Wishing, Decision Theory, and Two-Dimensional Content*

Kyle Blumberg

September 15, 2021

Abstract

This paper is about two requirements on wish reports whose interaction motivates a novel semantics for these ascriptions. The first requirement concerns the ambiguities that arise when determiner phrases, e.g. definite descriptions, interact with 'wish'. More specifically, several theorists have recently argued that attitude ascriptions featuring counterfactual attitude verbs license interpretations on which the determiner phrase is interpreted relative to the subject's beliefs. The second requirement involves the fact that desire reports in general require decision-theoretic notions for their analysis. The current study is motivated by the fact that no existing account captures both of these aspects of wishing. I develop a semantics for wish reports that makes available belief-relative readings but also allows decision-theoretic notions to play a role in shaping the truth conditions of these ascriptions. The general idea is that we can analyze wishing in terms of a two-dimensional notion of expected utility.

Our topic is the nature of our mental states, and the meaning of propositional attitude reports. There has recently been a considerable amount of work in two areas: (i) developing accounts of *desire ascriptions*, e.g. reports of the form ⌜S wants p⌝; and (ii) developing theories of *counterfactual attitude reports*, e.g. imagination reports of the form ⌜S imagines p⌝.[1] In this

*Pre-final draft. Final version to appear in *The Journal of Philosophy*. I am grateful to Sam Carter, Jeremy Goodman, Ben Holguín, Harvey Lederman, Kristina Liefke, Matt Mandelkern, the audience at the Colloquium on Philosophy of Logic, Language, and Information 2021 at Ruhr University Bochum, and the Virtual Language Work in Progress group for their thoughtful feedback. Special thanks to Dmitri Gallow, Simon Goldstein, John Hawthorne, and two referees for *The Journal of Philosophy* for their very helpful comments.

[1]Recent work on desire includes (Villalta, 2008; Wrenn, 2010; Crnič, 2011; Lassiter, 2011; Rubinstein, 2012; Anand & Hacquard, 2013; Graff Fara, 2013; Condoravdi & Lauer, 2016; Drucker, 2017; Grano, 2017; Phillips-Brown, 2018; Blumberg & Holguín,

paper, I aim to contribute to both of these areas by focusing on a construction that lies at their intersection, namely counterfactual *wish reports*, i.e. ascriptions of the form ⌜S wishes p⌝.

A fertile source of material for philosophical and linguistic theorizing concerns the semantics of attitude reports. In particular, there has recently been a considerable amount of work in two areas: (i) developing accounts of *desire ascriptions*, e.g. reports of the form ⌜S wants p⌝; and (ii) developing theories of *counterfactual attitude reports*, e.g. imagination reports of the form ⌜S imagines p⌝.[2] In this paper, I aim to contribute to both of these areas by focusing on a construction that lies at their intersection, namely counterfactual *wish reports*, i.e. ascriptions of the form ⌜S wishes p⌝.

More specifically, this paper is about two requirements on wish reports whose interaction motivates a novel semantics for these ascriptions. The first requirement concerns the ambiguities that arise when determiner phrases, e.g. definite descriptions, interact with 'wish'. Philosophers and linguists have long maintained that attitude reports generally give rise to the *de dicto*/*de re* ambiguity. However, several theorists have recently argued that this distinction isn't exhaustive. They show that some attitude reports exhibit a *three-way* ambiguity. More specifically, attitude ascriptions featuring counterfactual attitude verbs—e.g. 'wish', 'suppose', 'imagine', and 'dream'—allow not just for *de dicto* and *de re* construals, but also interpretations on which the determiner phrase is interpreted *relative to the subject's beliefs*. For instance, Blumberg (2018) provides the following case:

> *Burgled Bill*: Bill wakes up to find a trail of muddy footprints leading to his study. Fearing the worst, he runs to his study to check on his safe. He discovers the safe door open, and the safe emptied of its contents. His valuable collection of silverware is nowhere to be found. Given all of the evidence, Bill is quite certain that he's been burgled. As it happens, Bill wasn't robbed. His partner removed the silverware from the safe so that it could be cleaned; and the muddy footprints belonged to Bill—he made them unknowingly the night before.

2019; Jerzak, 2019; Phillips-Brown, Forthcoming; Blumberg & Hawthorne, forthcomingb,f; Blumberg, forthcoming). Recent work on counterfactual attitudes includes (Ninan, 2008; Yanovich, 2011; Maier, 2015; Ninan, 2016; Blumberg, 2018; Pearson, 2018; Liefke & Werning, 2021; Liefke, forthcoming).

