1. Introduction

In this paper I aim to make progress on two fronts, the semantics of indicative conditionals and the Newcomb problem, by investigating an interesting type of sentence, the *indicative dominance conditional*.\(^1\) There are two main claims.

The first claim is that evidence drawn from indicative dominance conditionals favors a Stalnaker-type\(^2\) semantics for indicative conditionals over its main rivals — the strict-conditional semantics due to Angelika Kratzer and the probabilistic theories relying on Adams’ Thesis (Adams theories, for short). I will argue for this by investigating indicative dominance conditionals, sentences like the following:

\[
(1) \text{ If I drink the (entire) contents of cup A, I will ingest more than if I drink the (entire) contents of cup B.}
\]

Intuitively, (1) is true in any situation in which we know that cup A has more tea than cup B. Of particular interest are *variably dominant* scenarios: cases where we know that cup A has more tea than cup B, but there is no amount X such that we know, of X, that the amount in cup A is greater than X and the amount in cup B less. I will argue that Kratzer’s semantics predicts that (1) is false in some variably dominant scenarios, and that Adams theories predict that (1) is unassertible in some such scenarios.

---

1. This paper grew out of my dissertation; I am in debt to my advisors Matti Eklund, Harold Hodes, and Nico Silins. Thanks to Michelle Kosch, Sally McConnell-Ginet, Zach Abrahams, David Liebesman, and Will Starr. Malte Willer and Charles Cross commented on parts of this paper at the 2012 and 2013 Central APA meetings.
2. I lay out a version of Stalnaker’s original 1975 semantics for indicatives in §4. But there are several views that agree with Stalnaker’s semantics in simple cases, yet diverge over more complicated conditional constructions (e.g., McGee [1985], and Heim [1992]). I will call the family of theories that agrees with Stalnaker’s truth-conditions for what I call *straight indicatives* *Stalnaker-type theories*, where a straight indicative has the form ‘if P, then Q’, where P and Q do not involve any modal material (in particular, no nested conditionals). The argument I present in §2–6 is not an argument for Stalnaker’s particular approach, but rather for a whole family of theories — the Stalnaker-type theories. The question of which Stalnaker-type theory is preferable requires separate treatment.
scenarios. But Stalnaker’s semantics predicts that (1) is true in all such cases. So, I claim, the predictions of Stalnaker-type semantics fit better with our intuitions. Since Stalnaker-type semantics is the only theory of indicatives that makes the right predictions, dominance indicatives provide an argument in favor of Stalnaker-type semantics and against its rivals (that’s the first main claim).

It has long been recognized that conditionals play a central role in motivating both the one-boxing and the two-boxing answers to the Newcomb problem. But the question of how conditionals guide us to these answers has not been posed, or answered. The question is of interest independently of whether one is convinced that one-boxing or two-boxing is the right answer. My second main claim is that the how question is answered by a principle I will call the Dominance Norm (DN), connecting the truth of dominance indicatives with rational courses of action. Given DN, our answer to the Newcomb problem can be tightly correlated with our views on the semantics of indicative conditionals, specifically, with our predictions for the Newcomb dominance conditionals:

(2) If I choose one box, I will win more than if I choose two boxes.
(3) If I choose two boxes, I will win more than if I choose one box.

(call these the one-boxing and the two-boxing dominance conditionals, respectively)

After defending DN against a potential objection stemming from DeRose (2010), I will look at what Adams theories and Stalnaker’s semantics tell us about the Newcomb problem in the light of DN. In particular, I argue that Stalnaker’s semantics for indicatives predicts that (3) is known to be true, and so commits one to two-boxing via DN. Combined with the argument for a Stalnaker-type semantics, this yields a new argument for two-boxing in the Newcomb scenario. But DN does not yield a sound argument for one-boxing on the supposition that the Adams semantics is right, contrary to a widely cited suggestion by Lewis (1981b). If DN is indeed the bridging principle that connects conditionals and rational deliberation, one-boxers cannot appeal to conditionals to help their case.

2. Dominance Indicatives

I am going to argue that, of the dominant theories of indicative conditionals, only Stalnaker-type semantics gives plausible verdicts for the type of dominance indicatives I will call variably dominant. I will restrict my attention to three prominent views on the semantics of indicatives: Stalnaker (1975), Kratzer (1986), and Adams theories — the theories based on Adams’ Thesis, like Edgington (1995). I propose to investigate the predictions of the three theories for indicative dominance conditionals, sentences like the following:

(1) If I drink the (entire) contents of cup A, I will ingest more than if I drink the (entire) contents of cup B.

When is (1) true? As it turns out, the three theories under investigation deliver interestingly different verdicts. I will proceed by comparing the predictions of the three theories for (1) in a few hypothetical scenarios. The main argument will revolve around the following scenario:

**Tea-party 1 (T1)**

Two cups of tea are on the table — cup A and cup B. We know that either cup A contains 2ml and cup B 1ml of tea, or cup A contains 4ml and cup B 3ml of tea.

The distinguishing feature of T1 is that, in T1, we know that cup A has Lewis’ influential discussion of the problem (Lewis [1981b]). For those already in the two-boxing camp, the main point of my argument in §7–9 is to clarify the connections between semantics and decision theory, and pin-point the sources of our two-boxing intuitions.

3. Does the two-boxing answer need a new argument in its favor? The recent writers on the Newcomb problem tend to agree that two-boxing is the right answer (e.g., Egan [2007], Wedgwood [2011], DeRose [2010], Joyce [2007]). But the recent unanimity does not go beyond acknowledging the intuitive force of
more tea than cup B, but there is no amount X such that we know, of X, that cup A has more tea than X, and cup B less. Let us call such a case a case where the amount in cup A \textit{variably dominates} the amount in cup B. In T1, one amount dominates another in the \textit{epistemic context of utterance} — that is, throughout the set of worlds consistent with what is known, or taken for granted, in the context of utterance.\textsuperscript{5}

The basic intuition that I am going to appeal to is that \textbf{(1)} is true in T1. Of course, the intuition is clearly stronger than that, since it is clear that \textbf{(1)} should also be true in all the cases where the amount in cup A dominates the amount in cup B. Now let's turn to our three theories.

3. Kratzer’s semantics

Let's start with Kratzer. Kratzer gives the following truth-conditions for what I will call \textit{straight indicatives}, that is, non-nested indicatives without an overt modal in the consequent:

\textit{Kratzer’s semantics for straight indicatives}

\textit{⌜If P , then Q⌝} is true just in case all P-worlds in C are Q-worlds\textsuperscript{6} (where C is the epistemic context of utterance, or the \textit{modal base}).

Since dominance conditionals are comparative constructions, we need a semantics for comparatives to derive Kratzer’s predictions for T1. Unfortunately, the semantics of comparatives is a matter of some dispute.\textsuperscript{7}

For the sake of simplicity of exposition, I will adopt Schwarzschild’s 2008 proposal, since its presentation does not need the apparatus of formal semantics. In the Appendix, I show that the relevant predictions are the same on the more standard degree-based and the alternative interval-based approach to the semantics of comparatives.

On Schwarzschild’s \textit{A-not-A} account, the meaning of

\textbf{(4)} \hspace{0.5cm} A is more expensive than B.

is given by

\textbf{(5)} \hspace{0.5cm} There is some expense threshold \(\theta\) such that A meets or exceeds \(\theta\), but B does not meet or exceed \(\theta\).

The following two examples from Schwarzschild (2008) illustrate an interesting flexibility of the threshold approach when it comes to dealing with quantifiers in the than-clause:

\textbf{(6)} \hspace{0.5cm} The balloon is higher today than it has been on any other day.

\textbf{(7)} \hspace{0.5cm} The balloon is higher today than it has been on at least one other day.

According to the A-not-A analysis, (6) and (7) both say that the balloon met or exceeded a threshold \(\theta\) today. But, further, (6) says that

\textbf{(8)} \hspace{0.5cm} The balloon did \textit{not} meet or exceed \(\theta\) on any other day.

while (7) says that

\textbf{(9)} \hspace{0.5cm} There was at least one other day on which the balloon did \textit{not} meet or exceed \(\theta\).

Schwarzschild’s proposal is that in order to provide adequate truth-conditions for comparatives with quantifiers in the than-clause, the A-not-A theory needs to allow the quantifiers to be interpreted sometimes inside (as in (8)) and sometimes outside the negation (as in (9)). The position outside the negation is the normal case. But negative polarity items and some modals are interpreted inside the negation. Here is an example with a modal:
The balloon is higher than it is allowed to be.

