Forthcoming in the *Journal of Moral Philosophy*

Feasibility and Normative Penetration[1]

Matthew Lindauer (Brooklyn College and CUNY Graduate Center)
Nicholas Southwood (Australian National University)

An important theme in recent experimental philosophy, associated especially with the pioneering work of Joshua Knobe, is that certain judgements are subject to a kind of *normative penetration* whereby, in spite of a not-obviously-normative subject matter, they turn out to be sensitive to, and co-vary with, our normative attitudes in interesting and surprising ways (see Knobe 2010). Examples where there is empirical evidence of such normative penetration include our judgements about *intentional action* (Knobe 2003a; 2003b; 2004), *causation* (Fraser and Knobe 2008), *freedom* (Phillips and Knobe 2009), *knowledge* (Beebe and Buckwalter 2010), *happiness* (Phillips, Nyholm, and Liao 2014), and *doing and allowing* (Cushman, Knobe, and Sinnott-Armstrong 2008; Barry, Lindauer, and Øverland 2014).

[1] For invaluable comments on, discussion about, and suggestions for how to improve previous versions of this article we would like to thank Mark Alfano, Erik Angner, David Braddon-Mitchell, Krister Bykvist, Mark Budolfson, Ryan Cox, Garrett Cullity, Keith Dowding, Ant Eagle, Brian Earp, Lisa Ellis, Eva Erman, Steve Finlay, Bob Goodin, Toby Handfield, Brian Hedden, Karen Jones, Renée Jorgensen, Josh Knobe, Ole Koksvik, Holly Lawford-Smith, Kristie Miller, Geoff Sayre-McCord, Ryan Muldoon, Jonathan Phillips, Sam Shpall, Nick Smith, Kai Spiekermann, Zosia Stemplowska, John Thrasher, audiences at the Experimental Philosophy and Political Theory Workshop at CUNY, the PPE Conference in New Orleans, the Australasian Workshop in Moral Philosophy at Kioloa, the Methodology of Political Theory Workshop at LSE, the Experimental Philosophy and Normativity Workshop at ANU, and in seminars at Adelaide University, the University of Sydney, Utrecht University, and Stockholm University, and two astute and generous referees. We owe a particular debt of gratitude to Josh Knobe for his encouragement and detailed written comments. Southwood's research for the article was supported by an Australian Research Council Future Fellowship (FT160100409).

Our question is: Do our judgements about *feasibility* (also) exhibit normative penetration; if so, how, why, and in what form; and what if anything of philosophical significance follows from it? The last few years have witnessed an explosion of interest in the idea of feasibility, especially in political philosophy (see Southwood 2018). Questions about feasibility are, of course, pervasive in social and political life, where they are often treated as having an important bearing on questions about what ought to be done, what justice requires, and so on (see Valentini 2012). But political philosophers remain deeply divided about how to understand the idea of feasibility and its normative significance. One important issue in this context concerns whether judgements about feasibility can be entirely divorced from normative considerations. According to some theorists, judgements about feasibility are themselves partly normative judgements about what is achievable without normatively unacceptable or inappropriate effects (Räikkä 1998; Buchanan 2004; Miller 2013). By contrast, other theorists deny this and hold that judgements about feasibility concern considerations that are prior to, and independent of, normative considerations (Gilabert and Lawford-Smith 2012; Wiens 2015; Stemplowska 2016; Southwood 2016, 2022). Yet, in spite of this burgeoning interest in feasibility, there has been no attempt to study our judgements about feasibility *empirically* hitherto. As a result, we simply do have not have any empirical evidence to bring to bear on such questions.

Our aim is to remedy this significant omission. We shall argue for three main claims. First, there is compelling empirical evidence, deriving from a series of experimental studies we ran, that our judgements about feasibility do indeed exhibit

normative penetration in at least some cases. Second, the best interpretation of the empirical data from these studies involves attributing to people a certain kind of extensional error, whereby they tend to conflate the question at hand with another question that is more salient given their normative attitudes. Third, this has significant, though not straightforward, implications for our understanding of both normative penetration and feasibility, respectively.

1. The normative penetration thesis

Do our feasibility ascriptions exhibit normative penetration – in at least some cases? To help answer this question we ran a study, henceforth Study 1, involving a modified version of the Aristotelian "ship's captain" case employed by Phillips and Knobe (2009) in which a ship is a caught in a storm and its captain realizes that the vessel is too heavy and will flood and capsize unless he makes it lighter. Study 1 deployed a between-participant design (N = 104).[2] Participants were randomly assigned to one of two conditions. In the first condition, participants were told that the only way to lighten the vessel is to have the ship's *cargo* thrown overboard (henceforth the "Cargo condition"). In the second condition, participants were told that the only way to lighten the vessel is to have the ship's *passengers* thrown overboard (hereafter the "Passengers condition"). In both conditions, participants were asked to rate their level of agreement with the

[2] The participants in all of the studies described in the paper were MTurk users. We limited our participant pool to MTurk users from the United States with a 98% or better approval rating for their past work.

statement “It is feasible for the captain to save the ship” on a 7-point Likert scale. The text of the Cargo and Passengers conditions read as follows:

> While sailing on the sea, a ship encounters a violent storm. As the waves begin to grow larger, the ship’s captain realizes that the vessel is too heavy and will flood and capsize unless he makes it lighter, and that the only way to lighten the vessel is to have the ship’s [cargo/passengers] thrown overboard.
>
> Please rate your level of agreement with the following statement on a 7-point scale (1 = Not at All, 7 = Fully Agree):
>
> *It is feasible for the captain to save the ship*.

Responses to the two cases diverged greatly. Participants were very likely to *agree* (very strongly) with the statement “It is feasible for the captain to save the ship” in the Cargo condition (M = 6.24, SD = .99). By contrast, participants were quite likely to *disagree* (though not strongly) with the statement “It is feasible for the captain to save the ship” in the Passengers condition (M = 3.3, SD = 2.18). The difference in mean agreeing ratings between the two conditions was statistically significant.[3]

On the face of it, these results provide compelling evidence for the thesis that feasibility judgements are subject to normative penetration. The obvious difference between the Cargo and Passenger conditions is the moral valence of the acts that are

[3] Welch’s $t(73.2) = 8.89$, $p < .001$, $d = 1.74$.

required to save the ship. It is reasonable to suppose that the difference between participants' feasibility ascriptions about the cases is due to this moral difference. There is no other obvious difference between the Cargo condition and Passengers condition that would explain the significant difference in participants' mean ratings.

