6

In Defense of 'Ought Implies Can'

Derek Leben, University of Pittsburgh, Johnstown

1. The Evidence against Ought Implies Can

Consider the following intuitive judgment: it only makes sense to say that an agent 'ought to do action A' whenever that agent is actually capable of performing action A. This intuition is called the 'Ought Implies Can' (OIC) Principle, and is often attributed to Kant,[1] due to passages like the following:

> duty commands nothing but what we can do . . . For if the moral law commands that we ought to be better human beings now, it inescapably follows that we must be *capable* of being better human beings.   (Kant, 1793: 47–50)

Assuming that intuitions about the OIC Principle are robust across speakers, there are two possible hypotheses for what is causing it. The first hypothesis I will call the 'semantic' hypothesis:

> **Semantic Hypothesis**:   The intuitive appeal of the OIC Principle is caused by a necessary connection between the lexical concepts OUGHT and CAN.

The Semantic Hypothesis is a natural reading of what Kant means by "it inescapably follows that," and has been endorsed by philosophers like

---

[1] Although there is debate about whether Kant himself actually endorsed it; see Ranganathan (2010) for dissent.

Vranas (2007). On the other hand, there is another hypothesis for the origin of OIC, most famously endorsed by Sinnott-Armstrong (1984):

Pragmatic Hypothesis: The intuitive appeal of the OIC Principle is caused by conversational constraints on the way that speakers use terms like 'ought.'

There are several ways that pragmatic factors might create an implication between attributions of obligations and abilities. For example, Gricean Maxims about relevance may compel a speaker to offer advice only when people are capable of taking it. This would make it conversationally pointless to attribute obligations to agents who are incapable of following them, but not conceptually impossible.[2]

Two recent papers have attempted to test these hypotheses, and presented evidence against the Semantic Hypothesis. I aim to criticize these papers and defend the possibility of the Semantic Hypothesis. In the first paper, Buckwalter and Turri (2015) presented subjects with a wide variety of scenarios where an agent normally would have an obligation, but the agent is suddenly prevented (either psychologically or physically) from being able to fulfill that obligation. Here are three of the scenarios:

Walter Obligation
Walter promised that he would pick up Brown from the airport. But on the day of Brown's flight, Walter is in a serious car accident. As a result, Walter is not physically able to pick up Brown at the airport.

Michael Obligation
Michael is a playground safety worker. He sees some broken glass in an area where kids sometimes play barefoot. But he is stricken by a sudden paralysis in his legs. As a result, Michael is not physically able to pick up the glass.

---

[2] Both hypotheses endorse the claim that 'ought *implies* can,' but they disagree about the strength of this implication. The semantic hypothesis interprets it as conceptual entailment, while the pragmatic hypothesis interprets it as pragmatic implicature. Thus, the title of this chapter should actually be: 'In Defense of the Semantic Interpretation of Ought Implies Can,' but the convention in the literature is to refer to the semantic hypothesis as endorsing OIC and the pragmatic hypothesis as denying it, so I will follow suit.

Jessica Obligation

Jessica is a lifeguard at a remote ocean beach. Two struggling swimmers are about to drown. Jessica rushes in to save them. But because of the very far distance between the swimmers, it is physically impossible for her to rescue both swimmers. Jessica rescues the one swimmer but not the other.

The authors forced participants to consider both obligations and inabilities simultaneously, by picking the option that best applies from the following list:

1. [Agent] is obligated to [A], and [Agent] is physically able to do so.
2. [Agent] is obligated to [A], but [Agent] is not physically able to do so.
3. [Agent] is not obligated to [A], but [Agent] is physically able to do so.
4. [Agent] is not obligated to [A], and [Agent] is not physically able to do so.