Now, consider the dominance conditional

If I drink the (entire) contents of cup A, I will ingest more than if I drink the (entire) contents of cup B.

The A-not-A approach suggests two possible logical forms, depending on whether the negation out-scopes the conditional in the than-clause or not. On one reading, (1) asserts that

There is a threshold \( q \) such that if I drink the contents of cup A, I will ingest an amount that meets or exceeds \( q \), and it is not the case that if I drink the contents cup B, I will ingest an amount that meets or exceeds \( q \).

On the second reading, (1) asserts that

There is a threshold \( q \) such that if I drink the contents of cup A, I will ingest an amount that meets or exceeds \( q \), and if I drink from cup B, I will not ingest an amount that meets or exceeds \( q \).

(11) and (12) are independent of any particular semantics of indicatives. What are the predictions when (11) and (12) are combined with Kratzer’s theory? Combined with Kratzer’s semantics, (11) gives:

There is a threshold \( q \) such that for all the worlds in C in which I drink from cup A, I ingest an amount that meets or exceeds \( q \), and it is not the case that for all worlds in C in which I drink from cup B, I ingest an amount that meets or exceeds \( q \).

(13) is true just in case the minimum amount I drink in the worlds in which I drink from cup A is greater than the minimum amount I drink in the worlds in which I drink from cup B.

On the other hand, (12), combined with Kratzer’s theory, gives:

There is a threshold \( \theta \) such that for all the worlds in C in which I drink from cup A, I ingest an amount that meets or exceeds \( \theta \), and for all worlds in C in which I drink from cup B, I do not ingest an amount that meets or exceeds \( \theta \).

(14) is true just in case the minimum amount I drink in the worlds in which I drink from cup A is greater than the maximum amount I drink in the worlds in which I drink from cup B.

But now consider the predictions of (13) and (14). First, it is clear that (13) does not work: that one minimum is greater than the other is no guarantee that if I drink from cup A I will ingest more than if I drink from cup B. The following scenario illustrates the problem:

**Tea-party 2 (T2)**

Two cups of tea are on the table — cup A and cup B. We know that cup A contains at least 10ml of tea, and cup B at least 1ml of tea, and nothing else of relevance (in particular we do not know that cup A contains more tea than cup B).

Clearly, (1) may well be false in T2. Yet (13) predicts that (1) is true. This is the wrong result.

What about (14)? (14) predicts that (1) will be false in T2, which is perhaps the right result. But (14) makes the wrong prediction for T1:

**Tea-party 1**

Two cups of tea are on the table — cup A and cup B. We know that either cup A contains 2ml and cup B 1ml of tea, or cup A contains 4ml and cup B 3ml of tea.

Here the A-minimum is 2ml, and the B-maximum is 3ml, so (1) comes out false according to (14). Yet it is clearly true that, in T1, if I drink

---

8. It is perhaps the right result because it may be that the description of the T2 scenario underspecifies the situation, so that (1) is true in some situations conforming to the description in T2 and false in others.
from cup A I will ingest more than if I drink from cup B. So (14), too, gives the wrong predictions.

One might think that (11) and (12) do not exhaust the possible logical forms of (1) on Schwarzschild’s view. (11) and (12) both interpret the comparative as out-scoping both conditionals, so that the result is a comparison of the amounts I ingest if I drink from cup A with the amounts I ingest if I drink from cup B. But one might think that perhaps the right logical form for (1) has the comparative inside the consequent of the first conditional, so that the logical form looks as follows:

(15)  If I drink from cup A [I will ingest more than if I drink from cup B].

So there are two more options to consider. The A-not-A approach gives two possible interpretations. One is:

(16)  If I drink the contents of cup A, [then there is a threshold \( \theta \) such that I will ingest an amount that meets or exceeds \( \theta \), and if I ingest from cup B, I will not ingest an amount that meets or exceeds \( \theta \)].

and the other is

(17)  If I drink the contents of cup A, [then there is a threshold \( \theta \) such that I will ingest an amount that meets or exceeds \( \theta \), and it is not the case that if I ingest from cup B, I will ingest an amount that meets or exceeds \( \theta \)].

But (16) is vacuously true when combined with Kratzer’s semantics. The vacuity is independent of the contribution of the comparative, and is the same as the vacuity of

(18)  If I drink from cup A, then if I drink from cup B I will ingest some tea.

Since we are supposing that the drinking is a single event, the first and second antecedent of (18) are inconsistent: there are no epistemically possible worlds in which I drink from both cups. The nested conditional ‘if I drink from cup B I will ingest some tea’ is evaluated against a modal base that contains only the worlds in which I drink from cup A, and so (18) comes out vacuously true. The same goes for (16): combined with Kratzer’s semantics, it is vacuously true. When combined with Kratzer’s semantics (17) is, for the same reason, vacuously false. So the alternative logical forms (16) and (17) do not help to resolve the problem.

The argument above shows that the combination of Kratzer’s semantics with Schwarzschild’s account of comparatives makes the wrong predictions. But could it be that it is the semantics of comparatives that is in need of revision? Part of the answer is contained in the Appendix, where I consider two other prominent theories of the comparative and show that they yield the same conclusions I have reached above. Furthermore, in §6 I will offer a diagnosis of the problem that suggests that it is the strict-conditionality of Kratzer’s view that is the culprit. Kratzer’s semantics gives the wrong predictions for variably dominant dominance conditionals.

4. Stalnaker’s semantics

Now, let us turn to Stalnaker-type theories. Stalnaker’s 1975 semantics

9. An anonymous reviewer suggested a possible escape route for Kratzer’s analysis: paraphrase (1) as:

If I drink the (entire) contents of cup A, I will ingest more than I would if I drank the (entire) contents of cup B.

An analysis of this ‘subjunctivised’ comparative along the lines of (16) would deliver the right truth-conditions in the cases I consider in this paper. However, the indicative version (1) is perfectly good English, and it is enough for my purposes to show that Kratzer’s theory does not handle (1) as it stands. Furthermore, it is not clear that the subjunctivised version of (1) with Kratzer’s semantics would in fact give the same truth-conditions as the indicative (1) with Stalnaker’s truth-conditions, since the subjunctivised variant is silent about the worlds in which I drink from cup B. It would be strange but perhaps possible that in all the worlds in which I drink from cup A, cup A contains more, but in some of the worlds in which I drink from cup B, cup B contains more, and that (1) should be true in such a scenario.
for indicatives relies on the similarity metric familiar from the standard Stalnaker-Lewis semantics for counterfactuals.\textsuperscript{10} The simplest formulation goes as follows:

\textit{Stalnaker’s semantics}\textsuperscript{11}

An indicative “if P then Q” is true at w iff the closest P-world to w in C is a Q-world.

(here C is the epistemic context of utterance: the set of worlds consistent with what is taken for granted, or known, in the context of utterance).

It should not be surprising that Stalnaker’s semantics predicts the truth of (1) in T1. First, recall that the A-no-A theory gives two possible logical forms for (1):

(11) There is a threshold \( \theta \) such that if I drink the contents of cup A, I will ingest an amount that meets or exceeds \( \theta \), and it is not the case that if I drink from cup B, I will ingest an amount that meets or exceeds \( \theta \).

and

(12) There is a threshold \( \theta \) such that if I drink the contents of cup A, I will ingest an amount that meets or exceeds \( \theta \), and if I drink from cup B, I will not ingest an amount that meets or exceeds \( \theta \).

Combined with Stalnaker’s semantics, the two logical forms yield:

(19) There is a threshold \( \theta \) such that the closest world w in which I drink the contents of cup A is such that, in w, I drink an amount that meets or exceeds \( \theta \), and it is not the case that the closest world \( w' \) in which I drink from cup B is such that, in \( w' \), I drink an amount that meets or exceeds \( \theta \).

and

(20) There is a threshold \( \theta \) such that the closest world w in which I drink the contents of cup A is such that, in w, I drink an amount that meets or exceeds \( \theta \), and such that the closest world \( w' \) in which I drink from cup B is such that, in \( w' \), I do not drink an amount that meets or exceeds \( \theta \).