[2]Recent work on desire includes (Villalta, 2008; Wrenn, 2010; Crnič, 2011; Lassiter, 2011; Rubinstein, 2012; Anand & Hacquard, 2013; Graff Fara, 2013; Condoravdi & Lauer, 2016; Drucker, 2017; Grano, 2017; Phillips-Brown, 2018; Blumberg & Holguín, 2019; Jerzak, 2019; Phillips-Brown, Forthcoming; Blumberg & Hawthorne, forthcomingb,f; Blumberg, forthcoming). Recent work on counterfactual attitudes includes (Ninan, 2008; Yanovich, 2011; Maier, 2015; Ninan, 2016; Blumberg, 2018; Pearson, 2018; Liefke & Werning, 2021; Liefke, forthcoming).

(1) Bill wishes that the person who robbed him had never robbed anyone.

(1) has a true reading in this scenario. However, it can be shown that this reading corresponds to neither the *de dicto* nor *de re* construal.[3] Instead, the relevant reading is intuitively one where the definite description 'the person who robbed Bill' is interpreted relative to Bill's beliefs. On this construal, the report can be roughly paraphrased as follows: 'Bill wishes that the person *who he thinks robbed him* had never robbed anyone'. The first requirement on an adequate theory of wish reports, then, is that it be able be able to capture belief-relative readings of these ascriptions.

The second requirement involves the fact that desire reports in general require decision-theoretic notions for their analysis. This can be illustrated by considering an example inspired by Levinson (2003):

> *Insurance Want*: Sue is deciding whether to take out house insurance. She estimates that the chances of her house burning down are $\frac{1}{1000}$. But the results would be calamitous: she'd lose her home which is valued at $\$1,000,000$. Comprehensive home insurance would cost her $100. Sue has a meeting with her insurance broker this afternoon, so she needs to decide what she wants to do.

(2) Sue wants to buy insurance.

If Sue is like most of us, (2) is true: even though she thinks it's likely that her house won't burn down, there is a small possibility that it does, and the badness of this possibility outweighs the cost of buying insurance. We can construct similar examples involving wish reports:

> *Insurance Wish*: Sue met with her broker, but they spent the whole time discussing Sue's life insurance policy. Sue forgot to bring up the issue of home insurance. The broker is going on a month-long holiday and won't be available for consultations.

(3) Sue wishes she had bought house insurance.

Again, if Sue is like most of us, (3) is true: even though it is highly likely that Sue would have wasted money had she bought home insurance, it is still reasonable for her to rue her missed opportunity, given the costs of a potential fire.

[3] See §1 for discussion.

So, an adequate semantics for wish reports needs to be able to generate belief-relative readings, in order to handle ascriptions such as (1), and it must be decision-theoretic, in order to handle reports such as (3). The current study is motivated by the fact that no existing account achieves both of these goals. On the one hand, existing decision-theoretic accounts of wanting maintain that S wants p just in case the *expected value* of p, for S, outweighs the expected value of $\neg p$. Expected value is defined in terms of conditional subjective probability (Jeffrey, 1965). However, since subjects can wish for things that they believe to be false, the expected value of objects wished true will be undefined. Thus, existing decision theoretic accounts can't even handle (3), let alone (1).

On the other hand, existing accounts of belief-relative readings aren't decision-theoretic. Consequently, these theories predict that, e.g. (3) will be false in context. Moreover, these accounts can't easily be tweaked to capture such cases. Indeed, there appears to be a considerable challenge in trying marry the insight driving theories of belief-relative counterfactual attitudes with decision theory. The key idea on this approach to counterfactual attitude reports is that the internal argument to, e.g. 'wish' is "two dimensional". This is usually spelled out by maintaining that the embedded clause denotes a set of *pairs* of worlds, or a function from worlds to sets of worlds, rather than just a set of worlds. These objects are called *two-dimensional intensions* (Ninan, 2008), or *paired propositions* (Blumberg, 2018). So, for instance, on the relevant reading of (1), the meaning of the complement can be represented as follows:

> **Meaning of the complement in** (1)**:**
> $p^*(w') = \{w'' \mid$ there exists a unique person who robbed Bill in w', and this individual never robs anyone in $w''\}$

The challenge is to come up with a notion suitably related to expected value that is defined relative to such two-dimensional entities.