Is there some *unobvious* difference that could instead be driving the effect? Maybe. One possibility is that "the ship" is being interpreted by participants in a broad way to encompass not just the vessel but also the ship's passengers. Thus, whereas saving the vessel (and presumably the ship's passengers) at the cost of the ship's cargo does constitute saving the ship in the relevant sense, saving the vessel at the cost of the ship's passengers might *not*.[4] A second possibility is that participants' feasibility ascriptions are tracking their judgements about the *captain's* normative attitudes as opposed to their own. That is to say, they might be assuming that the captain will be unable to bring himself to do what is necessary in order to save the ship because having the passengers thrown overboard would violate his core moral convictions. A third possibility is that participants' feasibility ascriptions are simply tracking their judgements about the relative *difficulty* of having passengers as opposed to cargo thrown overboard. After all, unlike cargo, passengers are presumably quite likely to resist being thrown overboard.

To determine whether these other factors might explain the difference in feasibility ascriptions about these cases, we ran a modified version of Study 1, henceforth Study 1a (N = 134), designed to control for each of the three factors. First,

[4] We are grateful to […] and […] both of whom independently raised this possibility.

we changed “save the ship” to “save the vessel.” Second, we described the captain as “widely known to be a ruthless man” to make it clear that he would have few compunctions about having the passengers thrown overboard. Third, we specified that the passengers were “a large group of recently orphaned infants” – to make it clear that there would be no chance of their resisting. For parity in terms of the specificity of the descriptions of the entities being thrown overboard between the cases, we also specified that the ship’s cargo was “a cargo of iron ore” in the revised Cargo condition.

The same effect that we observed in Study 1 was found again here. Participants were very likely to agree (very strongly) with the statement “It is feasible for the captain to save the ship” in the Cargo condition (M = 6.01, SD = 1.39) and quite likely to disagree (though not strongly) with the statement “It is feasible for the captain to save the ship” in the Orphans condition (M = 3.06, SD = 2.18).[5] This strongly suggests that these other factors cannot explain away the apparent normative penetration on feasibility judgements observed in Study 1.

We also ran a mediation study[6] to examine whether there is further evidence that moral judgements are driving the difference between responses to the Cargo and Passenger cases. In this study, participants were randomly assigned to either the Cargo or Passengers case, and in addition to rating the feasibility statement rated the statement “It would be morally wrong for the captain to save the ship” on a 7-point scale. The order of these questions was randomized. As we predicted, moral wrongness

[5] The difference here was also statistically significant. Welch’s $t(107.5) = 9.3$, $p < .001$, $d = 1.6$.
[6] Participants in this mediation study were 125 Prolific users from the United States.

judgements fully mediated the relationship between the conditions and feasibility judgements, accounting for 83% of the effect of condition on feasibility ratings.[7]

Thus, we conclude that the results of Study 1 do indeed provide empirical support for the normative penetration thesis. To be sure, it is only one study, with a follow-up study to control for possible confounds and a mediation study.[8] Moreover, the ship's captain case is a pretty extreme one; and we might reasonably wonder whether normative penetration will extend to less extreme cases where our feasibility ascriptions involve actions that are taken by participants to be wrong without being morally heinous. Yet even if it doesn't extend to less extreme cases, this would not show that there isn't normative penetration in the case of our feasibility ascriptions, merely that it is restricted to extreme cases.

---

[7] The regression coefficient between condition and feasibility rating and the regression coefficient between moral wrongness rating and feasibility rating were significant. The indirect effect was (-3.64)*(-0.58) = 2.11. We used bootstrapping procedures to test the significance of this indirect effect. Unstandardized indirect effects were computed for each of 1,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized indirect effect was 2.11, and the 95% confidence interval ranged from 1.43 to 2.92. Thus, the indirect effect was statistically significant ($p < .001$). We are very grateful to an anonymous reviewer for suggesting that we run this mediation study.

[8] Additionally, while we do not wish to rest our case on it, Study 1 participants' explanations of their responses in the Passengers condition also strongly suggested that their judgements regarding this condition were driven by moral considerations. The vast majority of participants in this condition whose ratings were below the midpoint, indicating disagreement with the feasibility statement, explained their response in terms of the immorality or unethical nature of throwing the passengers overboard. Some excerpts from these explanations include "It would not be feasible because the requirement to save the ship from sinking would be one that is completely immoral and under no circumstance should it be done," "The captain could throw people overboard but it would be highly immoral even to save the ship," and "I think that it is not feasible for the captain to save the ship since the only way to do so would be to kill all of the passengers, which is a very evil and immoral act. One should never resort to the murder of human beings to save inanimate objects, such as a ship; since the value of human beings is infinitely greater than the value of any inanimate object." These and other responses suggest that most participants who judged that saving the ship by throwing the passengers overboard would be infeasible did so on the basis of moral considerations. We are grateful to an anonymous reviewer for suggesting that we mention participants' responses in noting that the main difference between the Cargo and Passengers conditions is their moral content.

2. How not to explain the normative penetration thesis

We have presented empirical evidence of normative penetration in the case of our judgements about feasibility. But how should we *explain* it?

A. The constitutive interpretation

One possibility is that our feasibility ascriptions are co-varying with our normative attitudes because they *just are* normative attitudes (albeit not-obviously-normative attitudes). Call this the *constitutive interpretation*. The constitutive interpretation follows straightforwardly from certain prominent normative accounts of feasibility, which hold that our feasibility judgements are judgements about what is achievable without normatively unacceptable or inappropriate effects (Räikkä 1998; Buchanan 2004; Miller 2013).[9] For instance, David Miller claims that feasibility has an "inescapable normative element" (1993, 32). On his account, the limits of feasibility are set "not just by physical and sociological laws, but by implicit assumptions about what, for us, would count as a tolerable or intolerable outcome" (*ibid.*). He illustrates this point by discussing Rawls' assumption in *A Theory of Justice* (1971) of the existence of the family as a social institution. It is possible that if we replaced the kinds of small family units that are typical in liberal democratic societies with a system of collective childrearing, justice might be promoted in various ways, such as through improvements to equality of

[9] We had previously referred to these views as "cost-based" accounts of feasibility, following some descriptions in the literature. A reviewer made the helpful point that what is at issue for these accounts isn't cost *per se*, and led us to see that certain natural confusions arise if we describe these accounts in this way. Hence, we instead refer to normative accounts of feasibility here. We are grateful to the reviewer for helping us to see why this terminology is to be preferred.

opportunity. But whatever benefits there might be to such an arrangement, imposing it as a society-wide policy would be "wholly unacceptable to us" (Miller 1993, 33). The idea is that while various reforms to the family as a social institution may be feasible, as we have seen with pushes for gender equality and the inclusion of same-sex marriages, promoting justice in a way that would forcibly get rid of the family would be infeasible because this would be a normatively unacceptable effect.