But both (19) and (20) yield the same result: both are true in T1.\textsuperscript{12} So, suppose that we are evaluating (1) at a world w in accordance with (19). Suppose that w is a 2ml-1ml world in which I drink from cup A. Then there is a threshold, 2ml for example, such that the closest A-world (which, by hypothesis, is w itself) is such that in it I drink an amount that meets the threshold. What about the closest B-world? Here one has to appeal to the central feature of the Lewis-Stalnaker similarity metric — that it tracks (more or less) causal dependence and independence. In particular, it is plausible that the closest world to w in which I drink from cup B is a 2ml-1ml world, just like w itself. With that assumption, it follows that in the closest B-world to w I drink an amount that does not meet 2ml — our chosen threshold. So (1) is true in w on Stalnaker’s semantics. But this reasoning generalizes to all the worlds in C. So (1) is true throughout C. Stalnaker’s semantics gives the right predictions from cup B, I will not ingest an amount that meets or exceeds \( \theta \).

\textsuperscript{10} I incorporate the uniqueness assumption, the assumption that for all worlds w, there always is a unique closest P-world to w, into the semantics. Although I think that there are good reasons for accepting the uniqueness assumption, nothing in what follows hinges on it. Ties for closeness are usually motivated with examples like tossing a fair coin — the intuition being that the world in which the coin lands heads is as close as the world in which it lands tails. But T1 is obviously not symmetrical in this way: it is clear that had I drunk from the other cup, the amounts in the cups would have been the same as they actually are.

\textsuperscript{11} The view defended in Stalnaker (1975) does not include the explicit restriction on the context incorporated in semantics in the text. The main motivation for formulating things as I do is to avoid complicated questions about the presuppositions of indicatives that Stalnaker’s actual view depends on. In any case, our interest is in Stalnaker-type theories, and the variant I offer is just the simplest kind.

\textsuperscript{12} Recall that I am assuming, for simplicity, that there are no ties for closest world. If ties are allowed, (19) and (20) may diverge.
for variably dominant scenarios.\textsuperscript{13}

5. Adams’ Thesis

Adams’ Thesis is the claim that:

\textit{Adams’ Thesis}

An indicative \textit{if P, then Q} is assertible iff \( p(Q|P) \) is high.

There are two kinds of theory that explicitly subscribe to Adams’ Thesis — the material conditional accounts of Lewis (1976) and Jackson (1987), and the so-called \textit{No-Truth-Value} theories of Edgington (1995) and Bennett (2003). But in what follows, all that matters is Adams’ Thesis itself, so I will just speak of all the theories that subscribe to Adams’ Thesis, collectively, as \textit{Adams theories}. Some Adams theorists deny that indicative conditionals have truth-values, preferring to speak about their assertibility in context, and this creates severe difficulties for deriving the predictions of such an approach for dominance indicatives, because one cannot appeal to the standard compositional framework of truth-conditional semantics, as I have done above. So to derive the predictions of Adams theories one must use a more indirect approach.

What kind of constraints are there on what the Adams theorist can say about dominance indicatives? First, it seems that the assertibility of (1) can depend only on conditional probabilities \( p(I \text{ will ingest } d\text{-much} | I \text{ drink the contents of cup A/B}) \) for all \( d \). The intuition behind this assumption is simple: whatever the exact Adams semantics of the dominance conditional is, it ought to be compositional, and a dominance indicative is somehow composed of two conditional constructions and a comparative. According to Adams theorists, the contribution of the conditional is the conditional probability, hence the assumption that the assertibility of (1) can depend only on conditional probabilities \( p(I \text{ will ingest } d\text{-much} | I \text{ drink the contents of cup A/B}) \) of the current probability distribution \( p() \). Given this assumption, the argument is straightforward. Compare T1:

\textit{Tea-party 1}

Two cups of tea are on the table — cup A and cup B. We know

\textit{ingest } d\text{-much} | I \text{ drink the contents of cup A/B} \text{ for all } d \text{. Let’s call this assumption the minimal assumption of compositionality.}

Further, it is reasonable to assume that Adams dominance conditionals are \textit{extensional}, in the sense that their assertibility-value depends only on the conditional probabilities of the subject’s \textit{current} probability distribution and not on any other probability distributions.\textsuperscript{14}

Of course, even with the extensionality and compositionality assumptions accepted, we still need an answer to the question of \textit{how} the conditional probabilities just mentioned determine the assertibility of (1). In the absence of a compositional semantics, one could try to guess the assertibility-conditions that an Adams theorist might wish to assign to dominance indicatives. I do not have any good candidates to offer.\textsuperscript{15} But it can be shown that \textit{any} Adams semantics for dominance conditionals will have problems with variably dominant cases, given the extensionality and compositionality assumptions.

Together, the two assumptions amount to the assumption that the assertibility of (1) depends only on the conditional probabilities \( p(I \text{ will ingest } d\text{-much} | I \text{ drink the contents of cup A/B}) \) of the current probability distribution \( p() \). Given this assumption, the argument is straightforward. Compare T1:

\textit{Tea-party 1}

Two cups of tea are on the table — cup A and cup B. We know

\textit{ingest } d\text{-much} | I \text{ drink the contents of cup A/B} \text{ for all } d \text{. Let’s call this assumption the minimal assumption of compositionality.}

Further, it is reasonable to assume that Adams dominance conditionals are \textit{extensional}, in the sense that their assertibility-value depends only on the conditional probabilities of the subject’s \textit{current} probability distribution and not on any other probability distributions.\textsuperscript{14}

Of course, even with the extensionality and compositionality assumptions accepted, we still need an answer to the question of \textit{how} the conditional probabilities just mentioned determine the assertibility of (1). In the absence of a compositional semantics, one could try to guess the assertibility-conditions that an Adams theorist might wish to assign to dominance indicatives. I do not have any good candidates to offer.\textsuperscript{15} But it can be shown that \textit{any} Adams semantics for dominance conditionals will have problems with variably dominant cases, given the extensionality and compositionality assumptions.

Together, the two assumptions amount to the assumption that the assertibility of (1) depends only on the conditional probabilities \( p(I \text{ will ingest } d\text{-much} | I \text{ drink the contents of cup A/B}) \) of the current probability distribution \( p() \). Given this assumption, the argument is straightforward. Compare T1:

\textit{Tea-party 1}

Two cups of tea are on the table — cup A and cup B. We know

\textit{ingest } d\text{-much} | I \text{ drink the contents of cup A/B} \text{ for all } d \text{. Let’s call this assumption the minimal assumption of compositionality.}

Further, it is reasonable to assume that Adams dominance conditionals are \textit{extensional}, in the sense that their assertibility-value depends only on the conditional probabilities of the subject’s \textit{current} probability distribution and not on any other probability distributions.\textsuperscript{14}

Of course, even with the extensionality and compositionality assumptions accepted, we still need an answer to the question of \textit{how} the conditional probabilities just mentioned determine the assertibility of (1). In the absence of a compositional semantics, one could try to guess the assertibility-conditions that an Adams theorist might wish to assign to dominance indicatives. I do not have any good candidates to offer.\textsuperscript{15} But it can be shown that \textit{any} Adams semantics for dominance conditionals will have problems with variably dominant cases, given the extensionality and compositionality assumptions.

Together, the two assumptions amount to the assumption that the assertibility of (1) depends only on the conditional probabilities \( p(I \text{ will ingest } d\text{-much} | I \text{ drink the contents of cup A/B}) \) of the current probability distribution \( p() \). Given this assumption, the argument is straightforward. Compare T1:

\textit{Tea-party 1}

Two cups of tea are on the table — cup A and cup B. We know

\textit{ingest } d\text{-much} | I \text{ drink the contents of cup A/B} \text{ for all } d \text{. Let’s call this assumption the minimal assumption of compositionality.}

Further, it is reasonable to assume that Adams dominance conditionals are \textit{extensional}, in the sense that their assertibility-value depends only on the conditional probabilities of the subject’s \textit{current} probability distribution and not on any other probability distributions.\textsuperscript{14}

Of course, even with the extensionality and compositionality assumptions accepted, we still need an answer to the question of \textit{how} the conditional probabilities just mentioned determine the assertibility of (1). In the absence of a compositional semantics, one could try to guess the assertibility-conditions that an Adams theorist might wish to assign to dominance indicatives. I do not have any good candidates to offer.\textsuperscript{15} But it can be shown that \textit{any} Adams semantics for dominance conditionals will have problems with variably dominant cases, given the extensionality and compositionality assumptions.