For example, 'Walter is obligated to pick up Brown at the airport, and Walter is physically unable to do so,' etc. The results that Buckwalter and Turri found are striking. In almost every condition, the vast majority of participants selected the 'obligated but unable to' option (Walter: 80%, Michael: 88%, Jessica: 93%). The authors found that altering the scenarios from first to third person and altering the wording from 'ought' to 'duty' to 'obligation' also made little difference in these results. The conclusion that Buckwalter and Turri draw from this data is: "Common-sense moral cognition rejects the principle that ought implies can."

In another recent paper, Chituc et al. (2016) attempted to show that judgments about blame can lead speakers to attribute obligations in the absence of ability.[3] They first presented subjects with a scenario similar to the 'Walter' condition above, where an agent named Adams agrees to meet his friend Brown at 12 o'clock noon, but it takes him 30 minutes to reach the lunch destination (it's always Brown who gets stood up in these experiments). In a low-blame condition, the agent's car breaks down at

---

[3] For the rest of the chapter, instead of using the longer phrases like 'attributions of ability' or 'judgments about ability,' I will simply talk about 'ability,' and allow the context to disambiguate whether I am talking about abilities themselves (de re) or speaker judgments about abilities (de dicto).

11:45, and in a high-blame condition, he simply decides not to go. Participants were then asked to rate their agreement with the following: "At 11:45, it is still true that Adams ought to meet Brown at noon." The authors found that people were significantly more likely to agree with this statement in the high-blame condition, despite the fact that at 11:45, neither agent was technically *able* to meet his friend at noon (remember, it takes 30 minutes to get to the lunch destination). A second experiment presented the following scenario:

> **Adams Obligation**
> Brown is a CEO of a large company in the economic boom in the middle of the 20th Century. At 2 o'clock, Brown has a meeting in the city to make a significant financial decision that will decide the future of his company. Since so much money is at stake, he asks his trusted personal advisor, Adams, to meet him on the 12 o'clock train. On the train, he plans to discuss his decision on the ride into the city, where Brown will go straight to his 2 o'clock meeting. Adams promises to meet Brown on the train at noon. It takes Adams thirty minutes to drive to the train station, park, purchase a ticket, and board the train. However, fifteen minutes after leaving at eleven-thirty, Adams' car breaks down unexpectedly. Because his car is not working at the time, Adams cannot meet Brown at noon, as promised. Since cell phones have not been invented yet, Adams has no way to contact him.

Participants were then asked to rate their agreement with the following statements:

Ought: "At 11:45 AM, Adams ought to keep his promise"
Ability: "Adams can keep his promise"
Blame: "Adams is to blame for not keeping his promise"

The researchers found that 'ought' judgments are correlated with 'blame' judgments, but not 'can' judgments (although, interestingly, 'can' and 'blame' judgments were correlated). Finally, a third experiment carried over these designs to non-moral applications of 'ought' and found similar results. They concluded that attributions of blame are a better predictor of obligation than ability, and thus the Semantic Hypothesis is false.

Before moving on to criticisms of these two experiments, I think it's worth dwelling on just how surprising their results are. Imagine that I go

to visit Walter in the hospital after his car accident, flowers in hand, and say "I'm really sorry to hear about your two broken legs, but you really should have picked up Brown." To my own ears, this sounds absurd, and yet Buckwalter and Turri are reporting huge numbers of normal speakers (up to 92%) making this judgment. Furthermore, I can't think of any examples when I have blamed someone for an action that I genuinely believed that person was incapable of doing, and yet this is what large numbers of people appear to be doing in Chituc et al.'s experiments. Of course, I might be so blinded by philosophical theories that I am unable to imagine judgments that are perfectly normal for speakers beyond the walls of academia. However, you don't have to be a Bayesian to agree that we should be more critical of surprising results than unsurprising results.