The primary aim of this paper is to show that this challenge can be met. I develop a semantics for wish reports that employs paired propositions, but also allows decision-theoretic notions to play a role in shaping the truth conditions of these ascriptions. The general idea is that we can analyze wishing in terms of a notion of *expected utility*, where instead of appealing to a subject's conditional credence, we appeal to their *subjunctive*, or *counterfactual* credence. More specifically, I'll make use of the notion of revising a probability function through the process of *imaging*. Roughly, imaging a probability function by a proposition p moves probability mass from worlds in which p is false, to the most similar worlds in which p is true. This means that even if a probability function C assigns some proposition p zero probability, the image of C on p (denoted C^p) can assign p non-zero probability

(in fact, $C^p(p) = 1$ so long as p isn't a contradiction). For instance, even though Sue's subjective probability of buying insurance in the *Insurance 2* scenario is 0, the image of Sue's probability function on the proposition that she bought insurance assigns this same proposition non-zero probability (in fact, the imaged function assigns this proposition probability 1). As I show, this allows us to capture (3).

As standardly presented, the imaging operation over probability functions is only defined relative to propositions, i.e. sets of worlds. So, in order to handle reports such as (1), I extend this notion by allowing functions to be imaged relative to paired propositions. As we will see, some care is required in formulating this operation. But the overall result yields an elegant semantics for wish reports that is able to capture both the decision-theoretic aspects of desire as well as the belief-relativity of counterfactual attitudes.

The paper is structured as follows. In §1 I present the basic ideas underlying two-dimensional approaches to counterfactual attitudes, and argue that existing proposals can't capture the ways in which wishes interact with subjective probability. Then in §2 I discuss existing decision-theoretic approaches to non-counterfactual desire, e.g. wanting, and show that these don't carry over to counterfactual wishing. In §3 I develop my positive account of wish reports that makes use of a revised, two-dimensional notion of imaging. Finally, §4 concludes by discussing the relationship between wishing and desire reflection principles.

1 Belief-relative desire

In this section, I show that existing approaches to belief-relative desire ascriptions, namely the two-dimensional Hintikkian account (Maier, 2015) and the two-dimensional comparative desirability account (Blumberg, 2018), cannot handle insurance cases. I begin by presenting some background on belief-relative readings of counterfactual attitudes (§1.1). Then I present each account (§§1.2-1.3), and argue that neither approach is sufficiently sensitive to subjective probabilities (§1.4).

1.1 Two-dimensional content

To a first approximation, Φ is a counterfactual attitude so long as a subject can coherently believe $\neg p$ but still be in a Φ-state with the content p (putting Frege puzzles to one side). Thus, wishing, imagining, dreaming, and supposing are all counterfactual attitudes. For instance, 'Sue wishes

that she had bought insurance' can be true even when Sue is certain that she didn't buy insurance.[4]

Philosophers and linguists have long maintained that there are ambiguities that arise when determiner phrases, e.g. definite descriptions, interact with attitude verbs. In particular, it is generally assumed that attitude reports give rise to the *de dicto*/*de re* ambiguity. The basic observation here goes back at least to Quine (1956). Quine argued that a sentence such as (4) can mean different things. On the one hand, it can report a fairly trivial belief of Ralph's. This is the *de dicto* reading, and is given in (4a). On the other hand, it can report that Ralph could have some information that would be valuable to the authorities. This is the *de re* reading, and is given in (4b).

(4) Ralph believes that someone is a spy.
   a. Ralph believes that there are spies.
   b. Someone is such that Ralph believes they are a spy.

However, several theorists have argued that the *de dicto*/*de re* distinction isn't exhaustive.[5] They show that some attitude reports exhibit a *three-way* ambiguity. More specifically, attitude ascriptions featuring counterfactual attitude verbs allow not just for *de dicto* and *de re* construals, but also interpretations on which the determiner phrase is interpreted *relative to the subject's beliefs*.[6] This is known as the *de credito* reading (Yanovich, 2011). To illustrate, let us repeat the case from Blumberg (2018):

> *Burgled Bill*: Bill wakes up to find a trail of muddy footprints leading to his study. Fearing the worst, he runs to his study to check on his safe. He discovers the safe door open, and the safe emptied of its contents. His valuable collection of silverware is nowhere to be found. Given all of the evidence, Bill is quite certain that he's been burgled. As it happens, Bill wasn't robbed.