Together, the two assumptions amount to the assumption that the assertibility of (1) depends only on the conditional probabilities \( p(I \text{ will ingest } d\text{-much} | I \text{ drink the contents of cup A/B}) \) of the current probability distribution \( p() \). Given this assumption, the argument is straightforward. Compare T1:

\textit{Tea-party 1}

Two cups of tea are on the table — cup A and cup B. We know

\textit{ingest } d\text{-much} | I \text{ drink the contents of cup A/B} \text{ for all } d \text{. Let’s call this assumption the minimal assumption of compositionality.}

Further, it is reasonable to assume that Adams dominance conditionals are \textit{extensional}, in the sense that their assertibility-value depends only on the conditional probabilities of the subject’s \textit{current} probability distribution and not on any other probability distributions.\textsuperscript{14}

Of course, even with the extensionality and compositionality assumptions accepted, we still need an answer to the question of \textit{how} the conditional probabilities just mentioned determine the assertibility of (1). In the absence of a compositional semantics, one could try to guess the assertibility-conditions that an Adams theorist might wish to assign to dominance indicatives. I do not have any good candidates to offer.\textsuperscript{15} But it can be shown that \textit{any} Adams semantics for dominance conditionals will have problems with variably dominant cases, given the extensionality and compositionality assumptions.

Together, the two assumptions amount to the assumption that the assertibility of (1) depends only on the conditional probabilities \( p(I \text{ will ingest } d\text{-much} | I \text{ drink the contents of cup A/B}) \) of the current probability distribution \( p() \). Given this assumption, the argument is straightforward. Compare T1:

\textit{Tea-party 1}

Two cups of tea are on the table — cup A and cup B. We know

\textit{ingest } d\text{-much} | I \text{ drink the contents of cup A/B} \text{ for all } d \text{. Let’s call this assumption the minimal assumption of compositionality.}

Further, it is reasonable to assume that Adams dominance conditionals are \textit{extensional}, in the sense that their assertibility-value depends only on the conditional probabilities of the subject’s \textit{current} probability distribution and not on any other probability distributions.\textsuperscript{14}

Of course, even with the extensionality and compositionality assumptions accepted, we still need an answer to the question of \textit{how} the conditional probabilities just mentioned determine the assertibility of (1). In the absence of a compositional semantics, one could try to guess the assertibility-conditions that an Adams theorist might wish to assign to dominance indicatives. I do not have any good candidates to offer.\textsuperscript{15} But it can be shown that \textit{any} Adams semantics for dominance conditionals will have problems with variably dominant cases, given the extensionality and compositionality assumptions.

Together, the two assumptions amount to the assumption that the assertibility of (1) depends only on the conditional probabilities \( p(I \text{ will ingest } d\text{-much} | I \text{ drink the contents of cup A/B}) \) of the current probability distribution \( p() \). Given this assumption, the argument is straightforward. Compare T1:

\textit{Tea-party 1}

Two cups of tea are on the table — cup A and cup B. We know
that either cup A contains 2ml and cup B 1ml of tea, or cup A contains 4ml and cup B 3ml of tea.

with

Two cups of tea are on the table — cup A and cup B. We know that either cup A contains 2ml and cup B 3ml of tea, or cup A contains 4ml and cup B 1ml of tea.

Let us stipulate further that in both T1 and T3 the relevant possibilities are equiprobable: so that

\[ p_{T1}(I \text{ drink from cup } A \text{ and cup } A \text{ has } 2\text{ml and cup } B \text{ has } 1\text{ml}) = p_{T1}(I \text{ drink from cup } B \text{ and cup } A \text{ has } 2\text{ml and cup } B \text{ has } 1\text{ml}) = p_{T1}(I \text{ drink from cup } A \text{ and cup } A \text{ has } 4\text{ml and cup } B \text{ has } 3\text{ml}) = p_{T1}(I \text{ drink from cup } B \text{ and cup } A \text{ has } 4\text{ml and cup } B \text{ has } 3\text{ml}) = .25 \]

in T1, and similarly in T3:

\[ p_{T3}(I \text{ drink from cup } A \text{ and cup } A \text{ has } 2\text{ml and cup } B \text{ has } 3\text{ml}) = p_{T3}(I \text{ drink from cup } B \text{ and cup } A \text{ has } 2\text{ml and cup } B \text{ has } 3\text{ml}) = p_{T3}(I \text{ drink from cup } A \text{ and cup } A \text{ has } 4\text{ml and cup } B \text{ has } 1\text{ml}) = p_{T3}(I \text{ drink from cup } B \text{ and cup } A \text{ has } 4\text{ml and cup } B \text{ has } 1\text{ml}) = .25 \]

T1 is variably dominant, while T3 is not. It is clear that (1) is assertible in T1 and unassertible in T3.\(^6\) But in T1 and T3 the relevant conditional probabilities are identical: \(p_{T1}(I \text{ will ingest d-much } | I \text{ drink the contents of cup } A/B) = p_{T3}(I \text{ will ingest d-much } | I \text{ drink the contents of cup } A/B)\). So if (1) is not Adams-assertible in T3, it is not Adams-assertible in T1, either.

To sum up: although we do not have a direct route to a compositional

Adams semantics for dominance conditionals, the argument appealing to T1 and T3 shows that it is bound to give the wrong predictions, no matter what it is, so long as it respects the two assumptions of minimal compositionality and extensionality made above.

6. The argument in favor of Stalnaker-type semantics

Our basic result, then, is this: Kratzer’s semantics and Adams theories give the wrong verdict on (1) in T1 (or T2 or T3), while Stalnaker predicts that it is true (and assertible) in T1, false in T3, and true or false in T2, depending on the actual amounts in A and B. Since (1) is, plausibly, true (and assertible) in T1, false in T3, and true or false depending on the amounts in A and B in T2, we have an argument in favor of Stalnaker’s semantics.

Intuitively, the reasoning in §3–5 is a dramatization of the following basic fact: both Kratzer’s semantics and Adams theories are what one might call *global*: the truth/assertibility of an indicative conditional is determined entirely by the epistemic context of utterance, and is insensitive to which world in the epistemic context is the world of evaluation. Stalnaker’s semantics, on the other hand, is what one might call *local*: it is sensitive both to the epistemic context and to the identity of the world of evaluation. Because they are global, Kratzer and Adams dominance indicatives report, in T1, on the relations between the entire set of amounts I ingest in the worlds in which I drink from cup A, and the entire set of amounts I ingest in the worlds in which I drink from cup B. But, in variably dominant scenarios like T1, no such relation — neither the relation between the minima of both sets, nor between minima and maxima — is informative enough; no such relation tracks the truth of dominance conditionals in variably dominant scenarios. The argument against Adams theories above demonstrates this particularly clearly, by appealing to a pair of scenarios (T1 and T3) which are indistinguishable from the point of view of any Adams theory of dominance indicatives. The argument against Kratzer’s view above is also very similar, although the details about the semantics of comparatives may obscure this: the argument works, essentially, by showing that the Kratzer semantics *does*
not pass on enough information to the semantics of comparatives, and, as a result, the Kratzer dominance conditionals are not sensitive enough to distinguish between the peculiar variably dominant cases, in which (1) is true, and the ordinary non-dominant ones, in which (1) is false.

Local theories, by contrast, can give the comparative semantics enough information. So, for Stalnaker, the truth of (1), evaluated in a given world w in the epistemic context, depends on a comparison between w and another world (the closest world in which I drink from the other cup). And so (1) is assertible, or known, if (1) is true in most, or in all, of the worlds in the epistemic context. Thus, while global theories track the relations between sets of degrees (=the amounts I ingest in A-worlds and B-worlds), local theories like Stalnaker’s track the sets of pairwise relations between worlds. If this diagnosis is right, it seems plausible that no future semantics of comparatives could rescue global theories like Kratzer’s account (or Adams theories).

7. The Newcomb Problem
Recall the Newcomb problem. Two boxes stand in front of John. One is transparent, and contains $1,000. The other is opaque, and contains either nothing or $1,000,000. John is offered a choice — pick the opaque box (‘pick one box’), or pick both boxes. John is to receive the money contained in the boxes he chooses. John is also told that a very reliable (say, .99) computer has already set the amount in the opaque box in the following way: if it predicted that John will pick one box, it sets the amount to be $1,000,000, and if it predicted that John will pick two boxes, it set the amount to $0. The question is: Should John pick two boxes, or should he pick one?

The debate on the Newcomb problem continues, but one aspect of it has been neglected. It has long been recognized that one’s views on the Newcomb problem are somehow connected with, and motivated by, certain conditionals. So, Lewis expressed the basic one-boxing intuition as follows: ‘[One-boxers] are convinced by indicative conditionals: if I take one box I will be a millionaire, but if I take both boxes I will not’ (Lewis [1981b], p. 377). The proponents of two-boxing, by contrast, often appeal to subjunctives, as Lewis himself did: ‘We [two-boxers] are convinced by counterfactual conditionals: If I took only one box, I would be poorer by a thousand than I will be after taking both’ (Lewis [1981b], p. 377). But, surprisingly, the nature of this connection has not been addressed head-on. How, precisely, do conditionals convince us of one or the other answer to the Newcomb problem? That’s the question I want to pursue.