While I applaud the researchers on their empirical approaches to the question of OIC, I believe that *both* papers do indeed contain experimental designs that are testing different interpretations of 'ought'-claims. There are several ways the term 'ought' may have different interpretations in a vignette. Chituc et al. describe this objection and their reply to it:

Philosophers still committed to [Semantic OIC] may respond that our experiments don't capture the relevant concept of 'ought,' perhaps because 'ought' is polysemous—that is, it has multiple meanings. While this may be true, it is entirely ad hoc to insist that the participants in our experiments are utilizing a different meaning of 'ought.' There is no reason to posit different meanings of 'ought' other than to preserve the truth of [the Semantic Hypothesis] in light of contrary evidence. Until defenders of the principle supply an independent argument for the claim that 'ought' has a different meaning in our examples . . . this response merely begs the question.

Chituc et al. are correct that, if one is going to appeal to 'ought' having a different meaning in these experiments, it is important to clearly identify *why* it has a different meaning and bring in additional evidence to show that this difference in meaning leads to their experimental results.

One cause of different interpretations of 'ought'-claims is temporal context-sensitivity. For most of the conditions in both experiments, there is some point (A) at the beginning of the vignette when the character is able to fulfill his or her obligations, and another point (B) when the character is unable to do so. Perhaps the claim '*x* ought to A' is temporally underspecified, and most participants are applying it to time A rather than time B. Clearly, the authors of both papers were aware of this potential confusion and attempted to control for it. Buckwalter and

Turri worded their statements in the present tense and constructed the responses as conjunctions of obligations and abilities, such that participants were forced to select 'obligated and unable.' The only time at which the character was both obligated and unable was time B, so that seems to resolve the temporal underspecification. Chituc et al. explicitly specified a time in the story, 11:45, which seems to be located in time B, also apparently eliminating the temporal underspecification. However, I'll argue that there are some problems with the ways in which both sets of authors have attempted to control for temporal context-sensitivity, and when the time of the story is made more explicit, the results are significantly different.

In addition to temporal underspecification, which is a general feature of any claim, there may be polysemy in the semantics of 'ought' itself. This wouldn't be surprising, since the most influential current semantic theory for 'ought' already posits such polysemy (Kratzer, 1977, 1981).[4] We might add to this polysemy a naïve theory about how speakers make judgments about obligations: they consider both prior obligations and the present circumstances together in a weighting process that produces as its output an 'all-things-considered' judgment. Prior obligations are those that previously existed and would normally apply in the current situation, *ceteris paribus* (these are also called *prima facie* or *pro-tanto* obligations). For example, if I have children, then (all else being equal) I probably have an obligation to see that they are taken care of. Although these *originate* in the past, they are not *indexed* to the past, and it is perfectly grammatical to express them using the present tense: 'I *have* a prior obligation.'[5] On the other hand, all-things-considered obligations can be described as the product of weighing prior obligations against each other and the current circumstances. All-things-considered 'oughts'

---

[4] Kratzer's theory seeks to unify epistemic, deontic, teleological, and other interpretations of 'ought' with a core lexical meaning (universal quantification over sets of possible worlds) and two parameters for information supplied by conversational context, what she calls a 'modal base' (for restricting quantification) and an 'ordering source' (for ranking sets of worlds in a partial ordering). Kratzer has described deontic necessity modals as having a circumstantial base and an ordering source of relevant norms. If abilities are part of the circumstantial base, then a claim like 'Walter ought to pick up Brown' could be true of higher-ranked worlds outside the modal base. The naïve theory of obligation-judgments discussed here can be formulated in terms of Kratzer's theory, but it is also consistent with other semantic theories, such as force-dynamic approaches and probabilistic approaches (see Portner, 2011, for an overview of current theories).

[5] Similarly, if a speaker utters 'I have eaten lunch,' this statement is indexed to the present time, but part of its truth-conditions include an event in the past.

are non-defeasible and guiding rather than just reasons in favor of an action.[6] In Plato's example of borrowing a weapon from a friend who subsequently goes insane, an agent might still have a *prior* (prima facie) obligation to his friend, but this obligation can be outweighed by the circumstances and other obligations to fellow citizens. A speaker can say without contradiction: 'you have reason (prior obligations) to give the weapon back, but given the circumstances, you have most reason not to give the weapon back (all-things-considered obligations).'