[4]There are uses of 'wish' that are not counterfactual, e.g. when the verb takes a non-finite clause in 'I wish to play tennis later'. It means something closer to 'hope' in examples like this. It is an interesting question how exactly these two types of wish report are related, but not one that I can take up at any length here. I'll just note that the "counterfactuality" of finite-clause embedding 'wish' seems to come from its additional layer of "fake past tense", e.g. 'I wish I had a car' means that I have a desire to possess a car *now*. Thus, a promising approach to trying to explain the counterfactual aspect of finite-clause embedding 'wish' could look to accounts of fake tense (Iatridou, 2000; Schulz, 2014; Mackay, 2019).

[5]In fact, it's been known since at least (Fodor, 1979) that this distinction isn't exhaustive. However, Fodor's readings won't be relevant in what follows, since in the cases at issue, they collapse into the *de re* construal.

[6]See, for example (Ninan, 2008, 2016; Yanovich, 2011; Maier, 2015, 2016, 2017; Blumberg, 2018; Pearson, 2018; Liefke & Werning, 2021).

His wife removed the silverware from the safe so that it could be cleaned; and the muddy footprints belonged to Bill—he made them unknowingly the night before.

(1) Bill wishes that the person who robbed him had never robbed anyone.

(1) has a true reading in this scenario.[7] However, the report cannot be read *de re* since a natural *de re* paraphrase of (1) is 'a unique person robbed Bill and Bill wishes that they had never robbed anyone'. This can't be true for the simple reason that nobody robbed Bill. As for the *de dicto* reading of the report, it is equivalent to 'Bill wishes that a unique person robbed him and never robbed anyone'. This has Bill wishing something that is obviously logically incoherent, which badly misrepresents the content of Bill's attitude. Instead, the relevant reading is one where the definite description 'the person who robbed Bill' is interpreted relative to Bill's beliefs. On this construal, the report can be roughly paraphrased as follows: 'Bill wishes that the person *who he thinks robbed him* had never robbed anyone'. This is the *de credito* reading of the ascription.

As mentioned earlier, theories that try to capture *de credito* readings of counterfactual attitude reports maintain that the objects of these attitudes are in a certain sense "two-dimensional". These objects are usually represented by building on the possible worlds approach to semantic content. On the standard possible worlds framework, propositions are represented by sets of possible worlds, or equivalently functions from worlds to truth-values. The proposition expressed by a sentence is the set of worlds in which that sentence is true, e.g. the meaning of 'Harry is a doctor' is the proposition that contains all and only worlds in which Harry is a doctor.

Theories of *de credito* readings build on this approach by maintaining that the objects of counterfactual attitudes aren't propositions, i.e. sets of worlds, but rather sets of *pairs* of worlds, or equivalently functions from worlds to sets of worlds. Following Blumberg (2018), we will call these objects *paired propositions*. The rough idea is that when a subject bears a counterfactual attitude to a paired proposition p^*, then where w' is a world compatible with the subject's beliefs, $p^*(w')$ represents the subject's attitude "relative to" their beliefs. It is worth being explicit about terminology: I will only ever use the term 'proposition' to mean a set of worlds; and I will only ever use the term 'paired proposition' to mean a set of pairs of worlds, i.e. a function

[7] If this reading is difficult to access because (1) seems to require that there actually be someone who robbed Bill, then consider 'Bill thinks that someone robbed him, and he wishes that the person who robbed him had never robbed anyone'. This sentence is acceptable, and in no way suggests that Bill was robbed; but its second conjunct raises exactly the same issues as (1).

from worlds to sets of worlds. As for notation, I will represent propositions with italicized lower-case letters, e.g. p, q, r, etc; and I will represent paired propositions with a superscripted $*$: p^*, q^*, r^*, etc. always stand for paired propositions.

Since our central concern is with developing a semantics for wish reports, in the main text I won't discuss how paired propositions can be represented syntactically at logical form. But the interested reader can consult the appendix for a fairly straightforward account. For our purposes, it will suffice to outline which kind of paired proposition is associated with each construal of a counterfactual attitude report. The idea is that *de dicto* and *de re* construals are captured by *constant functions*, i.e. functions that take any world to the same set of worlds. For instance, the paired proposition expressed by the complement on the *de dicto* reading of (1) can be represented as follows (this paired proposition is denoted by d^*):

Meaning of complement on *de dicto* reading of (1)**:**

$d^*(w') = \{w'' \mid$ there exists a unique person who robbed Bill in w'', and this individual never robs anyone in $w''\}$

Clearly, this function takes any world to the empty set. This is supposed to explain why the *de dicto* reading of (1) represents an incoherent wish of Bill's.