There are really two questions here. One is: Which conditionals are we to appeal to, indicatives or subjunctives? The second, main, question is: How are we guided by conditionals toward one or the other answer to the Newcomb problem? Let me start with the first question.

Lewis, because he saw the contest between the one-boxer and the two-boxer as a contest between appeals to two different kinds of conditionals, thought that the ‘debate is hopelessly deadlocked’ (Lewis [1981a], p. 5). But such pessimism is unjustified. Lewis appeared to think that whether one appeals to indicatives or to subjunctives is somehow a theoretical choice that one can make. But this choice is illusory. We in fact use indicative conditionals to think about what decisions to make: we ask, What will happen if I do X?, etc. — nearly always in the indicative mood. The suggestion that we might be wrong to use indicatives in deliberation carries the serious cost of positing a theory of practical deliberation according to which we systematically misuse the conceptual resources of English. What would recommend such a theory? I suspect that the only reason for Lewis’ (and Gibbard’s [1981]) appeal to subjunctives in their discussion of the Newcomb problem is their prior commitment to an Adams theory of indicatives, and their belief that Adams indicatives would guide us to the wrong (one-boxing) answer to the Newcomb

17. DeRose (2010) makes fundamentally the same point. But DeRose also holds that some indicatives are ‘deliberationally useless’ — an issue I address below in §8. There is also an important question here that I am not going to pursue: just what is the relation between future-oriented indicatives (‘what will happen if I do A?’) and future less vivid subjunctives (‘what would happen were I to do A?’). For one view of the matter, see DeRose (2010). For my purposes it is sufficient that we rightly use indicatives in a wide variety of decision situations, including the Newcomb problem.
Theodore Korzukhin

Dominance Conditionals and the Newcomb Problem

...I have argued above that indicatives do not have the Adams semantics; I will argue below that even if indicatives did have the Adams semantics, they would not guide us to one-boxing. But, even apart from these arguments, the cost of condemning so much of our ordinary practice is just too high. We should take our ordinary practice at face value and take indicatives to be the paradigmatic conditionals of rational deliberation. I will have nothing more to say about subjunctives, and turn to the role of indicatives in decision-theoretic reasoning.

So what role do indicative conditionals play in our thinking about what to do? How are we guided by conditionals toward an answer to the Newcomb problem? I will offer an answer to this question below, an answer that is non-partisan so far as the Newcomb problem is concerned. But while my main concern is to answer the how question, it will turn out that, with the help of the argument for Stalnaker’s semantics from §2–6, my answer to the how question also yields an argument for two-boxing in the Newcomb problem (§9).

My proposal is simple. When we deliberate about whether A or B is the right course of action, what we want to know is whether we will be better off if we do A than if we do B. In simple cases, being better off amounts to getting more of something — more money, for example. In such cases, that is, in cases where V-ing more is all that matters (getting more money, or whatever), what we want to know is whether we will V more if we do A than if we do B. Of course, we are often not in a position to come to know this, and if we are not, we need to use other conceptual resources at our disposal to decide what to do. But, in the best-case scenario when we can come to know the relevant dominance indicative, I suggest that we find it of use because we implicitly recognize the truth of the following decision-theoretic norm:

**Dominance Norm (DN)**

If the indicative dominance conditional ‘if I do A, then I will V more than if I do B’ is true in every world in the epistemic context of utterance, and if V-ing more is all that matters in the context, then option A is rationally preferable to option B.\(^{18}\)

A couple of comments about DN, before we turn to its consequences.

i) DN is of course a kind of dominance principle — a decision-theoretic principle that says that in some cases one course of action is preferable to another no matter what the relevant probabilities are. In particular, since I will eventually argue that DN can be used as part of an argument for two-boxing in the Newcomb problem, it is important to compare DN with the dominance principle sanctioned by causal decision theory:\(^{19}\)

**Causal Dominance (CD)**

Let \(\{E_1, E_2, E_3, \ldots\}\) be a partition of events that the decision maker regards as causally independent of her choice between A and B. If she weakly prefers A to B given \(E_j\) for every \(E_j\), then she should weakly prefer A to B. Moreover, if any of these preferences is strict (and the associated event is nonnull), then she should strongly prefer A to B. (Joyce [1999], p. 151)

It is uncontroversial that CD leads to the two-boxing answer in the Newcomb problem. If we take \{There is $1,000,000 in the opaque box, There is $0 in the opaque box\} to be the partition, it is plausibly causally independent of one’s choice (since the amount, whatever it is, has already been deposited in the opaque box). It then follows from CD that two-boxing is the right choice. But CD can hardly be used as part of an argument in favor of two-boxing, because the two-boxing intuition in

\(^{18}\) Why introduce the mention of V-ing more into DN, rather than appealing, straightforwardly, to being better off? Because the notion of being better off is normative, and so may give rise to the suspicion that the truth of such a dominance indicative itself depends on which decision theory is right. Thus it may turn out that two parties to a decision-theoretic disagreement give different truth-values to the same better-off dominance conditional. This would prevent, or at least muddle, the use of dominance conditionals to settle some decision-theoretic disagreements. By contrast, the formulation of DN given in the text distinguishes the normative assumption, that V-ing is all that matters, from the purely factual assumption, that we shall V more if we do A than if we do B.

\(^{19}\) For more on causal decision theory, see Joyce (1999).
the Newcomb problem is itself supposed to be one of the main pieces of evidence in favor of CD.\textsuperscript{20}

But DN is importantly different from CD.\textsuperscript{21} First of all, DN makes no mention of a partition of the space of states of nature or of causal independence, as CD does. Second, DN is not a recognized norm of causal or evidential decision theory, simply in virtue of the fact that standard formulations of decision theory make no mention of English indicative conditionals. Third, and this is crucial, the attractiveness of DN does not rest on any prior decision-theoretic commitments, or on judgements of rationality in contested scenarios (a charge that can be brought against CD). DN is plausible simply because it is plausible that there can be no decision situation in which the relevant dominance indicative is known to be true, and yet the course of action it recommends is the wrong course of action. DN can be plausibly considered to be constitutive of rational decision-making. So DN should be thought neutral on the Newcomb problem.\textsuperscript{22}

ii) The natural default view is that a proposition P is known iff P is true in every world in the epistemic context of utterance. So it would be natural to formulate DN by appeal to knowledge of the relevant dominance conditional, rather than to its truth throughout C. But this is contested ground: the relation between knowledge and epistemic possibility is not clear. For example, Greco (2013) argues that knowledge is compatible with epistemic probability less than 1, and so compatible with the falsehood of the proposition known in some worlds in C. Further, it may well be that truth throughout C is insufficient for knowledge. For these reasons I have formulated DN without appeal to knowledge. I will, however, purely for the sake of readability, continue to speak as though DN appeals to knowledge of the relevant dominance conditionals, with the understanding that, if the natural default view is false, a more precise formulation should always appeal to truth throughout C instead, and make no reference to knowledge.

iii) The comparison of DN with CD brings up another important point. CD, as I have given it in Joyce’s formulation, is silent on what is the set of worlds over which the relevant partition is to be given. Clearly, that set of worlds is the set of possibilities relevant to decision-making. On the other hand, the formulation of DN appeals to C, the epistemic context of utterance. The notion of epistemic context comes from the work on the semantics of the epistemic modals, most notably the epistemic ‘possibly’, ‘probably’, and the indicative ‘if’. So it is not obvious that the possibilities in C are precisely the possibilities that are decision-theoretically relevant. Without the assumption that C contains precisely the decision-theoretically relevant possibilities, DN is clearly wrong. It is, however, quite plausible that C contains all and only possibilities relevant for decision-making: if it might be the case that P, then surely P is one of the possibilities that is relevant in decision-making, and if it is (epistemically) impossible that P, then surely the possibility that P is not decision-theoretically relevant. So the assumption that C contains precisely the decision-theoretically relevant possibilities has to be understood as part of the proposal I am putting forward.

iv) It is crucial that DN appeals to knowledge (i.e., truth throughout C), rather than mere probability of the relevant dominance conditional (i.e., probability < 1). DN manifestly fails if the dominance conditional is merely probable, because in that case the relative sizes of payoffs matter. To illustrate, imagine choosing between two lotteries. Lottery A pays $1,000,000 with probability .1, and $0 with probability .9. Lottery B pays $1 with probability .9 and $0 with probability .1. Now consider

20. Pollock (2010), p. 6, makes this same point explicitly. See ibid. for discussion of the relation between the Newcomb and the Smoking Lesion cases.
21. DN is of course also different from the unrestricted dominance principle that was employed in the original formulation of the Newcomb problem (i.e., the principle we get from CD by omitting the requirement of causal independence).
22. One further question remains: What is the relation of DN and CD on the supposition that Stalnaker’s semantics is the right semantics for indicatives? Clearly, since DN does not require causal independence, it has a wider application than CD, and does not follow from it. But does CD follow from DN, given Stalnaker’s semantics? Perhaps it does. The question is: Supposing that $E_i$ is causally independent of A and B, and supposing that the actual world is an $E_i$ world, does it follow that if A, then $E_i$, and if B, then $E_i$? Do the indicative conditional and causal independence always go hand in hand? We certainly know that they usually do, but whether it is always so is unclear.
(21) If I play lottery B I will win more than if I play lottery A.