Although both prior and all-things-considered obligations are indexed to the current time, there is a *logical* ordering to them. If our naïve theory described above is correct, then prior obligations (along with the current circumstances) serve as input to a decision-procedure that results in all-things-considered obligations. Thus, it would make sense to say something like: 'First, there is an obligation to give my friend's weapon back (along with other obligations not to do harm to fellow citizens). Second, there is the fact that he has gone insane and will hurt people with the weapon. Finally, after appropriately weighing (1) and (2), there is an all-things-considered judgment that I have an obligation not to give the weapon back.' Once again, this ordering is not necessarily a temporal ordering, but a logical ordering in the process of making ought-judgments. Of course, prior obligations do originate in the past, so making an earlier time salient may serve to highlight this part of the process (e.g., if I say 'remember that you did promise to give the weapon back,' this will highlight prior obligations).

The next two experiments will try to show that, when context-sensitivity and polysemy are made more explicit, the results of the experiments discussed in this section are undermined. The first experiment focuses on the temporal context-sensitivity, while the second experiment tests the order polysemy specific to the semantics of 'ought.'

2. Timeline Experiment

When I informally presented Buckwalter and Turri's stimuli to some groups of students, they initially responded the same way as most

---

[6] I will not take a stand on any of the more specific debates about all-things-considered 'ought,' such as whether it includes epistemic obligations (Booth, 2012), exactly how different types of obligations are weighted (Chang, 2004), or whether all-things-considered 'ought' is synonymous with rational behavior.

subjects in the original experiment (almost all chose 'obligated, and unable to'). I then followed up by asking them if they were really saying that, *after the car accident*, Walter was still obligated to pick up his friend. Most of them quickly modified their answer, explaining: "no, we were talking about before the accident." Of course, the students may have been responding to the pressure of their professor second-guessing their responses, so this can't be considered as valid counter-evidence. But at least it provides some initial suspicions that temporal underspecification is still not resolved.

In almost all of the vignettes, the character initially had an ability to fulfill his or her obligation (time A), but later lacked the ability (time B). For example, in the Walter scenario, he had the ability to pick up Brown before the car crash, but not after it. Subjects were asked about Walter's obligation without specifying a particular point in the story, leaving it underspecified between time A and time B. In the Chituc et al. stories, the authors were careful to specify the moment of the car breaking down (at 11:45), but this is actually the exact dividing point between time A and time B. This is an odd time to ask about ability; it strikes me as odd to ask whether someone is 'disabled' the exact moment of an accident, although it is obviously true after the accident and false before it. Thus, to clarify both sets of experiments, I ran a version of the Walter experiment with an illustrated timeline intended to more clearly distinguish between the two. Participants in one condition saw an arrow that indicates an obligation prior to the car crash, while those in another condition saw an arrow indicating the obligation after the car crash.

Eighty participants were randomly assigned to one of two conditions, to match the design of the original experiments. Participants were selected from the United States using Amazon's Mechanical Turk service and SurveyMonkey. Participants were paid $0.30 for an average rate of $11.47/hr. Both groups were presented with the original Walter scenario, and were then shown a screen with the description: "below is an illustrated timeline of the events in this story." Both timelines had arrows (yellow in the original experiment), but one arrow indicated a point prior to the car crash (Figure 6.1), while another arrow indicated a point after the car crash (Figure 6.2).

Participants were then asked:

