claiming, in the gender case, that we shouldn’t care about the consequences of definitional choice.

<sup>13</sup> “Purely epistemic considerations” can, of course, include moral considerations. As discussed

(epistemically inert) consequentialist consideration<sup>14</sup> as a tie-breaker seems entirely appropriate.

### 3.4 The disjunctive option

That is, we have a candidate for an easy amelioration *if* there aren't even better options around. If there is a third alternative that better fits our intuitions than either of the two above, an ameliorative justification of the performative account may still be possible, but it won't be easy in the sense above: the value considerations will necessarily function as more than mere tie-breakers.

So is there such an option? Vera Bergelson (2014) offers one: a disjunctive account that takes consent to be sometimes attitudinal and sometimes performative. She offers a series of hypotheticals meant to show this; two are relevant here (renumbering from the original, p. 174):

1. Polly is secretly in love with Mick and would like to have sex with him. At a party at Mick's home, she sneaks into his bedroom, and pretends to be completely drunken and asleep. Mick finds Polly in his bed and has sex with her under the mistaken impression that she is not aware of his actions. In the end, both are quite happy. Did Mick commit sexual assault?
2. Polly hates her long nose. She would love to have a nose job but she is too proud to openly admit that. In fact, she has a secret wish that her fiancé Mick (who is a plastic surgeon) single-handedly put her under anesthesia and fix her nose. Mick also hates Polly's nose but he is not aware of Polly's inner thoughts.

---

in the previous chapter, the question here — which understanding ought we adopt? — requires considering the epistemic responsibility of the proposals, and, in particular, goodness of fit with our intuitions. When considering moral concepts, this will invariably require checking goodness of fit with our moral intuitions.

<sup>14</sup>Of course, consequentialist considerations aren't always epistemically inert. In the *Brain Tumor* case of the previous chapter, the epistemic consequences, broadly understood — that is, the consequences for the effectiveness of our scientific inquiries — are part of what makes the change epistemically responsible (thanks to Steve Yablo for this point). Here I'm assuming for the sake of argument that the comparative harms to survivors of the competing consent models are not epistemically relevant in this way.

Nevertheless, one day he puts Polly under anesthesia and performs a surgery on her. Outraged by his conduct, his nurse reports him to the authorities. Is Mick guilty of battery?

Our intuition is supposed to be that in (1), Mick does not commit sexual assault; in (2), Mick does commit battery. This is surprising, because there seems to be a structural symmetry between the two. In each of them:<sup>15</sup>

- (a) Polly has a hidden desire that Mick *A*, and so
- (b) Polly consents to Mick's *A*ing on the attitudinal model;
- (c) Polly does *not* express her desire to Mick, and Mick is unaware of her desire, and so
- (d) Polly does *not* consent to Mick's *A*ing on the performative model;
- (e) Mick's *A*ing is criminal<sup>16</sup> if and only if Polly does not consent to it;
- (f) Mick *As*

Now it follows from (e) and (f) that Mick commits a crime if and only if Polly does not consent. So if Mick's *A*ing is criminal in case (1), where *A* is *having sex*, but not in case (2), where *A* is *performing surgery*, it seems that:

- In case (1), Polly's pro-*A* attitude suffices to show that she consents, and her lack of pro-*A* expression does not undermine this consent;
- In case (2), the reverse is true: Polly's lack of pro-*A* expression suffices to show she does not consent, and her pro-*A* attitude does not undercut this.

---

<sup>15</sup>I have here condensed Bergelson's argument into what I take to be its minimal form.

<sup>16</sup>To avoid any confusion about whether this is philosophical or legal analysis, it's worth emphasizing the role that criminality is playing here. Although Bergelson writes in the context of legal theory, and so is predictably unconcerned about this distinction, in this reconstruction I'm stipulating that the legal question "is Mick's behavior criminal?" is wholly determined by the factual question "does Polly consent?" In particular, I'm intentionally not translating this argument into purely non-legal moral terms, because I take it to be quite obvious that Mick's behavior is seriously wrong in both cases.