Although one needs some further assumptions about the semantics of 'probably' to demonstrate the fact, it is plausible that (21) is probable (and assertible), but not known to be true, given the description of the scenario. But clearly one ought to play lottery A, not lottery B. So the knowledge requirement cannot be relaxed.

I offer DN as the answer to the question raised earlier: How do conditionals guide us in reasoning about the Newcomb problem? What good is DN? DN is of use because it establishes a substantive connection between semantics and decision theory. In cases in which V-ing more is all that matters, DN connects the (known) truth of a factual claim (the relevant dominance indicative) with rational preference. In the case of the Newcomb problem it is clear that the normative assumption, that money is all that matters, is not in dispute. So, any semantics of indicatives that predicts that the one-boxing, or the two-boxing, dominance indicative is known to be true commits us, via DN, to one-boxing, or to two-boxing.23

To the extent that the debate over the Newcomb problem is a debate about what course of action conditionals recommend (as Lewis suggested), DN can help us formulate that debate precisely: as a disagreement over whether the right semantics of indicative conditionals commits us, via DN, to one- or two-boxing. In §9 I will look at the connection between two-boxing and Stalnaker’s semantics, and one-boxing and Adams theories, in the light of DN. But before I turn to that, there is an objection to DN that needs to be considered.

8. Deliberationally useless indicatives

DeRose (2010) also holds, as I do, that indicative conditionals are rightly used in practical deliberation. But he thinks that there are cases when conditionals cannot be used to guide action — they are deliberationally useless (DeRose [2010], p. 20ff.). A conditional is deliberationally useless if it is known, and it recommends some course of action, but it would be irrational to heed the conditional’s advice.24 Although DeRose does not mention dominance conditionals, the threat of deliberationally useless indicatives to my account is immediate: if some indicatives, in particular, if some dominance indicatives, are deliberationally useless, then DN falls, and with it the claimed connection between semantics and decision theory, at least in the strong form in which I suggest it. So we need to examine DeRose’s case carefully. DeRose’s scenario is as follows:25

Risk It!

Sly Pete is playing a new card game called Risk It! against Gullible Gus. Largely because your henchmen have been hovering about the game and helping him to cheat, the unscrupulous Pete has already won £1,000 from Gus as they move into the final round of the game. The final round of this game is quite simple. A special deck of 101 cards, numbered 0–100, is brought out, shuffled, and one card is dealt to each of the two players. After each player gets a chance to view his own card, but not his opponent’s, the player who is leading going into the final round — in this case, Pete — gets to decide whether he wants to ‘play’ or ‘quit’. If he decides

23. An anonymous reviewer suggests that the one-boxer may not be swayed by DN. What if, for example, the dominance conditional that we should appeal to in deliberation is not the one- or the two-boxing dominance conditionals that feed into DN, but instead

(*) If I choose one box, the expected value of my prize will be greater than if I choose two boxes.

I am assuming that (*) is true just in case the expected value of choosing one box is greater than the expected value of choosing two. In this case, (*) is true, and recommends one-boxing. But the plausibility of (*) as a guide in deliberation depends on the plausibility of the relevance of expected value. The dominance conditionals that DN appeals to, by contrast, have no such dependence, and that is what makes DN plausible.

24. DeRose does not draw the line between conditionals that are known to be true and those that are merely probable, and so counts many conditionals that are merely probable as giving bad advice — for example, in the Newcomb problem. But in the context of discussing DN, the merely probable cases are irrelevant.

25. DeRose’s example is inspired by Gibbard’s Sly Pete case (Gibbard [1981]).
to ‘quit’, then he simply keeps the money he has won before this final round — in this case, £1,000. If he instead decides to ‘play’, then his winnings are either doubled or cut to nothing depending on which player holds the higher card: both players show their card, and if the leader’s (Pete’s) is the higher card, the leader’s winnings are doubled — in this case, to £2,000. But if the leader decides to play, and his card is the lower one, he walks away with nothing.

In our first version of the story, your henchman Sigmund (the signaller) has seen what card Gus is holding, has signalled to Pete that Gus’s card is 83, and has received Pete’s return sign confirming that Pete got the message, and knows that Gus is holding 83. Sigmund does not know what card Pete is holding, and so does not know which player holds the higher card, but because he knows that Pete knows what both cards are, and because he is certain that Pete is not stupid enough to ‘play’ if his card is the lower one, it is clear that Sigmund knows that, and is in a position to report to you that:

(O) If Pete plays, he will win

Such information is helpful to you, because, we may suppose, you are making derivative bets on the results of Pete’s game. (DeRose [2010], p.20–21)

But, DeRose points out,

But though Sigmund seems to know that, and seems in a position to report to you that, Pete will win if he plays, Pete cannot use this conditional that Sigmund knows in Pete’s deliberation about whether or not to play. If Pete overhears Sigmund reporting to you that ‘If Pete plays, he will win’, it would be disastrous for Pete to reason as follows: ‘Well, Sigmund seems to know that I’ll win if I play; so, I should play.’ And if Pete knows what Sigmund’s grounds are for his claim, Pete will know not to reason in that disastrous way, if he is a competent consumer of indicative conditionals. This is a case where using a straightforward FDC [future-directed conditional, i.e., an indicative like (O)] as a conditional of deliberation leads to trouble: where the conditional would constitute bad advice if used in deliberation. (DeRose [2010], p.21)

DeRose concludes that (O) is deliberationally useless. If it is known that Pete will win if he plays, then the dominance conditional

(22) If Pete plays, he will win more money than if he does not.

will be known as well. So, if matters are as DeRose describes, DN fails. However, I think DeRose’s scenario does not show what it purports to show.

Let’s grant that in the scenario Sigmund knows that if Pete plays, he will win. But is Sigmund’s knowledge consistent with the possibility that Pete nevertheless goes on to play even though his card is the lower one — by mistake, because perhaps his eyesight is poor and he misreads his own card, or intentionally, because he is suddenly overcome with remorse, and wants to make it up to poor Gus? If the answer is yes, then knowledge and truth throughout C can diverge. Call this the weak reading of Risk It! On the other hand, if the answer is no, then knowledge entails truth throughout C: Sigmund’s knowledge rules out the possibility of Pete playing and losing. Call this the strong reading of Risk It!

Now, my response to the weak reading of Risk It! is just that if some worlds in C are worlds in which Pete plays and loses, then the relevant dominance indicative

(22) If Pete plays, he will win more money than if he does not.

will not be true throughout C — it will be false at least in the worlds in which Pete plays and loses. But then DN does not apply, and Risk It! is not a counterexample to DN.

Now let’s turn to the strong interpretation of DeRose’s example. On the strong interpretation we are imagining that Sigmund’s knowledge
rules out the possibility of Pete playing and losing. But now there is a different problem. What reason do we have for thinking that, in the strong case, the indicative (O), and the dominance indicative (22) recommend the wrong course of action? Why does DeRose, for instance, think that (O) recommends the wrong course of action in this case? Unfortunately, all DeRose says is that it would be ‘disastrous’ for Pete to use (O) to guide his actions. One would like to know what ‘disastrous’ means here. ‘Disastrous’, I take it, means that if Pete follows the advice of (O) (that is: play no matter what), he will do worse than if he follows a different strategy: to play if he has a higher card. But what would make us think that the first strategy is worse than the second? Presumably, that it is (epistemically) possible that the first strategy produces a worse outcome than the second — that is, that there are cases where Pete has a lower card, and plays nevertheless (and so loses). But if Sigmund knows, as per the strong interpretation, that there are no such cases, then the two strategies are in fact equally good in the strong case: if it is not possible to play and lose, Pete might as well play.