At the time of the yellow arrow, do you agree with the following statement: Walter ought to pick up Brown at the airport.
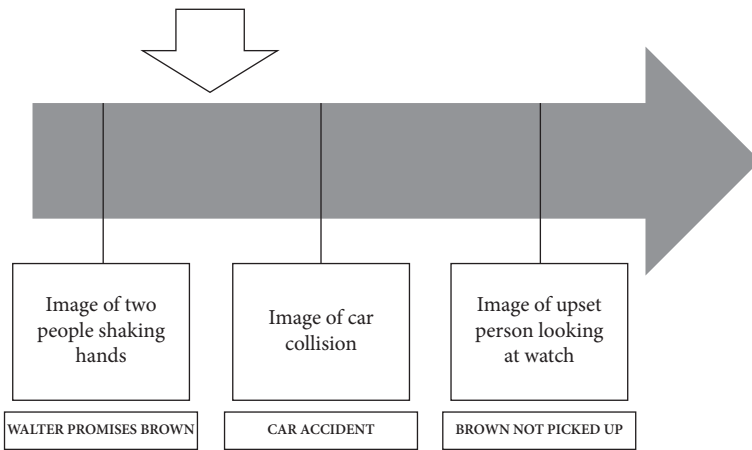
Figure 6.1 Timeline diagram, prior condition. Images used in the original experiment omitted for copyright reasons.
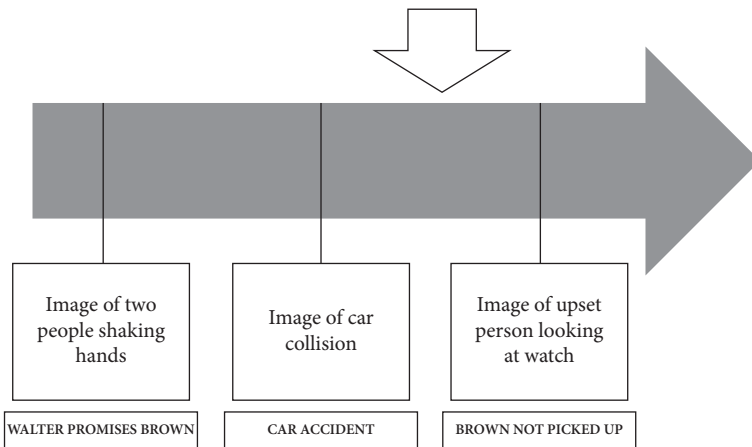

Figure 6.2 Timeline diagram, current condition. Images used in the original experiment omitted for copyright reasons.

A Likert scale was presented from 1 ('completely disagree') to 6 ('completely agree').[7] The questions about ability and responsibility were

---

[7] In the previous experiment, a 7-point Likert scale was used because Buckwalter and Turri also used a 7-point scale, and the goal was to match their design as closely as possible. Here, a 6-point scale is used because I believe that eliminating the middle option (i.e., 4, 'not sure')
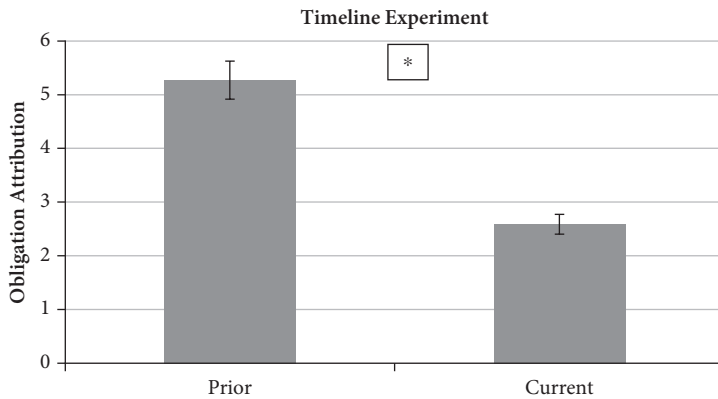
Figure 6.3 Timeline experiment results.

omitted in this version. The same demographic questions were presented. Responses of the *prior* group were significantly higher than those in the *current* group, $t(79) = 6.74$, $p < 0.001$ (Figure 6.3), $M_{prior} = 5.27$, SD = 2.1. $M_{current} = 2.58$, SD = 1.28.

These results demonstrate that, when the time is explicitly specified as either definitively before or after the accident, practical ability has a significant influence on whether the participants judge that the agent 'ought' to pick up Brown.