Since in both cases, Polly consents on the attitudinal model, and does not consent on the performative model, Bergelson concludes, neither model can be adequate; the correct theory of consent should say not that (only) her attitude matters or that (only) her expression matters, but that (only) her attitude matters in case (1), and (only) her expression matters in case (2).

What's the difference between the cases? Given the formal equivalence, the only room for difference lies in *A* itself and the related crime. So, Bergelson concludes (177): "different models of consent should be used depending on whether the role of consent in a particular case is inculpatory or exculpatory."

Now there is room to push back on this example from the very beginning. For one thing, I'm not sure I agree at all with her intuitive assessment, that in case (1) Polly consents in a way that renders Mick's behavior non-assaultive. For our purposes here, the question isn't resolved by observations about how the law *is in fact* applied — it may well be applied *poorly*. Rather, we must ask: if the law (rightly or wrongly) criminalizes sex without consent, does Mick's behavior count as criminal? While I do share the intuition that there is something *somewhat less disturbing* about Mick's behavior in case (1) than in case (2), that's about it: it's not at all obvious to me that Mick's behavior isn't an assault in (1).

Moreover, the disjunctive proposal is only plausible if there's an independent way of understanding the key differentiator, "inculpatory." But it's not clear that there is. Bergelson begins with the characterization: "It is inculpatory if it is an element of an offense; it is exculpatory if consent is a defense." She continues,

In the first instance, the perpetrator's act is prima facie morally neutral; it becomes criminal due to the attending circumstances (non-consent). Theft, rape, trespass, and kidnapping provide examples of the [*sic*] inculpatory non-consent. Absent the victim's non-consent, there is nothing wrongful or regrettable in the act of taking the property of another or having sex or visiting someone's home. . . . In contrast, when consent plays an exculpatory role, the perpetrator's act is prima facie wrongful. . . . Homicide, maiming, and battery are examples of such conduct. (p. 177)



Whether the role of consent is inculpatory, then, is a matter of how the crimes are *formally defined*, and of what defenses are allowed under the law. Take trespass, for instance: it is not a crime to walk across someone's land; it is a crime when they have posted "No Trespassing" signs. Intrusion is criminal only when consent is explicitly denied. In the case of battery, on the other hand, it *is* a crime to cut into someone with a knife; it is *not* a crime in special circumstances such as consensual surgery. Cutting into someone is non-criminal only when consent is explicitly given.

This much makes sense. And it shows why the law should sometimes focus on the perpetrator, and what he knew, and why it should sometimes instead focus on the victim, and what she did. But it seems to me deeply confused to think that this is because of a difference in whether the victim consents. There's a much simpler explanation at hand: we have different default presumptions we make about consent absent other evidence. If it's sensible to presume someone does not consent to some *A* absent evidence to the contrary, then it's sensible to criminalize *A* and offer consent as an affirmative defense; if it's sensible to presume someone does consent to *A* absent evidence to the contrary, then it's sensible to criminalize only non-consensual *As*.<sup>17</sup>

Put another way, Bergelson's approach confuses cause for effect. I see no reason to think that we can't perfectly well explain Bergelson's observations by starting from what ought to be criminalized, and moving to how consent should be legally incorporated. And doing this doesn't require moving to a disjunctive notion of consent.

Moreover, it makes clear that the disjunctive notion has an important drawback: it makes consent depend upon legal contingencies in a way that seems implausible. Surely, part of why non-consent is an element of trespass is that we have social norms that default to permitting free movement across land. Had our social defaults been switched, we'd expect that the inculpatory standard for trespass would have been switched as well: instead of making trespass a crime only when non-consent is expressed, we would make trespass a crime unless consent were expressed. And in this switched case, it seems that *nothing differs* when it comes to consent to having

---

<sup>17</sup>If this were a law review paper, some further remarks about differential evidentiary burdens for affirmative defenses *versus* elements of a crime would be in order. Fortunately, it's not.

their land traversed.