So my response to the DeRose worry in the strong case is that Pete can indeed trust the advice given by (O). And so again there is no counterexample to DN. Now let’s come back to the connection between decision theory and semantics established by DN.

9. The semantics of indicatives and the Newcomb problem
What can be learned about the Newcomb problem given the results in semantics from §2–6, and the foregoing defense of DN? Let’s take it for granted that the two-boxers and the one-boxers are both guided by indicative conditionals. Let’s further assume that they are both guided by indicative conditionals with the help of DN. Here there are two interesting connections, one between Stalnaker’s semantics and two-boxing, the other between one-boxing and Adams theories.

Let’s start with Stalnaker. The Newcomb problem is a variably dominant scenario: the amount of money in both boxes dominates the amount of money in one (since it is epistemically impossible for the two boxes to contain an amount less than or equal to the amount of money in the opaque box). The two relevant dominance conditionals are

(2) If I choose one box, I will win more than if I choose two.
(3) If I choose two boxes, I will win more than if I choose one.

The reasoning here is perfectly parallel to the reasoning given for T1 and (1) in §5. Since the choice of one or two boxes does not causally influence the amounts of money in either box, (3), the two-boxing dominance conditional, is true, and known to be true, in the Newcomb case. So, by DN, Stalnaker’s semantics commits us to two-boxing.

What about Adams theories? Does the present discussion bear out Lewis’ claim that one-boxers are guided by indicative conditionals, provided that we grant Lewis’ unspoken assumption that these in turn get

26. One might propose that the use of (O) is disastrous if Pete would lose if he were to play. But it is not clear why Pete should be concerned with the advice being disastrous in this way. If it is epistemically impossible for him to play and lose, why should it be of concern how he would do were he to play?
27. Admittedly, it would be rather strange for Pete to overhear Sigmund’s utterance of (O), and resolve to play. But the strangeness of this course of events is all due to the unrealistic nature of the strong reading: if Pete knows (O), he knows that he will not play and lose, even before he looks at his own cards.

28. For a similar observation, compare Pollock (2010), p. 5.
the Adams semantics? Here matters are more complicated. Let’s grant the Adams theorist that the one-boxing dominant conditional

(2) If I pick one box I will win more than if I pick two boxes.

is assertible. But since we cannot assign a truth-value to (2), DN, as I have formulated above, does not apply. Furthermore, the natural variation on DN:

**Dominance Norm — Adams (DNA)**

If the indicative dominance conditional ‘if I do A, then I will V more than if I do B’ is assertible, and if V-ing more is all that matters in the context, then option A is rationally preferable to option B.

clearly does not work, for the reason given in §7. So, if lottery A pays $1,000,000 with probability .1, and $0 with probability .9, and lottery B pays $1 with probability .9 and $0 with probability .1, then

(21) If I play lottery B I will win more than if I play lottery A.

is assertible, and so DNA recommends playing lottery B. But clearly one should play lottery A. So DNA is wrong.

There do not seem to be any plausible DN-like principles that would connect with (2) to yield the recommendation of one-boxing. So either it is a mistake to think that Adams theories of indicatives favor one-boxing, or they do favor one-boxing, but with the help of some other principle besides DN. I am inclined to think that the first possibility is true, but the question needs further study.

To sum up: if DN is the way to understand how indicatives are involved in practical deliberation, then we have two results, one positive and one negative. The positive argument appeals to DN and the argument for Stalnaker’s semantics from §2-6: if DN is right, then if Stalnaker-type semantics is right, we should two-box. The negative result is that Adams indicatives do not recommend one-boxing via DN, or any similar principle.

10. Conclusion

The argument of this paper can now be put succinctly. The dominance indicative (1) is true in the key variably dominant scenario T1. But only one type of theory of indicatives can predict this: a Stalnaker-type semantics. Further, we should accept DN, the principle that tells us that, in the right circumstances, the known truth of the relevant dominance indicative recommending option A entails that option A is in fact the rational course of action. But DN in combination with a Stalnaker-type semantics entails that two-boxing is the rational choice in the Newcomb problem.

---

29. I am ignoring here the differences between Lewis’s 1976 sophisticated material implication analysis of indicatives and the more common Edgington-style view.

30. One can argue as follows: If I pick one box, I will in $1,000,000; if I pick two boxes, I will win $1000; therefore, if I pick one box, I will win more than if I pick two boxes. The premises of this argument are assertible on Adams semantics because the corresponding conditional probabilities are high. Further, the argument is valid. So, the conclusion, the one-boxing dominance indicative, is also assertible.

31. On the other hand, if we try to fix DNA by stipulating that the dominance conditional is to have assertibility 1, the resulting principle will no longer apply in the Newcomb case, since (2) plausibly does not have assertibility 1.
Appendix

The two main theories of the comparative are the degree-based and the interval-based approaches. Here I will lay out the predictions of each theory for (1) in combination with Kratzer’s semantics and, for comparison, with Stalnaker’s theory. The end result is the same as that reached above: Kratzer’s theory makes the wrong predictions either for T1 or T2, and Stalnaker’s semantics gives the right predictions, no matter which theory of the comparative is used to derive them.

Let’s start with the standard degree-based account of comparatives. On the standard account,

(23) Mary is taller than Paul.

compares two sets of heights — those which Mary meets or exceeds, and those that Paul meets or exceeds — and is true just in case some specified relation holds between these two sets of degrees. It is natural to think that the relation is that the maximal element in the first set is greater than the maximal element in the second set.

The standard semantics consists of a semantics for gradable expressions, and a semantics for the comparative morpheme ‘-er’. It can be summarized as follows (the following exposition follows Heim [2006]).

\[\text{[tall]} = \lambda d.\lambda x. x\text{'s height} \geq d\]
\[\text{[-er]} = \lambda P_{<d,t>}.\lambda Q_{<d,t>}. \text{max}(Q) > \text{max}(P)\]

So the Logical Form (LF) for (23) will look like this:

(26) [-er than wh₁ [Paul is t₁ tall]]₂ [Mary is t₂ tall]³³

which leads to the following truth-conditions:

\[\text{max}\{d: \text{if I drink the contents of cup A, I will ingest at least } d\text{-much}\} > \text{max}\{d: \text{if I drink the contents of cup B, I will ingest at least } d'\text{-much}\}\]

which is true just in case Mary’s height is greater than Paul’s, as desired.

Now let us see what the standard account predicts for dominance indicatives on Kratzer’s theory. Let us take

(1) If I drink the contents of cup A, I will ingest more than if I drink the contents of cup B.

The LF will be as follows:

(28) [-er than wh₁ [if I drink the contents of cup B, I will ingest t₁]]₂ [If I drink the contents of cup A, I will ingest t₂]

which gives the following truth-conditions:

\[\text{max}\{d: \text{if I drink the contents of cup A, I will ingest at least } d\text{-much}\} > \text{max}\{d: \text{if I drink the contents of cup B, I will ingest at least } d\text{-much}\}\]

which, if we now add Kratzer’s semantics for the indicative conditional, becomes

(30) \[\text{max}\{d: \forall w (I drink from cup A in w ⊃ I will ingest at least } d\text{-much in w)}\} > \text{max}\{d: \forall w (I drink from cup B in w ⊃ I will ingest at least } d'\text{-much in w)}\]

So, according to the standard account of comparatives in combination with Kratzer’s semantics, (1) is true just in case the maximum amount d such that ‘I drink the contents of cup A ⊃ I will ingest at least d-much’ is true in all w∈C is greater than the maximum amount d’ such that ‘I drink the contents of cup B ⊃ I will ingest at least d’-much’ is true in all w∈C. In other words, according to (30), (1) is true just in case the minimum I drink in all the worlds in which I drink from cup A is greater than the minimum I drink in all the worlds I drink from cup B. But (30), as can be easily seen, gives the wrong predictions. So consider...
the following:

Tea-party 2

Two cups of tea are on the table — cup A and cup B. We know that cup A contains at least 10ml of tea, and cup B at least 1ml of tea, and nothing else of relevance (in particular we do not know that cup A contains more tea than cup B).