Although the mean responses for current obligations were below the dividing point, a critic might reply that this is still far higher than the 0% that the semantic explanation for OIC would predict. I agree, but as suggested in section 1, temporal context-sensitivity is not the only cause of differences in meaning here. I think there is also a non-temporal ordering of obligations, even indexed to the current time, that is also causing problems for the experimental results.

3. Ordering Experiment

In every condition of Buckwalter and Turri's experiments, the question places an ought-claim sequentially before an ability claim:

forces participants to make a judgment. Use of the forced-judgment approach is controversial, but the results here are large enough that I suspect they will replicate on a 7-point scale as well.

Walter is obligated to pick up Brown, and he is physically unable to do so.

If our naïve theory is correct, this suggests that the obligation the authors are referring to is the *prior* obligation, since it is sequentially before the clause about ability. After all, it sounds normal to say, "I ought to help out, but I can't," while it sounds odd to say, "I can't help out, but I ought to." Although Buckwalter and Turri seem to be using the word 'and' to indicate logical conjunction, most people typically interpret it as progression,[8] so it is important to see if this ordering effect is present.

To test for the effect of ordering, I re-ran three of the conditions from Buckwalter and Turri's study along with three versions of the same stories with the ordering of obligation and ability reversed. Thus, rather than "*x* is obligated to A, and(s)he is unable to do so," the new question read: "*x* is unable to do A, but she is still obligated to do so." I also changed the wording to try and emphasize this change of ordering. The connectives 'but still' were used to emphasize that the question is referring to all-things-considered obligations rather than prior ones. Other phrases were also used to emphasize this in the alternative options, including 'no longer obligated to do so.' Thus, the experiment should be considered to be both a change in the wording and in the sequence of propositions (perhaps future replications will distinguish if one of these changes is driving the effect more than another).

To follow the original design as closely as possible, I assigned 240 participants randomly to one of six conditions. All were selected from the United States using Amazon's Mechanical Turk service and Survey-Monkey. Participants were paid $0.30 for an average rate of $18.00/hr. Three of the six conditions replicated Buckwalter and Turri's original design, while three of them kept the same story and used the reverse wording. The stories were the same three described above, Walter, Michael, and Jessica. As with the original design, the order of questions (1–4) was randomized. The original and reversed Walter scenarios were as follows:

Walter Obligation (original)
Walter promised that he would pick up Brown from the airport. But on the day of Brown's flight, Walter is in a serious car accident. As a result, Walter is not physically able to pick up Brown at the airport.

---

[8] E.g., "they got pregnant and got married" strongly implies a shotgun wedding, while remaining logically constant with virginity at the wedding.

1. Walter is obligated to pick up Brown at the airport, and Walter is physically able to do so.
2. Walter is obligated to pick up Brown at the airport, but Walter is not physically able to do so.
3. Walter is not obligated to pick up Brown at the airport, but Walter is physically able to do so.
4. Walter is not obligated to pick up Brown at the airport, and Walter is not physically able to do so.

Walter Obligation (reversed)

Walter promised that he would pick up Brown from the airport. But on the day of Brown's flight, Walter is in a serious car accident. As a result, Walter is not physically able to pick up Brown at the airport.

1. Walter is physically able to pick Brown up at the airport, and he is obligated to do so.
2. Walter is physically able to pick Brown up at the airport, but he is not obligated to do so.
3. Walter is not physically able to pick Brown up at the airport, so he is no longer obligated to do so.
4. Walter is not physically able to pick Brown up at the airport, but he is still obligated to do so.