Normative accounts of feasibility, if correct, can easily explain the results of Study 1. Whether or not the captain's saving the ship will count as achievable without normatively unacceptable or inappropriate effects – and, hence, feasible by the lights of these accounts – will obviously depend on what is required to achieve it. If saving the ship requires throwing all the ship's *cargo* overboard (as in the Cargo condition), then it is plausible to suppose that saving the ship will count as achievable without anything normatively unacceptable or inappropriate occurring. Given the alternative (the loss of the ship and potentially the death of everyone onboard), it doesn't seem that there are any unacceptable or inappropriate effects of doing so by the lights of moral or other norms. By contrast, if saving the ship requires throwing all the ship's *passengers* overboard (as in the Passengers condition), then it is plausible to suppose that saving the ship will *not* count as feasible on such accounts. Here, while the alternative (the loss of the ship and potentially the death of everyone onboard) is extremely significant, having the ship's passengers thrown overboard is plausibly even worse and morally unacceptable.

Unfortunately, however, the constitutive interpretation is not well supported by certain additional empirical evidence. The natural way to test it is to consider whether the normative penetration we observed in Study 1 is *nullifiable*: that is, whether participants' feasibility ascriptions and the normative attitudes to which they appear to be sensitive can be, as it were, *decoupled*.[10] If the constitutive interpretation is correct – that is, if our feasibility ascriptions are co-varying with our normative attitudes because they just are normative attitudes (say, judgements about what is achievable without normatively unacceptable or inappropriate effects) – then we should obviously *not* expect to find decoupling taking place. If we do find decoupling, this provides us with compelling evidence that the constitutive interpretation is mistaken.

To determine whether the normative penetration we observed in Study 1 is nullifiable in this way we ran a second study, henceforth Study 2. Like Study 1, Study 2 employed a between-participant design (N = 170) with participants being randomly assigned to one of two conditions. Participants in both conditions were presented with the same vignette as the Passengers version of the vignette in Study 1 and were asked to rate their level of agreement with a statement on a 7-point Likert scale. However, the two conditions involved presenting subjects with different statements. In the first condition, participants were asked to rate their level of agreement with the *original statement*: "It is feasible for the captain to save the ship" (henceforth the "Original Passengers condition"). In the second condition, participants were asked to rate their

[10] We are grateful to […] for urging us to use a neutral such as "nullifiability," rather than the more familiar term, "cancellability," on the grounds that the latter is obviously associated with the sort of conversational implicature explanation that we consider next.

level of agreement with a new *nullifying statement*: "It is feasible for the captain to save the ship but, if he does this, he will be doing something very wrong" (henceforth the "Nullifying Passengers condition").

Once again, the results of Study 2 were striking. In the Original Passengers condition, participants were quite likely to *disagree* (though not strongly) with the original statement**,** "It is feasible for the captain to save the ship" (*M* = 3.32, *SD* = 2.19). As expected, the mean agreement rating given to the original statement here was similar to the mean agreement rating given to the statement in the Passengers condition in Study 1 (3.3). However, in the Nullifying Passengers condition, participants were likely to *agree* (quite strongly) with the nullifying statement, "It is feasible for the captain to save the ship by having the passengers thrown overboard but, if he does this, he will be doing something very wrong" (*M* = 5.6, *SD* = 1.96).[11] On the face of it, Study 2 provides strong evidence that the normative penetration that our feasibility ascriptions appear to exhibit is nullifiable. Here, if anything, the fact that it is an extreme case makes the evidence all the stronger. Thus, the constitutive interpretation is not well supported by the available empirical evidence.

B. The conversational implicature interpretation

The lesson from our foray into the constitutive interpretation is that we need an interpretation of the normative penetration thesis that can make sense of nullifiability.

[11] The difference between the mean agreement ratings for the nullifying statement and the original statement was statistically significant. Welch's $t(167.75) = 7.16$, $p < .001$, $d = 1.1$.

An obvious thought is that our feasibility ascriptions are sensitive to our normative attitudes when they are because of certain normative *conversational implicatures* associated with feasibility ascriptions in particular conversational contexts. Rather than normativity being any part of the semantic content of our feasibility ascriptions, our normative attitudes are playing a purely *pragmatic* role in regulating our feasibility ascriptions. We will sometimes be reluctant to assent to certain feasibility statements, not because we think they are false (we may well think they are true), but because of implicatures associated with assenting to the statements within a certain conversational context – the content of which would be at odds with our normative attitudes. In particular, assenting to statements to the effect that it is feasible for an agent to perform an act will often implicate that we *condone* (or don't sufficiently strongly *condemn*) the agent's performing the act, or that we do not regard her as blameworthy insofar as she performs the act. Call this the *conversational implicature interpretation*.

Is the conversational implicature interpretation well supported by the available evidence? It offers a persuasive explanation of the results of Study 1. The conversational context in both versions of Study 1 is such that agreeing to the statement would conversationally implicate that one judges that the captain's saving the ship has a certain normative status: that it is morally permissible for him to save the ship, or not obviously morally impermissible. In the Cargo condition participants are perfectly willing to implicate this since they think it's true. By contrast, in the Passengers condition participants are unwilling to implicate this given that they judge that saving the ship would require the captain to do something morally heinous, namely to throw the

passengers overboard. The only way for them to distance themselves from the implicature is to *disagree* with the statement.

Moreover, unlike the constitutive interpretation, the conversational implicature interpretation appears to be tailor-made to explain the results of Study 2. Nullifiability – in particular, *cancellability* – is, of course, one of the key markers of conversational implicature (Grice 1975; Blome-Tillmann 2013). The conversational implicature interpretation therefore predicts that participants' unwillingness to ascribe feasibility to the captain's saving the ship when doing so requires having the passengers thrown overboard will evaporate if we provide participants with the opportunity to explicitly register (and perhaps to be seen to explicitly register) their strong moral disapprobation vis-à-vis the captain's saving the ship by having the passengers thrown overboard. This is exactly what we saw in the Nullifying Passengers condition in Study 2.