As for the paired proposition expressed by the complement on the *de re* reading of (1), it can be represented as follows (this paired proposition is denoted by r^*):

Meaning of complement on *de re* reading of (1)**:**

$r^*(w') = \{w'' \mid$ there exists a unique person who robbed Bill in $w_@$, and this individual never robs anyone in $w''\}$

$w_@$ represents the actual world. So, this function takes any world to the set of worlds w'' such that the person who actually robbed Bill never robs Bill in w''. Since nobody actually robbed Bill, this is supposed to explain why the *de re* reading of (1) can't be true.

Finally, the meaning of the complement on the target *de credito* reading can be expressed as follows (this paired proposition is denoted by c^*):

Meaning of complement on *de credito* reading of (1)**:**

$c^*(w') = \{w'' \mid$ there exists a unique person who robbed Bill in w', and this individual never robs anyone in $w''\}$

Note that c^* is *not* a constant function: if w_1 and w_2 differ as to who robs Bill at these worlds, then $c^*(w_1)$ and $c^*(w_2)$ will be distinct. For instance, if Joe robs Bill at w_1, then $c^*(w_1)$ is the set of worlds where Joe never robs anyone. And if Steve robs Bill at w_2, then $c^*(w_2)$ is the set of worlds where Steve never robs anyone. As we'll see below, it is precisely this variation in the output of c^* that is used to capture the *de credito* reading of (1).

Hopefully the reader now has a better sense of what it means for accounts of counterfactual attitudes to be "two-dimensional". Next I turn to some semantics for wish reports in this vein.

1.2 The Hintikkian account

In order to handle *de credito* readings of desire ascriptions, Maier (2015) builds on the notion of belief-relative imagining developed by Ninan (2008).[8] Maier's account can ultimately be understood as a two-dimensional variant of the classic Hintikka-style approach to attitude reports.

On Hintikka's (1962) semantics, attitude verbs are given a quantificational semantics involving a lexically-determined accessibility relation. For example, 'believe' denotes a relation that holds between an agent S and a proposition p just in case every world compatible with what S believes is one in which p is true, i.e. a p-world. This set of worlds compatible with everything the subject believes is the subject's *belief set*, and is denoted by $\text{Dox}_{w,S}$. Similarly, on Hintikka-style semantics for desire verbs, subjects are assigned a set of "ideal worlds" compatible with their desires.[9]

On Maier's semantics, each subject S in a world w is assigned a paired proposition $\text{Bul}^*_{w,S}$ whose domain includes at least S's belief set in w. $\text{Bul}^*_{w,S}$ is supposed to represent S's "belief-relative" desires in w: when $w' \in \text{Dox}_{w,S}$, then $w'' \in \text{Bul}^*_{w,S}(w')$ iff w'' is compatible with what S desires in w *relative to* w'.

Maier maintains that the notion of a world being compatible with what a subject desires "relative to" one of her belief worlds must ultimately be taken as a primitive. However, he also gives it the following "counterfactual" gloss (220):[10]

> The new primitive notion Bul*, describing an agent's "belief-relative buletic alternatives", requires some explanation...The

[8] Ninan's semantics has also been taken up by Yanovich (2011) and Pearson (2018).

[9] See (von Fintel, 1999) and (Crnič, 2011) for examples of this sort of approach to desire ascriptions.

[10] Maier uses contexts, rather than worlds, as the intensional parameter in his account. Since the considerations that motivate using contexts rather than worlds are orthogonal to our present concerns, where Maier talks about contexts, I've replaced this with talk of worlds.

> motivation for the extra [world] parameter is that we need our model to give us a set of buletic alternatives relative to what the agent believes. More precisely, $\text{Bul}^*_{w,S}(w')$ is the set of [worlds] compatible with what the agent S in w would desire if her belief set were the singleton $\{w'\}$...Imagine you're agent S at w. Let w' be one of your doxastic alternatives. Now imagine that w' is your only doxastic alternative, i.e., you're convinced that you inhabit context w'—free of any uncertainty. In that situation, if you consider a w'' to be compatible with your desires, then $w'' \in \text{Bul}^*_{w,S}(w')$.