The original and reversed Michael and Jessica scenarios had the same format. After each condition, participants were then asked whether they agreed with the following statement:

[Walter/Michael/Jessica] is literally unable to [action]

where the actions were either "pick up Brown from the airport," "pick up the glass," or "save both of the swimmers." After this, the next question asked to rate agreement with the following statement:

[Walter/Michael/Jessica] deserves to be blamed for the fact that [result]

where the results were either "the fact that Brown was not picked up," "the fact that the glass was not picked up," or "the fact that both swimmers were not saved." The ratings were on a Likert scale from 1 ("strongly agree") to 7 ("strongly disagree"). All conditions then asked demographic questions about age, gender, and education level.
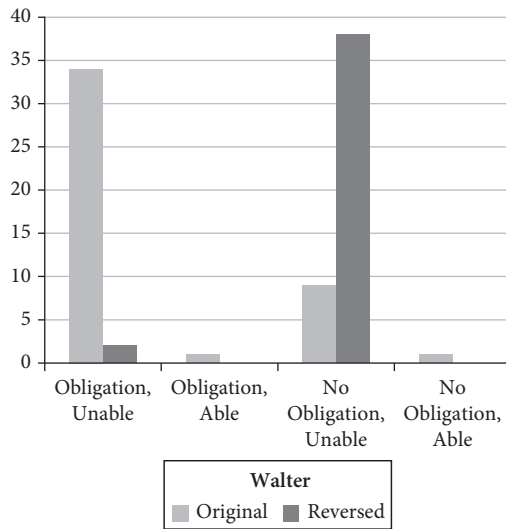
Figure 6.4 Ordering results Walter condition.

In the original conditions, the results of Buckwalter and Turri were very similar, with the vast majority of participants selecting the "obligated but unable" option. However, in the conditions where wording was reversed, the effects were significantly different in the Walter case (Figure 6.4), X^2 $(3, N = 85) = 48.21, p < 0.0001$, the Michael case (Figure 6.5), $X^2 (3, N = 78)$ $= 18.97, p = 0.0003$, and the Jessica case (Figure 6.6), $X^2 (3, N = 78) = 23.08$, $p < 0.0001$. For the inability and responsibility questions, responses were similar to the original experiment across all conditions.

These results demonstrate that the order of obligation and ability is extremely important. When the question is constructed to indicate obligation before inability, Buckwalter and Turri's results are replicated. However, when the question is constructed to indicate inability before obligation, the results are significantly different in all conditions, and almost reversed in the Walter and Jessica cases. Interestingly, the Michael case still generates an almost 50/50 split, with more participants still violating OIC. However, if this were to be taken as evidence against OIC, one would expect it to carry over to the other two cases (Chituc et al.'s hypothesis doesn't help here, since there are no significant differences in
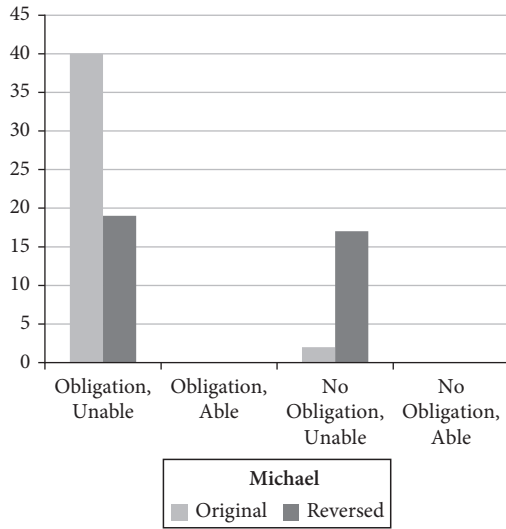
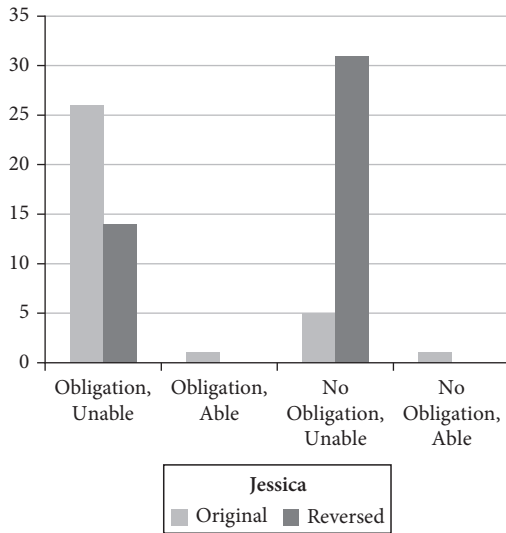**Figure 6.5** Ordering results Michael condition.