### 3.5 Deflationism

So it seems that the performative model retains its claim to maximal (if not perfect) fit with our intuitions, and serves our ameliorative purposes far better than the attitudinal model. However, there is a final option to consider: rejecting a unitary notion of *consent*. Why shouldn't we, instead of taking the two models of consent to be different theories of *what consent is*, or even of *what consent is as used in the law*, take them instead to be articulations of *two different kinds of consent*?

In reply, some of the considerations in §5.2 of the first chapter are worth reiterating. There I pointed out that in many cases consent may be given, but we may nonetheless have morally objectionable (and morally objectionable on grounds of consent) behavior. This is so because the moral landscaping isn't done by consent *simpliciter*, it's done by consent *to the right thing*. In ever so many cases, we are apt to mistakenly focus on the question of *whether* consent was given, rather than the question of *to what* consent was given. To extend that thought here: it may be entirely right that there are two different kinds of consent, attitudinal consent and performative consent; this expands our possible space by a factor of two. But the question then remains: what kind of consent does moral work? And, in the legal case, what kind of consent (and when) should be contemplated by consent-related laws?

That is to say, whether we conceive of them as competing models of a unitary kind of consent or as different kinds of consent determines very little as far as the interesting questions go. Any of the considerations we've considered can be rephrased in a way amenable to the deflationary picture: rather than asking which model of consent the law should use, we instead ask which kind of consent the law should reference. As far as I can see, there is no deep matter to be decided here.

## 3.6 Radical consent

Everything thus far has been, I take it, consistent with mainstream liberal approaches to consent; as such, while I've taken care not to too deeply conflate legal questions with moral and semantic ones, I have not made any use of intuitions around consent outside the legal context.

In this remaining section, I'll depart from this restriction, and argue that the performative model has a further virtue: it makes conceptual space for a kind of claim advanced by radical feminists. Many such feminists profess a commitment to a sort of strong deference to survivors<sup>18</sup> when it comes to characterizing encounters. Here I focus on, and sketch a defense of, one claim made in this context: survivors' posterior assessments of whether they were sexually assaulted can determine whether they were in fact sexually assaulted.

It's important first off to distinguish this from weaker and more intuitively plausible claims. First, I am not claiming (merely) that we should listen to survivors, in the sense that we should stand against the systematic discrediting of sexual assault claims by casting survivors as confused, regretful, or vengeful. Of course we should, but I take that to be obvious. Second, I am not claiming that as a matter of law, policy, or personal habit we should take survivors' accounts of their own experiences as (defeasibly) determinative — that, except perhaps in extremely narrow circumstances, we should take a survivor's claim that she was assaulted as overwhelmingly strong evidence that she was assaulted. I believe there may be good reasons for this position, but this is not what I'm addressing here. Third, I am not claiming that, as an epistemic matter, it may be in some cases impossible to know whether someone has been assaulted until they retrospectively reflect on the encounter. This is true, and will follow from my claim, but it is a strictly weaker claim.

What I am defending here is, rather, the superficially crazy view that

---

<sup>18</sup>Here and in the sequel, I'll use 'survivor' to mean 'potential survivor' — that is, in cases where there is a sexual assault, it should be understood to indicate the person who is assaulted; in cases where it is indeterminate whether an encounter is a sexual assault, it should be understood as denoting the person who is assaulted if there is an assault; in cases where there is not, it should be understood as denoting the person about whom there is most doubt as to whether they were assaulted.

(PA) There are (possible) sexual encounters for which,

1. at the time of the encounter, there is no determinate fact of the matter as to whether the encounter constitutes sexual assault,
2. it is possible for the survivor's after-the-fact evaluation of the encounter to make it the case that the encounter constituted sexual assault, and
3. it is possible for the survivor's after-the-fact evaluation of the encounter to make it the case that the encounter did not constitute sexual assault.