Let’s evaluate (1) in T2, assuming that (30) gives the right truth-conditions. Clearly, max({d: if I drink from cup A, I will ingest at least d-much})=10ml (because for all d>10ml it is false that in all the worlds in which I drink from cup A, I ingest at least d-much) and max({d: if I drink from cup B, I will ingest at least d-much})=1ml. So the standard semantics for comparatives, coupled with Kratzer’s semantics for indicatives predicts that (1) is true in T2. But this is the wrong result.

I think it is plausible that (28) is the LF of (1) because it is plausible that (1) is equivalent to

(31) The amount I will ingest if I drink from cup A is greater than the amount I will ingest if I drink from cup B.

—and in (31) the comparative clearly out-scopes both conditionals. But note that when the -er-phrase moves, it has two landing sites available, and the other possibility is

(32) [If I drink the contents of cup A [[-er than wh1 [if I drink the contents of cup B, I will ingest t1]]2 I will ingest t2]]

Combining (32) with Kratzer’s semantics raises a further issue. When I introduced Kratzer’s view in §3, I introduced the simple version of the theory for straight indicatives — indicative conditionals without modals in the consequent. But (32) is not a straight indicative, because its consequent contains another conditional. The question is whether the consequent of (32) is to be interpreted against the old modal base C, or against an updated modal base C∩[I drink the contents of cup A]. I think the second option is preferable in general (that is the route suggested by Gillies (2009)), but let me spell out the consequences of taking both options.

On the first option, (32) combined with Kratzer’s semantics yields:

(33) ∀w (I drink from cup A in w ⊃ (max({d: I will ingest at least d-much in w}) > max({d: ∀w (I drink from cup B in w ⊃ I will ingest at least d-much in w)}))

(33) makes (1) true just in case the minimum amount I drink in the worlds in which I drink from cup A is greater than the minimum amount I drink in the worlds in which I drink from cup B. These truth-conditions, as we have seen, give the wrong verdict in the case of T2.

If the consequent of an indicative conditional is evaluated in a shifted context, C∩[I drink the contents of cup A], then (32) gives

(34) ∀w (I drink from cup A in w ⊃ (max({d: I will ingest at least d-much in w}) > max({d: ∀w (I drink from cup A in w and I drink from cup B in w ⊃ I will ingest at least d-much in w)})))

But note that (34) is trivially true, given that C does not contain any worlds in which I drink from both cups. In sum, (32) does not fare better than (28).

Both Heim (2006) and Schwarzschild and Wilkinson (2002) have suggested that the standard account of comparatives gives the wrong predictions in cases like the following:

(35) John is taller than every girl in his class.

The standard semantics predicts that (35) is true iff

(36) max({d: every girl in his class is at least d-tall}) > max({d: John is at least d-tall})

that is, iff John is taller than the shortest girl in his class, which is clearly the wrong result. The standard semantics predicts a greater-than-minimum reading where a greater-than-maximum reading is called for.
Since Kratzer’s semantics for indicatives makes them essentially universal quantifiers (over worlds in the epistemic context), the difficulty with T2 illustrated above can be seen as due to the same kind of failure as the failure of (36), and so be blamed on the degree semantics of comparatives. So we need to see how Kratzer’s view fares on other approaches to comparative semantics. To remedy the problem with (36) Heim proposed an interval-based semantics of comparatives. Let me sketch it briefly.\(^{34}\)

First, the lexical entry for the adjective:

\[(37) \quad \text{[tall]} = \lambda D_{<d,t>} \lambda x_e. x\text{'s height} \in D\]

Here, D is a set of degrees. Next, the lexical entry for the comparative morpheme ‘-er’:

\[(38) \quad \text{[-er]} = \lambda d_d. \lambda d'_d, d' > d\]

We can see how these lexical entries work if we look at the example Heim discusses in (2006):

\[(39) \quad \text{The desk is wider than every couch is long.}\]

The LF is:

\[(40) \quad \text{[wh}_1 \text{[every couch is } t_1 \text{ long][}_2 \text{[the desk is [-er than } t_2 \text{ wide][}_3\text{]}\]

which leads to the following derivation:

\[(41) \quad \begin{align*}
\text{[\lambda D_{<d,t_>, \forall x_e[couch(x) \to x\text{'s length} \in D]}(\lambda d_d. \text{ the desk's width} & \in [\lambda d'_d. d' > d])] = [\lambda D_{<d_>, \forall x[couch(x) \to x\text{'s length} \in D]}(\lambda d_d. \text{ the desk's width} >} \\
\text{d)]} = \forall x[couch(x) \to x\text{'s length} \in [\lambda d_d. \text{the desk's width} > d]] \\
= \forall x[couch(x) \to \text{the desk's width} > x\text{'s length}]
\end{align*}\]

So Heim’s account predicts, correctly, that (39) is true just in case the desk’s width is greater than every couch’s length (and similarly for (35)). Now, let’s come back to our example:

\[(1) \quad \text{If I drink the contents of cup A, I will ingest more than if I drink the contents of cup B.}\]

The LF will be as follows:

\[(42) \quad [\lambda D_{<d,t_>} \text{.if I drink from cup B, I will ingest an amount } d \in D] \quad (\lambda d_d. \text{if I drink from cup A, I will ingest an amount } \in [\lambda d'_d. d' > d])]
\]

\[= [\lambda D_{<d_>, \forall w}(\lambda d_d. \text{I drink from cup B in } w \supset \text{I will ingest an amount } d \in D \wedge (\lambda d_d. \forall w(\text{I drink from cup A in } w \supset \text{I will ingest an amount } > d \text{ in } w))]\]

In other words, Kratzer’s semantics in combination with the Heim interval semantics for comparatives predicts that (1) is true just in case the minimum amount I drink in all the worlds in which I drink from cup A is greater than the maximum amount I drink in all the worlds in which I drink from cup B (if there is a maximum — if there is not, (1) will be false). So (42) predicts that (1) will be false in T2.

But (42) gives the wrong result for T1. In T1, the denotation of \([\lambda D_{<d_>, \forall w}(\lambda d_d. \forall w(\text{I drink from cup A in } w \supset \text{I will ingest an amount } > d \text{ in } w))]\) will be \([0, 2]\) (the minimum I drink in the A-worlds is 2ml). But \([0, 2] \notin \{D: 1ml \in D \text{ and } 3ml \in D\}\). So according to (42), (1) is false in T1. That is the wrong

---

\(^{34}\) I am here presenting the simpler of the two theories that Heim discusses in 2006. The simpler account makes the wrong predictions in some of the cases that the standard degree-based semantics gets right (e.g. ‘I may eat more than I have to’; the example was given to me by Will Starr). But this will not matter for our purposes, since Heim’s more complex H-operator theory gives the same results in our cases.

\(^{35}\) Here, the -er phrase has the right type, \(<d, t>, \) but the than-phrase, which is of type \(<d, t, t>\) has to move for type reasons.
result.

To sum up: when combined with either of the two main approaches to comparatives, Kratzer’s semantics predicts that (1) is false in T1 or true in T2. This is the wrong result.

It is not surprising that Stalnaker’s semantics gives the right predictions for T1 and T2 on either theory of the comparative. Recall that the degree approach gives the following LF:

\[(\text{28}) \quad [-\text{er than } w_1 \text{ if I drink the contents of cup B, I will ingest } t_1]_2 \\
|| \\
\text{[If I drink the contents of cup A, I will ingest } t_2]_2 \]

which gives rise to the following truth-conditions:

\[(\text{29}) \quad \max\{d: \text{if I drink the contents of cup A, I will ingest at least d-much}\} > \max\{d: \text{if I drink the contents of cup B, I will ingest at least d-much}\} \]

And Heim’s interval semantics gives the following LF:

\[(\text{43}) \quad [\lambda D_{<d,l>}. \text{ if I drink from cup B, I will ingest an amount } d \in D][\lambda d_d. \text{ if I drink from cup A, I will ingest an amount } d \in [\lambda d'_d. d' > d)] \\
= [\lambda D_{<d,l>}. \text{ if I drink from cup B, I will ingest an amount } d \in D][\lambda d_d. \text{ if I drink from cup A, I will ingest an amount } > d] \]

Both (29) and (43), in combination with Stalnaker’s semantics, are true at a world w in C just in case the amount I drink in the closest world to w in which I drink from cup A is greater than the amount I drink in the closest world to w in which I drink from cup B. This gives the right verdict for T1: (1) comes out true at every world in C. And the prediction for T2 is right, too: (1) will come out true in some worlds in C, namely those in which the amount in cup A is greater than the amount in cup B, and false in others, namely those in which the amount in cup A is less than or equal to the amount in cup B. So, while true in some worlds in C, (1) will not be assertible in T2.

References