Maier's semantics can be represented as follows (note that $[\![\text{p}]\!]$ denotes a paired proposition, and so $[\![\text{p}]\!](w)$ denotes a proposition):[11]

> **2D Hintikkian semantics for *wish***
> $[\![\text{S wishes p}]\!]^w$ = True iff for all w' in $\text{Dox}_{w,S}$: $\text{Bul}^*(w')_{w,S} \subseteq [\![\text{p}]\!](w')$

Let us focus on the *de credito* reading of (1) ('Bill wishes that the person who robbed him had never robbed anyone'). Recall that on this reading, the complement of the report expresses the following paired proposition:

> **Meaning of complement on *de credito* reading of** (1)**:**
> $c^*(w') = \{w'' \mid$ there exists a unique person who robbed Bill in w', and this individual never robs anyone in $w''\}$

Then (1) is true (on its *de credito* reading) at $w_@$ just in case for every one of Bill's belief worlds w', and every world w'' compatible with what Bill desires at $w_@$ relative to w', the person who robbed Bill in w' never robs anyone in w''. Using Maier's counterfactual gloss on belief-relative desire we can also express these conditions as follows: (1) is true at $w_@$ just in case for each of Bill's belief worlds w': if Bill was certain that he inhabited w', then every world w'' compatible with what he would desire would be one where the person who robbed Bill in w' never robs anyone in w''.

To make these truth-conditions a bit clearer, suppose that in the *Burgled Bill* scenario, Bill thinks that either Joe or Steve robbed him, but he isn't

[11] This is a static version of Maier's semantics. His original account is set in a dynamic framework (a variant of Discourse Representation Theory), since he tries to model presupposition projection patterns involving attitude verbs. It is also worth mentioning that Maier doesn't provide a semantics specifically for wish reports. Instead, he provides a "proto semantics" that is intended to capture the common features of desire reports in general.

sure which. Then Bill's belief set can be represented by two worlds: w_1, a world where Joe robs him, and w_2, a world where Steve robs him. In this case, (1) is true at $w_@$ iff if Bill was certain that w_1 was the actual world, then every world compatible with what he desires would be one where Joe never robs anyone; and if Bill was certain that w_2 was the actual world, then every world compatible with what he desires would be one where Steve never robs anyone.

1.3 The comparative desirability account

Unlike Maier, Blumberg doesn't build on a Hintikka-style account of the attitudes. Instead, Blumberg takes inspiration from Heim's (1992) theory of "comparative desirability" (which was in turn influenced by Stalnaker (1984)). The general idea on this approach is that desire reports are true when subjects prefer the most similar worlds in which the thing being desired holds to the most similar worlds in which it does not. By employing paired propositions, Blumberg allows the object of desire to vary with each of the subject's belief worlds that is being considered.

To make this precise, it is assumed that a subject's desires generate a preference ordering over possible worlds: for any subject S and world w: $w' >_{w,S} w''$ iff w' is more desirable to S in w than w''. $>_{w,S}$ is a strict partial order. This ordering over worlds is then "lifted" to an ordering over sets of worlds as follows: $X \subseteq W$, $Y \subseteq W : X >_{S,w} Y$ iff $w' >_{S,w} w''$ for all $w' \in X$, $w'' \in Y$.

Comparative desirability accounts also make use of a *similarity function* $\text{Sim}_w(\cdot)$. Intuitively, this function maps propositions to propositions, and takes each proposition p to the set of worlds maximally similar to w in which p is true. This function is assumed to obey the following constraints:

> **Constraints on $\text{Sim}_w(\cdot)$**
>
> *Success*: $\text{Sim}_w(p) \subseteq p$
>
> *Strong Centering*: $\text{Sim}_w(p) = \{w\}$, if $w \in p$
>
> *Uniformity*: If $\text{Sim}_w(p) \subseteq q$ and $\text{Sim}_w(q) \subseteq p$, then $\text{Sim}_w(p) = \text{Sim}_w(q)$

Blumberg's account can be expressed as follows (note again that $[\![\text{p}]\!]$ denotes a paired proposition, and so $[\![\text{p}]\!](w)$ denotes a proposition):[12]

> **2D comparative desirability semantics for *wish***
> $[\![\text{S wishes p}]\!]^w$ = True if and only if for each w' in $\text{Dox}_{w,S}$: $\text{Sim}_{w'}([\![\text{p}]\!](w')) >_{w,S} \text{Sim}_{w'}(\overline{[\![\text{p}]\!](w')})$

[12] I use overbar notation to denote the set complement operation.