This claim represents a strong form of the thought that survivors are the authorities on their own experiences. To motivate this, consider the following case:

(AS) Amy and Sharon are close friends. Sharon feels a strong sexual desire for Amy, but feels conflicted about this desire. In ordinary circumstances, Sharon would decline sex with Amy. Amy feels a strong sexual attraction to Sharon as well, but she feels no conflict about this — she would readily have sex with Amy given the opportunity. Both are open about their feelings to one another: Amy knows Sharon doesn't want to have sex, but dismisses this to herself as prudishness that Sharon will eventually see as such; Sharon knows that Amy wants to have sex with her, and would do so if Sharon were willing. Amy hatches a plan to have Sharon come over, drink alcohol, and offer to share her bed, hoping things will become sexual. Amy carries out her plan, and, in bed, asks "hey, do you want to hook up?" Sharon responds, "OK," and the two have sex while both are drunk.

The next morning, Sharon wakes up. She feels terrible, but isn't sure whether this is just because of the hangover. She leaves before Amy wakes. Some time later, Amy texts her with a message: "Not even a kiss goodbye?"

Now I take it that, thus far, what's been described is something like a borderline case of sexual assault. On the one hand, Amy is manipulating the situation, encouraging mutual intoxication, and aiming for sex to happen despite Sharon's sober

decisions; she is, in particular, hoping to *undermine* Sharon's usual resistance. On the other hand, far from forcing Sharon, Amy is not even *determined* to have sex — she knows beforehand that, even after alcohol, Sharon may still decline; she would not consider forcing Sharon into sex, or having sex with Sharon while she was unconscious, or having sex with Sharon if Sharon seemed genuinely unable to understand what was happening.

Consider now two possible extensions:

( $AS^+$ ) Sharon, getting home, takes a long shower; in the shower she listens to a *This American Life* about friendships blossoming into romance. Her hangover clears, and she begins to let herself think about the previous night. She sees Amy's text, laughs, and asks Amy out. The two end up dating happily. Some years later, Amy reflects on that night, and, feeling uncomfortable, asks Sharon whether she crossed a line. Sharon says no; she takes what happened that night not to have been a sexual assault.

The second extension is less happy:

( $AS^-$ ) Sharon starts to walk to the shower, but notices the “unread message” notification on her phone. She decides she needs a cup of coffee, only to realize she's all out. Headache and nausea raging, she reads Amy's text, and takes it as mocking. The night comes rushing back. She wonders to herself, “is what Amy did OK?” She takes a shower, but isn't in the mood to listen to a podcast. Afterward, she sits down and, thinking through the night, takes herself to have been manipulated into sex. Going online, she finds stories by other women that reflect her own experience. She comes to the conclusion that Amy assaulted her.

I claim it is plausible that Sharon is correct in both cases. In other words, she is correct in her assessment in ( $AS^+$ ) that Amy did not assault her, and she is *also* correct in her assessment in ( $AS^-$ ) that Amy did assault her. But we can stipulate that everything is the same about the cases up until Sharon gets home.<sup>19</sup> Then

---

<sup>19</sup>At least up to “small miracle” differences. Cf. (Lewis, 1979).

the only plausible grounding for the different assessments is the course of Sharon's reflection. Thus, this pair of cases is a witness to (*PA*).

To be clear, I don't take this sort of case to be commonplace. In fact, I think that most pairs of cases that might be described in these terms are not examples of this situation: most of the time, if we imagine a case in which Sharon later takes herself to have been assaulted, and a case in which she doesn't, we're actually imagining that there's something different, even if it's unobservable, about *Sharon at the time of the encounter* between the two cases. Cases in which the very same encounter could later turn out to have been, and could later turn out to not have been, assault may be exceedingly rare — even nonexistent. Despite this, I do think the sort of case described above — in which there is *no* difference at the time of the encounter — is *possible*; this is all that's needed for the claim.

Now the attitudinal model will have a hard time making sense of this result. This is true because among the things that do not differ at the time of the encounter are Amy and Sharon's mental states: their desires and wishes are the same in (*AS*<sup>-</sup>) and (*AS*<sup>+</sup>). Thus, the attitudinal model will give the same verdict as to whether Sharon consented in (*AS*<sup>-</sup>) and in (*AS*<sup>+</sup>). So if whether Amy assaulted Sharon turns on the question of whether Sharon consented to sex, and it's true that there *is* a difference in whether an assault occurred between the two, the attitudinal model cannot be correct.