Nonetheless, here too we suggest that there is additional experimental data that is at least somewhat at odds with the conversational implicature interpretation. For example, there is evidence that nullifying is possible *without* providing subjects with the opportunity to register their moral disapprobation vis-à-vis the captain's saving the ship by having the passengers thrown overboard. For example, we ran a different nullifying study, henceforth Study 3 (N = 150), structurally identical to Study 2, except that the second condition involved presenting subjects with the following *non-normative* nullifying statement: "It is feasible for the captain to save the ship, but it is not feasible for him to save the ship without having the passengers thrown overboard."

Again, we found a significant difference in mean agreement between the two conditions: the new Non-Normative Nullifying Passengers condition (*M* = 4.64, *SD* = 2.16) and the Original Passengers condition (*M* = 2.82, *SD* = 1.97).[12] This is not what we would expect if the conversational implicature interpretation were the correct one. If agreeing to the original statement, "it is feasible for the captain to save the ship" is supposed to carry with it the unwelcome implicature that one is somehow condoning saving the ship by having the passengers thrown overboard, it is hard to see how this would not also be true of agreeing to the statement, "it is feasible for the captain to save the ship, but it is not feasible for him to save the ship without having the passengers thrown overboard." Why should the implicature be cancelled merely by merely more *specific* about what is and isn't feasible? Indeed, if anything, the implicature associated with the non-normative nullifying statement might be thought to be even worse since it might seem to convey an openness to the idea of saving the ship *by* having the passengers thrown overboard. This suggests that something else is going on. The normative and non-normative nullifying statements are not serving to cancel any default *implicature* associated with the original statement, "it is feasible for the captain to save the ship." Rather, they are playing some other kind of role.

C. The contextualist interpretation

This brings us to what we take to be the most promising existing interpretative strategy and rival to our own favoured view, namely one that deploys an account of "feasible"

[12] Welch's $t(147.8) = 5.37$, $p < .001$, $d = .88$.

along the lines of the account of "can" suggested by Angelika Kratzer (1977). A Kratzerian account of feasibility would hold that our feasibility ascriptions express judgements about what is possible in light of certain considerations – considerations that are salient within a conversational context. These considerations may (and indeed typically will) include certain normative considerations. So, participants' feasibility ascriptions co-vary with their normative attitudes when they do because a) feasibility ascriptions express judgements about what is possible in light of considerations that are salient within a conversational context; and b) certain normative considerations are typically among the considerations that are salient within an ordinary conversational context. In consequence, c) feasibility ascriptions are typically (though not inevitably) going to be sensitive to participants' normative attitudes. Call this the *contextualist interpretation*.

The contextualist interpretation has the virtue of being able to explain the results of all the experimental studies discussed hitherto. First, it can explain the significant difference in mean agreement between the Cargo condition and the Passengers condition in Study 1. Whereas possible worlds in which the captain saves the ship by having the ship's cargo thrown overboard are not ruled out by the normative judgements that are part of the conversational context within the Cargo condition, worlds in which the captain saves the ship by having the ship's passengers thrown overboard *are* ruled out by the normative judgements that are part of the conversational context within the Passengers condition.

Second, the contextualist interpretation can also explain the significant differences in mean agreement between a) the Original Passengers condition and the Nullifying Passengers condition in Study 2 and b) the Original Passengers condition and the Non-Normative Nullifying Passengers condition in Study 3. Whereas within the Original Passengers condition worlds in which the captain saves the ship by having the ship's passengers thrown overboard are ruled out in light of certain normative considerations, those worlds are *ruled back in* in the Nullifying Passengers condition and the Non-Normative Nullifying Passengers condition.

However, and while we don't want to overstate the point, we suggest that there is additional experimental data that at least presents a *challenge* for the contextualist interpretation. We shall focus on data concerning how participants tend to respond to certain feasibility ascriptions by *others* as opposed to their own feasibility ascriptions: in particular, their ascriptions of correctness and incorrectness.[13] The underlying thought is that if contextualism is the correct view of what speakers mean by making feasibility utterances, then we should expect speakers to have dispositions, not merely to make and refrain from making feasibility ascriptions in a way that accords with the predictions that contextualism makes, but also to respond in particular ways to what are, by the lights of contextualism, correct and incorrect feasibility ascriptions by others. At least,

[13] While there are some important differences, the experimental design in Study 4 control, Study 4, and Study 5 bears some resemblance to that used in recent work that purports to identify empirical support for contextualist theories of the semantics of both epistemic modals (Khoo 2005; Khoo and Phillips 2018) and moral claims (Khoo and Knobe 2018). Notice that these results are perfectly consistent with our own results, which only purport to identify evidence against a contextualist interpretation of normative penetration in the context of our feasibility ascriptions.

we would expect this when considerations that are salient by the lights of contextualism are rendered suitably *transparent*.

To check, we first ran the following control study involving ascriptions of *costliness* Study 4 control, *N* = 92):

> Anna and Belinda are both competitors in the New York Marathon. After the race a journalist approaches Anna and asks her: "Tell me about what it takes to run a marathon. Is it a very costly thing to do?" Anna's response is: "Yes, it sure is very costly." Later, the journalist approaches Belinda and asks her the same question. Belinda's response is: "Not especially costly, no. But it does require huge sacrifices in terms of time and energy."
>
> Please select from the following four options:
>
> (a) Anna's response is the correct response.
>
> (b) Belinda's response is the correct response.
>
> (c) Both responses are correct.
>
> (d) Neither response is correct.

We hypothesised that participants would tend to choose option (c) – or perhaps (d) – but not (a) or (b). That's because participants would tend to recognise that Anna's and Belinda's utterances about the costliness of running the New York marathon are clearly operating with different implicit comparison classes. Given this recognition, it would be odd for participants to insist upon one of them being correct and the other being incorrect. The results of the study strongly supported this hypothesis with 75% of

subjects favouring response (c) – significantly greater than the percentage who favoured the other options.[14]

How about our feasibility ascriptions? Here, too, if the contextualist interpretation is correct, then we should expect participants to exhibit the same kinds of patterns of responses to feasibility ascriptions (see also Cappelen and Lepore 2005). To check, we ran the following modified version of the ship's captain case, henceforth Study 4 (*N* = 101):

> Participants have been randomly selected to take an online survey. First, participants in the survey are presented with the following text:
>
> > *While sailing on the sea, a ship encounters a violent storm. As the waves begin to grow larger, the ship's captain realizes that the vessel is too heavy and will flood and capsize unless he makes it lighter, and that the only way to lighten the vessel is to have the ship's passengers thrown overboard.*
>
> Next, participants are asked to type in their response to the following question:
>
> > *Is it feasible for the captain to save the ship?*
>
> Participant A says: "No, it is not feasible for the captain to save the ship."