Figure 6.6 Ordering results Jessica condition.

blame between the conditions).[9] I propose that the results of this experiment are best explained by positing polysemy between two distinct parts of the semantics of 'ought'-claims, one related to prior obligations, and the other related to all-things-considered obligations.

A critic might object that, at the very least, the experiments have shown that *prior obligations* do not entail ability. Perhaps this is true, but to my knowledge, everyone endorsing OIC has in mind all-things-considered obligations. As opposed to prior obligations, all-things-considered 'oughts' are non-defeasible and guiding rather than just reasons in favor of an action. Thus, these 'oughts' must be the target of anyone looking to investigate the relevant concepts in the philosophical debate.

4. Conclusion

Semantic OIC is an empirical hypothesis about the structure of our concepts, and the authors of the papers criticized here are exactly right for using the methods of the cognitive sciences to investigate it. This chapter has argued that the evidence offered against Semantic OIC is unconvincing once we consider differences in the meaning of 'ought' within the experimental designs. The timeline experiment suggests that speakers provide dramatically different responses once the time is explicitly specified; the ordering experiment shows that responses are also significantly different when wording and sequence are changed to suggest a different aspect of 'ought'-claims (prior vs. all-things-considered). However, it's easy to throw stones, so I will end on two positive suggestions. The first is that perhaps the authors are correct that prior obligations are independent of ability. This would be interesting, since it is not introspectively obvious what happens to prior obligations once they are defeated (e.g., Plato's obligation to return the weapon; does this obligation simply disappear or does it still exist in some non-guiding sense?). The other positive suggestion is that we have made progress in

---

[9] There are several possibilities why the Michael case might be different; perhaps participants simply do not believe that he really has 'sudden paralysis' in his legs, which is unexplained by the vignette, or they believe that he could still drag himself over and clean up the glass. This sounds silly, but many of my students who heard these stories in informal discussions insisted that Walter was still able to have someone pick up Brown, even if he were in a serious car accident. Of course, further manipulations would be necessary to test these possibilities.

clarifying and identifying exactly what sense of 'ought' is at issue, and this can help fuel future empirical work.

References

Booth, A. (2012). All Things Considered Duties to Believe. *Synthese* 187: 509–17.

Buckwalter, W. & Turri, J. (2015). Inability and Obligation in Moral Judgment. *PLoS ONE*, 10 (8).

Chang, R. (2004). All Things Considered. *Philosophical Perspectives* 18: 1–22.

Chituc, V., Henne, P., Sinnott-Armstrong, W. & De Brigard, F. (2016). Blame, Not Ability, Impacts Moral "Ought" Judgments for Impossible Actions: Toward an Empirical Refutation of "Ought" Implies "Can". *Cognition*, 150: 20–5.

Kant, I. (1793). *Religion Within the Limits of Reason Alone*. New York: Harper.

Kratzer, A. (1977). What *Must* and *Can* Must and Can Mean. *Linguistics and Philosophy*, 1: 337–55.

Kratzer, A. (1981). The Notional Category of Modality. In Hans-Jürgen Eikmeyer & Hannes Rieser (eds.), *Words, Worlds, and Contexts: New Approaches in Word Semantics* (Research in Text Theory 6), 38–74. Berlin: de Gruyter.

Portner, P. (2011). *Modality*. Oxford: Oxford University Press.

Ranganathan, S. (2010). Does Kant Hold That Ought Implies Can? In J. Sharma & A. Raguramaraju (eds.), *Grounding Morality*. London: Routledge.

Sinnott-Armstrong, W. (1984). "Ought" Conversationally Implies "Can." *The Philosophical Review*, 93: 249–261.

Vranas, P.M.B. (2007). I Ought, Therefore I Can. *Philosophical Studies*, 136: 167–216.