The performative model, in contrast, has more room. There is, I believe (following MacFarlane (2003, 2005)), independent reason to think that the truth value of an utterance can depend upon not just the context in which it is uttered, but also upon the context in which it is assessed.<sup>20</sup> This dual context-sensitivity makes it possible to say, in MacFarlane's example, that my utterance today that "there will be a sea battle tomorrow" is of indeterminate truth as assessed today, but, determinately true (if there is a sea battle tomorrow) or false (if there is not) as assessed tomorrow. The same tool can be used to make sense of taste disagreement. If you say "chocolate is

---

<sup>20</sup>I am under no illusion that this is settled, or that MacFarlane's view is universal. I in fact believe that the same case can be made under alternative approaches to future contingents; sadly, a full articulation of this argument will not fit here.

disgusting,” and I say “no, it’s delicious!” it would be a mistake to understand us as asserting two different things (you saying that chocolate is disgusting *to you*, me saying that it’s delicious *to me*), for this would mean we’re not disagreeing; clearly we are. Yet it seems equally a mistake to say that one of us is correct. Assessment-sensitivity gives a solution: my utterance is true as assessed from my judgment-frame and false as assessed from yours, and vice-versa.

We can use exactly this tool to make sense of ( $AS^\pm$ ), if the question of whether Sharon consented is a question about what Sharon *expressed*. Was her “OK” a performance of a felicitous consent speech act?

To recall the beginning of this chapter, the attitudinal account’s intuitive appeal comes from the alignment between will and bodily experience it secures. As discussed above, a plausible version of the performative account will take the expressor’s agential capacity into account; an “OK” by someone who is incapacitated is infelicitous; it fails to do the work of consenting. The reason for this is, it seems to me, fundamentally connected to *autonomy*. Part of what it takes for a consent expression to be felicitous is for it to be uttered autonomously, in accordance with the utterer’s will. In cases where a consent expression is uttered in an infelicitous way — because the utterer is acting, or because the utterer is coerced, or because the utterer “isn’t in her right mind,” or because the utterer is not the person for whom the expression purports to grant consent — it is this underlying autonomy condition that fails.

Putting this together with assessment sensitivity, we can see why the performative account is compatible with the radical feminist claim. Sharon, according to her self-conception as constituted in the world of ( $AS^+$ ), takes herself to have autonomously uttered “OK.” She recognizes the utterer as the same person as herself. As assessed in ( $AS^+$ ), her utterance was felicitous; as such, as assessed in ( $AS^+$ ), she can truly say she was not assaulted. In contrast, Sharon, according to her self-conception as constituted in the world of ( $AS^-$ ), does not take herself to have autonomously uttered “OK” . She does *not* identify with the person who uttered this phrase.<sup>21</sup> As assessed

---

<sup>21</sup>Although there is not space to discuss this in detail, I believe this “identification” is more than rhetorical, but is both normatively and metaphysically significant, for many of the reasons discussed in the first half of Mark Johnston’s *Surviving Death* (2011).

in  $(AS^-)$ , her utterance was infelicitous; so, as assessed in  $(AS^-)$ , she can truly say that she was assaulted.<sup>22</sup>

### 3.7 Conclusion

In the bulk of this essay, I made an ameliorative case for the performative model in legal contexts: we should adopt the performative model because it fits our intuitions as well or better than alternatives, and because there are strong social justice reasons to do so.

In the final section, I sketched how the performative model can be put to even more radical work, by making sense of what might, at first glance, be an absurd claim: that a survivor's understanding of their own experience can, sometimes, be the determining factor in whether that experience was one of sexual assault — even if that understanding coalesces *after* the experience.