[14] Vs. 2.2% for option (a), $\chi^2(1) = 100.07$, $p < .001$, $\varphi = 1.19$; vs. 20.7% for option (b), $\chi^2(1) = 52.88$, $p < .001$, $\varphi = .78$; vs. 2.2% for option (d), $\chi^2(1) = 100.07$, $p < .001$, $\varphi = 1.19$.

Participant B says: "Yes, it is feasible for the captain to save the ship by having the passengers thrown overboard but, if he does this, he will be doing something very wrong."

Please select from the following four options:

(a) Participant A's response is the correct response.

(b) Participant B's response is the correct response.

(c) Both responses are correct.

(d) Neither response is correct.

Yet here we found a completely different pattern of results. 19.8% of participants favoured option (a), 45.5% favoured option (b), 24.8% favoured option (c), and 9.9% favoured option (d). On the face of it, this is evidence against the contextualist interpretation. Whereas contextualism predicts that participants will go with option (c), only 24.8% of participants did so. 65.3% of participants favoured options (a) or (b). Moreover, the fact that this support was strongly concentrated on option (b) in particular, which received 45.5% of the support, significantly greater support than option (c),[15] provides evidence for a rival explanation, as we shall see.

What might the contextualist say in response?[16] One response is that, at most, Study 4 is testing participants' *theory* about what speakers mean when they make feasibility utterances,[17] whereas contextualism about feasibility is merely a view about what speakers *mean* when they make feasibility utterances. For contextualism to be the

[15] $\chi^2(1) = 9.45$, $p = .002$, $\varphi = .36$.

[16] For interesting responses to a similar objection to contextualism about other modal terms, see, for example, Finlay 2014, pp. 236-44 and von Fintel and Gillies 2011.

[17] We are grateful to […] for reminding us to heed this distinction.

correct view it is not necessary that participants must be themselves adherents of contextualism. However, this response rests on a mistake. Study 4 is not testing whether participants are adherents of contextualism. Rather, it is testing whether participants accept certain *linguistic norms* that behave as contextualism predicts. For there to be a linguistic norm it is *not* enough merely to look at whether speakers' feasibility ascriptions tend to behave as contextualism predicts. We must also look at subjects' dispositions to make correctness judgements regarding others' feasibility ascriptions.

Another possible response is that contextualism, in fact, successfully predicts the results of study 4. That's because participants cannot help but share the conversational context of participant B and, therefore, to interpret the first part of the nullifying statement as expressing the idea that saving the ship is compossible with the way things are even if very serious wrongs are thereby committed. Given this interpretation, participant B's response is, in effect, the correct response.[18] The question is: what *entitles* the contextualist to the claim that the conversational context of participant B is somehow *privileged* such that participants cannot help but share it? As we have seen, other contextualist-friendly predicates, such as "costly," do not behave in the same way. The natural explanation for why feasibility is different is that participants have picked up on the fact that participant B's interpretation is the right one and participant A's interpretation is mistaken. It is difficult if not impossible for participants to ignore this fact and treat an interpretation that they recognise to be mistaken as equally valid. But, of course, this explanation is not available to the contextualist.

[18] The discussion of this response benefitted greatly from input from […] and […].

Is this a conclusive refutation of the contextualist interpretation? Certainly not. Nonetheless, we suggest that a proponent of the contextualist interpretations at least incurs a burden that is not easy to discharge without surrendering at least some of what was initially appealing about contextualism. It is worth looking to see whether there is some *other* interpretation of the normative penetration thesis that can explain the relevant experimental data in an appealing way. It is to that task that we now turn.

3. The conflation interpretation

Our alternative suggestion is that feasibility ascriptions are subject to normative penetration when and because we are disposed to make a certain kind of extensional error that results from conflating the question at hand with another question that is more salient to us given our normative attitudes. Let's call this the *conflation interpretation*.

The conflation interpretation consists of three key claims. The first claim is that normative penetration involves relevant individuals' being disposed to make an *extensional error*: in particular, the error of ascribing infeasibility or resisting ascribing feasibility to acts that, as a matter of fact, are perfectly feasible. The conflation interpretation, therefore, offers an *error theoretic* interpretation of the normative penetration thesis. As such, it is quite different from the other three interpretative theses we encountered in the previous section, which seek to provide *vindicating* explanations of normative penetration – albeit vindicating explanations of very different

kinds. By contrast, the conflation interpretation seeks to provide a kind of *undermining* explanation.

The second key claim is that we are disposed to make the extensional error (of ascribing infeasibility or resisting ascribing feasibility to acts that are in fact feasible) when we are because we are disposed to *conflate* what is at issue – namely whether it is feasible for an agent to do something – with the question of whether it is feasible for the agent to do that thing *in a certain way*. To say that we are *conflating* questions is to say that, without noticing it, we are implicitly settling one question (whether x is F) by settling another intensionally distinct question (whether x is F* or whether x* is F).[19] Given that the questions are intensionally distinct, settling the one by settling the other involves a kind of *intensional slippage* that makes participants vulnerable to extensional errors.

The third key claim is that we are disposed to conflate these questions when and because, given our normative attitudes, the question of whether it is feasible for an agent to do something *in a certain way* is more salient to us than the question of whether it is feasible for the agent to do that thing *simpliciter*. Other things being equal, questions about what is feasible are salient to us insofar as they stand to provide us with information that would be useful for settling the question of what the agent ought to do. Some feasibility questions provide us with information that is simply more useful

[19] We are not, of course, suggesting that our feasibility ascriptions (necessarily or typically) involve individuals *explicitly considering* any question such as the question of whether it is feasible for an agent to do something.

for settling the question of what agents ought to do than the information that is provided by other feasibility questions.

We suggest that the conflation interpretation is better supported by the available empirical evidence than its rivals. Like its rivals it can easily explain normative penetration. Unlike its rivals, it can readily accommodate the results of the additional experimental studies that made problems for the rivals and another study that was designed to test the conflation interpretation in particular.

Let's start with Study 1. Take the Passengers condition. First, the conflation interpretation holds that a non-trivial number of participants in the Passengers condition are making an extensional error in resisting ascribing feasibility to the captain's saving the ship. Second, they are making this extensional error because they are conflating the question of whether it is feasible for the captain to save the ship with the intensionally distinct question of whether it is feasible for the captain to save the ship *without having the ship's passengers thrown overboard*. Third, they are conflating these questions because, given the participants' normative judgements, the question of whether it is feasible for the captain to save the ship without throwing the passengers overboard is *more salient* than the question of whether it is feasible for the captain to save the ship simpliciter. The first question is obviously highly relevant to settling the question of what the captain ought to do since plausibly an affirmative answer to it would imply that the captain ought to save the ship (obviously without having the passengers thrown overboard). By contrast, the second question is much less helpful in settling the question of what the captain ought to do since, even if the answer to it is

yes, it does not follow that the captain ought to save the ship (if, say, it is not feasible for the captain to save the ship without throwing the passengers overboard).

Next, take the Cargo condition. First, the conflation interpretation holds that most participants in the Cargo condition are *not* making an extensional error since they are not incorrectly failing to ascribe feasibility to the captain saving the ship. Second, they are not making this extensional error because they are not conflating the question of whether it is feasible for the captain to save the ship simpliciter with the question of whether it is feasible for the captain to save the ship without throwing the cargo overboard. And, third, they are not making this conflation because, the question of whether it is feasible for the captain to save the ship without throwing the cargo overboard is *not* the most salient question given the participants' normative beliefs. That's because the participants do not judge that it would be (very) wrong for the captain to have the cargo thrown overboard and, hence, they do not judge that the captain ought to save the ship only if he saves the ship without having the cargo thrown overboard. On the contrary, the feasibility question that seems most salient would seem to be the question of whether it is feasible for the captain to save the ship *by* having the ship's cargo thrown overboard. Plausibly, an affirmative answer to this question would entail an affirmative answer to the question of whether the captain ought to throw the cargo overboard.

How about the results of Study 2? How does the conflation interpretation explain the significant difference in mean agreement in the Nullifying Passengers condition and the Original Passengers condition? We suggest that the nullifying statement serves to

prevent the conflation from happening by, in effect, specifying that we are interested in whether there is *any way at all* – including some way that involves doing something very wrong – such that it is feasible for the captain to save the ship. Where the original statement invites the conflation given the relatively low salience of whether it is feasible for the captain to save the ship without having the passengers thrown overboard, the nullifying statement rules it out and forces conscientious participants to examine the matter that is actually at issue: the matter of whether it is feasible for the captain to save the ship.

How about the results of study 3? How does the conflation interpretation explain the significant difference in mean agreement in the Non-Normative Nullifying Passengers condition and the Original Passengers condition? This is even more straightforward. Whereas the normative nullifying statement in Study 4 implicitly invites participants to distinguish the question of whether it is feasible for the captain to save the ship with the question of whether it is feasible for the captain to save the ship without having the passengers thrown overboard, the non-normative nullifying statement in Study 3 *explicitly* distinguishes these questions.

Finally, how does the conflation interpretation explain the results of Study 4? Again, very easily. We saw that participants were inclined to describe Participant A's response (that it is feasible for the captain to save the ship by having the passengers thrown overboard but, if he does this, he will be doing something very wrong) as the correct response. The conflation interpretation implies that they are right to do so. It

implies that Participant B's response involves an extensional error: the error of describing something that is perfectly feasible (though wrong) as infeasible.

So the conflation interpretation can easily accommodate the results of all the experiments deployed hitherto. How else might we test it? The natural way to do so is to provide a modification of Study 4 based on the non-normative nullifying statement in Study 3. The idea is that if we are right about the nature of the error that participants are making in Study 1, then we should expect participants to agree that the non-normative nullifying statement is the uniquely correct response. To check, we ran a study of just this sort, henceforth Study 5 (N = 102). Study 5 was just like Study 4 except that Participant B says, "Yes, it is feasible for the captain to save the ship, but it is not feasible for him to save the ship without having the passengers thrown overboard." The conflation interpretation predicts that participants will tend to favour option (b). And that is exactly what we found. 18.6% of participants favoured option (a), 51% favoured option (b), 19.6% favoured option (c), and 10.8% favoured option (d). Of particular relevance, again, is the comparison between option (b) and option (c), where (b) received significantly greater support.[20] On the face of it, this is strong evidence for the conflation interpretation. Of course, it may yet turn out to be unable to explain *other* experimental data. However, we suggest that, given the experimental data such as it is, the conflation interpretation represents the best available explanation of the normative penetration we observe in the case of our feasibility ascriptions.

[20] $\chi^2(1) = 21.9$, $p < .001$, $\varphi = .55$.

4. Implications

We have argued that there is compelling evidence of normative penetration in the case of our feasibility judgements and that the best explanation of it is provided by the conflation view. These conclusions are significant in their own right. We shall now argue that they also have significant, though not straightforward, implications for our understanding of both normative penetration and feasibility, respectively.

A. Normative penetration

Let's start with normative penetration. What, if anything, can our conclusions teach us about the phenomenon of normative penetration more generally – over and above the fact that judgements about feasibility appear to be among the ever-growing class of judgements that are subject to it?

It depends. In particular, it depends on whether we accept what we shall call the *unity thesis*: the thesis that normative penetration must be susceptible to a *unified explanation* across all of its instances – and, hence, that we should be in the market for an account that succeeds in explaining each and every instance of it. This is Knobe's (2010) own preferred view, and he takes a single explanation of normative penetration to apply to judgements involving intentional action, causation, knowledge, freedom, happiness, and other concepts.

If the unity thesis is correct, then our conclusions have some rather striking *negative* implications. For they give us reason to reject alternative explanations of normative penetration in *other* domains, such as Knobe's competence-based

contextualist explanation of normative penetration in the case of our judgements about intentional action. If it is incumbent on a satisfactory explanation of normative penetration to explain each and every instance of it, then the evidence we have presented in favour of the conflation interpretation is also evidence against such alternative explanations.