In this last, I have not considered all possible alternative approaches. In particular, I have taken for granted the assumption that whether someone is assaulted hinges upon whether they consent. There is room to resist this. Stephen Schulhofer (1992), for instance, rejects a consent-based approach to assault altogether, instead arguing for foregrounding autonomy directly in understanding sexual assault. This may well be an equally attractive ameliorative option — but delving into it is a task for future work. For now, I'm content with the claim that, given that consent is to play a central role in our reasoning about assault, the performative approach is most promising, together with the claim, embraced in the first chapter, that however we limn their relationship, consent cannot be understood without also considering autonomy.

---

<sup>22</sup>Note that nothing here entails that her evaluation is the only determiner of whether she consented. There may be other felicity conditions, and there are certainly other truth conditions — she cannot truly say she was assaulted, for instance, if there was no encounter at all.



# Bibliography

- Anderson, E. (1995). Knowledge, human interests, and objectivity in feminist epistemology. *Philosophical Topics*, 23(2):27–58.
- Appiah, A. (1985). The uncompleted argument: Du Bois and the illusion of race. *Critical Inquiry*, 12(1):21–37.
- Austin, J. L. (1975). *How to Do Things with Words*. Oxford, Oxford, 2nd edition.
- Bachman, R. and Paternoster, R. (1993). A contemporary look at the effects of rape law reform: How far have we really come? *The Journal of Criminal Law and Criminology (1973-)*, 84(3):554.
- Bakunin, M. (2005). Statism and anarchy. In Guérin, D., editor, *No Gods, No Masters*, pages 194–197. AK Press, Oakland.
- Barnes, E. (2015). Realism and social structure. Given at the Feminist Ontology Workshop.
- Barnes, E. (2016). *The Minority Body: A Theory of Disability*. Oxford University Press, Oxford, United Kingdom.
- Begay v. United States* (2008). 554 U.S. 137.
- Billock, J. (2018). When rape survivors have rape fantasies. Online at [https://www.vice.com/en\\_au/article/3k5gey/when-rape-survivors-have-rape-antasies](https://www.vice.com/en_au/article/3k5gey/when-rape-survivors-have-rape-antasies).
- Blackstone, W. (2016). *Commentaries on the Laws of England*. Oxford University Press, the oxford edition of blackstone edition.
- Brennan, S. (1999). Recent work in feminist ethics. *Ethics*, 109(4):858–893.
- Brett, N. (1998). Sexual offenses and consent. *Canadian Journal of Law & Jurisprudence*, 11(1):69–88.
- Bromberger, S. (1966). Why-questions. In Colodny, R., editor, *Mind and Cosmos*, pages 86–111. University of Pittsburgh Press, Pittsburgh.
- Burgess, A. and Plunkett, D. (2013). Conceptual ethics i. *Philosophy Compass*, 8(12):1091–1101.

- Burwell v. Hobby Lobby* (2014). 573 U.S. 682.
- Byrne, A. (2020). Are women adult human females? *Philosophical Studies*.
- Chambers v. United States* (2009). 555 U.S. 122.
- CrimethInc. Ex-Workers' Collective (2012). Breaking with consensus reality: From the politics of consent to the seduction of revolution. Online zine. Available at <http://crimethinc.com/texts/recentfeatures/breakwith.php>.
- Debord, G. (1995). *The Society of the Spectacle*. Zone Books, New York.
- DeRose, K. (2004). Single scoreboard semantics. *Philosophical Studies*, 119(1):1–21.
- Díaz-León, E. (2018). On haslanger's meta-metaphysics: Social structures and metaphysical deflationism. *Disputatio*, 10(50):201–216.
- Dredd Scott v. Sandford* (1856). 60 U.S. 398.
- Dretske, F. (1972). Contrastive statements. *The Philosophical Review*, 81(4):411–437.
- Employment Division v. Smith* (1990). 494 U.S. 872.
- Fara, D. G. (2012). Specifying desires. *Nous*.
- Felker v. Turpin* (1997). 518 U.S. 651.
- Fowles, S. M. (2008). The fantasy of acceptable “non-consent”: Why the female sexual submissive scares us (and why she shouldn't). In Friedman, J. and Valenti, J., editors, *Yes Means Yes: Visions of Female Sexual Power and a World Without Rape*, chapter 9, pages 117–175. Seal Press, Berkeley.
- Freeman, J. (1989). The feminist debate over prostitution reform: Prostitutes' rights groups, radical feminists, and the (im)possibility of consent. *Berkeley Women's Law Journal*, 5:75.
- Friedman, M. (2003). *Autonomy, Gender, Politics*. Studies in Feminist Philosophy. Oxford University Press, Oxford.
- Gallie, W. B. (1956). Essentially contested concepts. *Proceedings of the Aristotelian Society*, 56(1):167–198.
- Garfinkel, A. (1981). *Forms of Explanation*. Yale University Press, New Haven.
- Garner, B. A., editor (2019). *Black's Law Dictionary*. American Law Institute, 11th edition.
- Glanzberg, M. (2003). Focus: A case study on the semantics/pragmatics boundary. In Szabo, Z., editor, *Semantics vs. Pragmatics*, pages 1–43. Oxford.
- Gonzales v. O Centro Espírita Beneficente União do Vegetal* (2006). 546 U.S. 418.