At the same time, it also means that we ourselves face a formidable challenge: to show that the conflation view can be *extended* in a way that would allow it to explain each and every instance of normative penetration. Extending the conflation interpretation in this way does not strike us as a hopeless prospect. Consider *intentional action*. The analogue of the conflation interpretation here would be something like this: a) Participants are making an extensional error in ascribing intentional action, e.g., in the case of the chairman intentionally harming the environment in Knobe's classic study because b) they are conflating the question of whether he intentionally harmed the environment with the question of whether he intentionally performed some other action (say, intentionally implemented the program that would result in the environment being harmed) and c) they are conflating these question because the second is more salient given their normative beliefs.

But even if the conflation interpretation cannot plausibly be extended in this way, this does not necessarily spell doom for it. For we might *reject* the unity thesis. Perhaps, on the contrary, we should expect to find that the best interpretation of normative penetration in one domain is quite different from the best interpretation of it in

another.[21] If this is right, then the fate of the conflation interpretation would *not* depend on our being able to extend it to each and every other instance of normative penetration. It would also mean that we cannot use our conclusions to cast doubt on alternative explanations in other domains.

There is one final possibility. We have been assuming that the question of how unified a phenomenon normative penetration is independent of our conclusions about normative penetration in the case of our judgements about feasibility. But suppose that you are convinced by our conclusions and yet also convinced of the fact that an alternative explanation is the correct one in some other domain (say, in the domain of intentional action). In that case, our conclusions here provide additional evidence against the unity thesis itself.

B. Feasibility

Our conclusions also have potentially significant implications for our understanding of *feasibility*. Consider, for example, the issue of whether feasibility has "an inescapable normative element" (Miller 2013, p. 32). As mentioned, some theorists hold that claims about feasibility are indeed partly normative (Raiika 1998; Buchanan 2004; Miller 2013). Others deny this and hold that claims about feasibility are prior to, and independent of, normative claims. These include proponents of prominent possibility-based accounts of feasibility (such as Wiens 2015), probability-based accounts (such as Gilabert and

[21] Hindriks (2014) argues against a unified interpretation of the asymmetries in judgements about different concepts that Knobe and others have identified. For an error theory of the original Knobe Effect, see Nadelhoffer (2004).

Lawford-Smith 2012), disposition-based accounts (such as Stemplowska 2016 and Southwood 2016); and function-based accounts (such as Southwood 2022).

Our conclusions suggest that normative accounts of feasibility are mistaken. At the very least, they are not well motivated by our actual linguistic practices. On the contrary, these accounts square poorly with such practices given that, as we have seen, normative penetration can be nullified. Similarly, opposition to non-normative accounts is not well motivated by such practices. On the contrary, such accounts are perfectly compatible with the linguistic data. They only *seem* incompatible insofar as we are tempted to give a *semantic* interpretation of the normative penetration thesis: to claim that normative considerations play a role in determining the semantic content of our feasibility ascriptions. However, the conflation interpretation instead holds that normative attitudes play a *causal* role with regard to our feasibility ascriptions; they cause us to conflate the question that is at issue (whether it is feasible to X) with some other question that is more salient in light of our normative attitudes (whether it is feasible to X in a certain way). Thus, to the extent that the case for normative accounts rests on an implicit appeal to our linguistic practices, it follows that they are not well motivated simpliciter. Moreover, if we think that fidelity to our linguistic practices is a legitimate constraint on any satisfactory account of feasibility, then our conclusions give us some reason to think, not merely that normative accounts are not well motivated, but that such accounts are false and that we should accept some non-normative account instead.

Our conclusions also serve to draw our attention to certain significant *dangers* associated with our practices of making and using feasibility claims. The conflation interpretation holds that we are disposed to make extensional errors about feasibility. To be sure, sometimes these errors may be benign or even beneficial. Plausibly, the case of the ship's captain is a case of just the latter sort. However, in other cases the errors may be anything but benign. Suppose that a U.S. citizen judges that it is not feasible for the United States to achieve a carbon-neutral economy because she judges that it is not feasible to achieve a carbon-neutral economy without changes to the lifestyles of ordinary U.S. citizens. Suppose, moreover, that the citizen conflates these two issues because they enjoy their current lifestyle and consumption patterns immensely and find political proposals that put the onus of change on ordinary people displeasing. Further, imagine that on the basis of their judgement that the U.S. achieving a carbon-neutral economy is infeasible they refrain from deliberating about which politicians and proposals to support that would help the U.S. get closer to achieving this goal. Suppose also that there is nothing especially wrong with expecting U.S. citizens to change their lifestyles, at least to some degree, for the purpose of achieving a carbon-neutral economy. In this case, the feasibility statement that is most salient to this U.S. citizen is not the right feasibility statement to consider in determining which politicians and proposals to support or, more broadly, whether the U.S. government ought to attempt to achieve a carbon-neutral economy.

We suggest that using feasibility ascriptions of this kind to inform our normative thinking is distinctively dangerous. Not simply because it will produce normative errors,

but because the normative errors are likely to be especially resilient and difficult to overcome. Why? First, because mistaken feasibility ascriptions that are the product of conflation are not counterfactually response to evidence in the same way as other kinds of mistaken feasibility ascriptions. Our judgements that p are typically at least minimally counterfactually responsive to evidence regarding p in the sense that if the evidence against p mounts up to a sufficient extent we tend to revise the judgements. But judgements that Xing is infeasible that are the product of conflating whether X is feasible, on one hand, with whether Xing in a certain way is feasible, on the other hand, are not like that. So long as one is conflating these two questions, one is likely to continue regarding Xing as infeasible unless there is evidence that Xing in the relevant way is feasible. Second, because the ultimate drivers of the mistaken feasibility ascriptions – the mistaken normative attitudes that dispose us to conflate whether it is feasible to X and whether it is feasible to X in a certain way – are hidden from view and thereby isolated from scrutiny. And third, because treating these feasibility ascriptions as determining the domain of deliberation means, in effect, that our mistaken feasibility ascriptions and normative attitudes are not subject to one familiar corrective, namely deliberation.

What is the upshot? First, we should be alert to a certain potential danger: that certain normative assumptions that may not themselves be subject to the requisite level of scrutiny can lead to an indefensible shrinking of the deliberative agenda. Second, to protect ourselves from this danger, we should try to be as clear and explicit about the content of our feasibility ascriptions as possible. In particular, we should be relatively

clear and explicit regarding *ways* of bringing about states of affairs, rather than allowing others, or even ourselves, to dismiss the achievement of important goals, such as a carbon-neutral economy, as infeasible simpliciter.

5. Conclusion

Normative penetration and feasibility are both philosophically important topics that have generated significant, though hitherto unconnected, research programs in their respective sub-fields. In this paper we have sought to connect them in the hope of shedding light on both. We have argued that there is compelling empirical evidence that our feasibility ascriptions are indeed subject to normative penetration, that the best explanation of it is provided by the conflation interpretation, and that these conclusions have potentially significant implications for our understanding of both normative penetration and feasibility.

References

Adams, Fred and Annie Steadman. 2004a. "Intentional action in ordinary language: core concept or pragmatic understanding?" *Analysis*, 64, 173-81.

Adams, Fred and Annie Steadman. 2004b. "Intentional action and moral considerations: still pragmatic," *Analysis*, 64, 268-76.

Adams, Fred and Annie Steadman. 2007. "Folk concepts, surveys, and intentional action." In *Intentionality, deliberation, and autonomy: the action-theoretic basis of*

practical philosophy, eds. Christoph Lumer and Sandro Nannini. Aldershot: Ashgate Publishers, 17-33.

Barry, Christian, Matthew Lindauer, and Gerhard Øverland. 2014. "Doing, allowing, and enabling harm: an empirical investigation," *Oxford Studies in Experimental Philosophy*, 1.

Beebe, James and Wesley Buckwalter. 2010. "The epistemic side-effect effect," *Mind and Language*, 25, 474-98.

Bicchieri, Cristina. 2006. *The Grammar of Society*. Cambridge: Cambridge University Press.

Blome-Tillmann, Michael. 2013. "Conversational implicatures (and how to spot them)," *Philosophy Compass*, 8, 170-85.

Brennan, Geoffrey, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood. 2014. *Explaining Norms*. Oxford: Oxford University Press.

Buchanan, Alan. 2004. *Justice, Legitimacy, and Self-Determination: Moral Foundations for International Law*. New York: Oxford University Press.

Cappelen, Herman and Ernie Lepore. 2005. *Insensitive Semantics*. Blackwell: Oxford.

Cushman, Fiery, Joshua Knobe, and Walter Sinnott-Armstrong. 2008. "Moral appraisals affect doing/allowing judgments," Cognition, 108, 353-80.

Finlay, Stephen. 2014. *Confusion of Tongues*. Oxford: Oxford University Press.

Fraser, Ben and Joshua Knobe. 2008. "Causal judgment and moral judgment: Two experiments." In *Moral psychology volume 2: The cognitive science of morality: Intuition and diversity*, ed. Walter Sinnott-Armstrong. MIT Press.

Gilabert, Pablo and Holly Lawford-Smith. 2012. “Political Feasibility: A Conceptual Exploration,” *Political Studies*, 60, 809-25.

Grice, H. Paul 1975. *Studies in the way of words*. Cambridge, Mass: Harvard University Press.

Hindriks, Frank. 2014. “Normativity in Action: How to Explain the Knobe Effect and its Relatives.” *Mind and Language*, 29, 51-72.

Khoo, Justin. 2015. “Modal disagreements,” *Inquiry*, 58, 511-534.

Khoo, Justin and Joshua Knobe. 2018. “Moral Disagreement and Moral Semantics.” *Noûs*, 52, 109-143.

Khoo, Justin and Jonathan Phillips. 2019. “New Horizons for a Theory of Epistemic Modals,” *Australasian Journal of Philosophy*, 97, 309-324.

Knobe, Joshua. 2003a. “Intentional action and side-effects in ordinary language,” *Analysis*, 63, 190-94.

Knobe, Joshua. 2003b. “Intentional action in folk psychology: an experimental investigation,” *Philosophical Psychology*, 16, 309-23.

Knobe, Joshua. 2004. “Intention, intentional action, and moral considerations,” *Analysis*, 64, 181-87.

Knobe, Joshua. 2010. “Person as scientist, person as moralist.” *Behavioral and Brain Sciences*, 33, 315-29.

Knobe, Joshua and Zoltán Gendler Szabó. 2013. “Modals with a taste of the deontic,” *Semantics & Pragmatics*, 6, 1-42.

Knobe, Joshua and Erica Preston-Roedder. 2009. “The ordinary concept of valuing,” *Philosophical Issues*, 19, 131-47.

Kratzer, Angelika. 1977. “What ‘Must’ and ‘Can’ Must and Can Mean,” *Linguistics and Philosophy*, 1, 337-355.

Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University Press.

Miller, David. 2013. *Justice for Earthlings: Essays in Political Philosophy*. Cambridge University Press.

Nadelhoffer, Thomas. 2004. “On Praise, Side Effects, and Folk Ascriptions of Intentionality.” *Journal of Theoretical and Philosophical Psychology*, 24, 196-213.

Nichols, Shaun and Joseph Ulatowski. 2007. “Intuitions and individual differences: the Knobe effect revisited,” *Mind and Language*, 22, 346-65.

Phillips, Jonathan and Joshua Knobe. 2009. “Moral judgments and intuitions about freedom,” *Psychological Inquiry*, 20, 30-6.

Phillips, Jonathan, Sven Nyholm, and Shen-yi Liao. 2014. “The good in happiness,” *Oxford Studies in Experimental Philosophy, 1*, 253-93.

Räikkä, Juha. 1998. “The Feasibility Condition in Political Theory,” *The Journal of Political Philosophy*, 1, 27-40.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Southwood, Nicholas. 2016. “Does ‘ought’ imply ‘feasible’?” *Philosophy and Public Affairs*, 44, 7-45.

Southwood, Nicholas. 2018. “The feasibility issue,” *Philosophy Compass*, 13, 1-13.

Southwood, Nicholas. 2022. “Feasibility as deliberation-worthiness,” *Philosophy and*

Public Affairs, 50, 121-62.

Stemplowska, Zofia. 2016. “Feasibility: individual and collective,” *Social Philosophy and Pollcy*, 33, 273-91.

Valentini, Laura. “Ideal vs. Non-Ideal Theory: A Conceptual Map.” *Philosophy Compass*, 7, 654-664.

von Fintel, Kai and Gilles, Anthony. “Might made right,” *Epistemic Modality*, eds. Andy Egan and Brian Weatherson. Oxford: Oxford University Press.

Wiens, David. 2015. “Political Ideals and the Feasibility Frontier,” *Economics and Philosophy*, 31, 447-77.