- Graeber, D. (2007). On the phenomenology of giant puppets: Broken windows, imaginary jars of urine, and the cosmological role of the police in American culture. In *Possibilities: Essays on Hierarchy, Rebellion, and Desire*, chapter 12, pages 375–418. AK Press, Oakland.
- Green, M. D. and Powers, W. C. (2010). *Restatement of the Law (3d) of Torts—Liability for Physical and Emotional Harm*, volume 1. American Law Institute.
- Grice, P. (1957). Meaning. *The Philosophical Review*, 66(3):377–388.
- Hardy, J. W. and Easton, D. (2017). *The Ethical Slut*. Random House.
- Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs*, 34(1):31–55.
- Haslanger, S. (2006). Philosophical analysis and social kinds. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 80:89–143.
- Haslanger, S. (2007). “But Mom, crop-tops *are* cute!” social knowledge, social structure and ideology critique. *Philosophical Issues*, 17(1):70–91.
- Haslanger, S. (2012a). *Resisting Reality*. Oxford University Press.
- Haslanger, S. (2012b). Social explanation and social structure. Handout from the 2012 Carus Lectures.
- Haslanger, S. (2013). Race, intersectionality, and method: A reply to critics. *Philosophical Studies*, 171(1):109–119.
- Holton, R. (2003). How is strength of will possible? In Stroud, S. and Tappolet, C., editors, *Weakness of Will and Practical Irrationality*, pages 39–67. Oxford, Oxford.
- Hume, D. (1994). Of the original contract. In Haakonssen, K., editor, *Political Essays*, Cambridge Texts in the History of Political Thought, chapter 23, pages 286–201. Cambridge, Cambridge.
- Hurd, H. (1996). The moral magic of consent. *Legal Theory*, 2:121–146.
- Jackson, F. (1985). On the semantics and logic of obligation. *Mind*, 94(374):177–195.
- Johnson v. United States* (2015). 576 U.S. 591.
- Johnston, M. (2011). *Surviving Death*. Princeton University Press.
- Katz, J. (1977). Informed consent—A fairy tale—Law’s vision. *U. Pitt. L. Rev.*, 39:137.
- Kazan, P. (1998). Sexual assault and the problem of consent. In French, S., Teays, W., and Purdy, L., editors, *Violence against Women: Philosophical Perspectives*, pages 27–42. Cornell University Press.

- Kripke, S. (1980). *Naming and Necessity*. Harvard University Press, Cambridge, Mass.
- Lange, M. (2000). *Natural Laws in Scientific Practice*. Oxford University Press, Oxford New York.
- Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, 13(4):455.
- Longino, H. E. (1990). *Science As Social Knowledge*. Princeton University Press.
- MacFarlane, J. (2003). Future contingents and relative truth. *The Philosophical Quarterly*, 53(212):321–336.
- MacFarlane, J. (2005). Making sense of relative truth. *Proceedings of the Aristotelian Society, New Series*, 105:305–323.
- McLaughlin, P. (2007). *Anarchism and Authority: A Philosophical Introduction to Classical Anarchism*. Ashgate, Burlington, VT.
- O'Neill, O. (1985). Between consenting adults. *Philosophy & Public Affairs*, 14(3):252–277.
- Pateman, C. (1988). *The Sexual Contract*. Blackwell, Oxford.
- Plunkett, D. and Sundell, T. (2013). Disagreement and the semantics of normative and evaluative terms. *Philosophers Imprint*, 13.
- Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy*, 70(19):699.
- Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*, 7:131–193.
- Quine, W. V. O. (2013). *Word and Object*. MIT Press Ltd.
- Rawls, J. (1996). *Political Liberalism*. Columbia, New York, 2nd edition.
- Raz, J. (1981). Authority and consent. *Virginia Law Review*, pages 103–131.
- Richard, M. (1995). Defective contexts, accommodation, and normalization. *Canadian Journal of Philosophy*, 25(4):551–570.
- Richard, M. (2019). *Meanings As Species*. Oxford University Press, Oxford, United Kingdom.
- Schaffer, J. (2004). From contextualism to contrastivism. *Philosophical Studies*, 119(1):73–103.
- Schaffer, J. (2005). Contrastive causation. *The Philosophical Review*, 114(3):327–358.
- Schulhofer, S. J. (1992). Taking sexual autonomy seriously: Rape law and beyond. *Law and Philosophy*, 11(1/2):35.

- Shiffrin, S. V. (2008). Promising, intimate relationships, and conventionalism. *Philosophical Review*, 117(4):481–524.
- Snedegar, J. (2012). Reason claims and contrastivism about reasons. *Philosophical Studies*.
- Snedegar, J. (2013). Contrastive semantics for deontic modals. In Blaauw, M., editor, *Contrastivism in Philosophy*, pages 116–133. Routledge, New York.
- Sober, E. (1986). Explanatory presupposition. *Australasian Journal of Philosophy*, 64(2):143–149.
- Stalnaker, R. C. (2004). Assertion revisited: On the interpretation of two-dimensional modal semantics. *Philosophical Studies*, 118(1/2):299–322.
- Stokeling v. United States* (2019). 586 U.S. \_\_.
- Taylor v. United States* (1990). 495 U.S. 575.
- Tchen, C. M. (1983). Rape reform and a statutory consent defense. *Journal of Criminal Law and Criminology*, 74(4).
- Thomson, J. (1990). *The Realm of Rights*. Harvard University Press, Cambridge, Mass.
- Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458.
- United States v. Booker* (2005). 543 U.S. 220.
- United States v. Harrison* (CA10, 2011). 639 F.3d 1273.
- van Fraassen, B. (1980). *The Scientific Image*. Oxford University Press, Oxford.
- Warren, C. (1895). *Volenti Non Fit Injuria* in actions of negligence. *Harvard Law Review*, 8.
- West, R. (2007). The harms of consensual sex. In Soble, A. and Power, N., editors, *Philosophy of Sex: Contemporary Readings*, pages 317–322. Rowman & Littlefield.
- Wiestler, B., Capper, D., Holland-Letz, T., Korshunov, A., von Deimling, A., Pfister, S. M., Platten, M., Weller, M., and Wick, W. (2013). ATRX loss refines the classification of anaplastic gliomas and identifies a subgroup of IDH mutant astrocytic tumors with better prognosis. *Acta Neuropathologica*, 126(3):443–451.
- Williamson, T. (2007). *The Philosophy of Philosophy*. John Wiley and Sons Ltd.
- Wolff, R. P. (1998). *In Defense of Anarchism*. University of California Press, Berkeley.
- Yablo, S. (2008). Coulda, woulda, shoulda. In *Thoughts*, pages 103–150. Oxford University Press.

Young, I. M. (1990). *Justice and the Politics of Difference*. Princeton University Press, Princeton.