Again, let us focus on the *de credito* reading of (1). (1) (on its *de credito* construal) is true at $w_@$ just in case for each of Bill's belief worlds w', a unique person robbed Bill at w', and Bill prefers the worlds most similar to w' in which the person who robbed Bill at w' never robs anyone, to the worlds most similar to w' in which the person who robbed Bill at w' robs someone. Given Strong Centering, and the fact that Bill thinks that he was robbed, these truth-conditions can be simplified: (1) is true at $w_@$ just in case for each of Bill's belief worlds w': every world w'' that is maximally similar to w' in which the person who robbed Bill at w' never robs anyone is such that Bill prefers w'' to w' at $w_@$.

To make these truth-conditions a bit clearer, suppose again that Bill's belief set is comprised of two worlds: w_1, where Joe robs Bill, and w_2, where Steve robs him. Then (1) is true at $w_@$ iff Bill prefers the most similar worlds to w_1 in which Joe never robs anyone to w_1, and Bill prefers the most similar worlds to w_2 in which Steve never robs anyone to w_2.

1.4 Probability and desire

The accounts of belief-relative desire that we considered in §§1.2-1.3 are sophisticated, and it is arguable that they are able to handle the range of cases that motivated their development. However, these semantics can't capture the ways in which desire reports are sensitive to subjective probability. To illustrate, consider (3) in the *Insurance Wish* scenario once again:

> *Insurance Wish*: Sue met with her insurance broker, but they spent the whole time discussing Sue's life insurance policy. Sue forgot to bring up the issue of home insurance. Sue estimates that the chances of her house burning down are $\frac{1}{1000}$. But the results would be calamitous: she'd lose her home which is valued at $\$1,000,000$. Comprehensive home insurance would have cost her \$100. Unfortunately, Sue's broker is going on a month-long holiday and won't be available for consultations.

(3) Sue wishes she had bought house insurance.

If Sue is like most of us, (3) is true: even though she thinks it's (overwhelmingly) likely that insurance wouldn't be needed, the enormous cost of a fire destroying her home makes it reasonable for her to rue her missed opportunity.

But it is difficult to see how (3) can be true on either of the semantics discussed above. We may assume that the meaning of the complement in

(3) is a constant function from worlds to the proposition that Sue buys insurance:[13]

Meaning of the complement in (3)**:**
$s^*(w') = \{w'' \mid \text{Sue buys insurance in } w''\}$

We can represent Sue's belief set by two worlds: w_1, a world where a fire occurs, and w_2, a world where no fire occurs. On the Hintikkian account, (3) is true just in case for each of Sue's belief worlds w': if Sue was certain that she inhabited w', then she would desire that she had bought insurance. Thus, (3) is true only if were Sue certain that she inhabited w_2, she would desire that she had bought insurance. But this isn't the case: if Sue knew that there would be no fire, then buying insurance would certainly be a waste of money, and she would *not* want to buy insurance. Of course, this construal of the truth-conditions of (3) relies on Maier's counterfactual gloss of the notion of "belief relative" desire alternatives. But I submit that on any reasonable understanding of this notion, the report will come out false.

As for the comparative desirability account, (3) is true just in case for each of Sue's belief worlds w': every world w'' that is maximally similar to w' in which Sue buys insurance at w' is such that Sue prefers w'' to w'. So, (3) is true only if Sue prefers the most similar worlds to w_2 in which she buys insurance, to w_2. But again this isn't the case: the closest worlds to w_2 in which Sue buys insurance are worlds where Sue wastes money, since no fires occur there. So, Sue does *not* prefer these worlds to w_2.[14] This argument assumes that the most similar worlds to w_2 are ones where Sue buys insurance, and nothing else relevant to her preference ordering occurs. That the similarity function is configured this way in the *Insurance 2* scenario is supported by the fact that sentences such as those in (5) are also true in context:

(5) a. Sue thinks that if she had bought insurance and there was no fire, then she wouldn't have needed the policy.
b. [Uttered by Sue:] I know I probably wouldn't have needed house insurance (but I still really wish I'd bought some).

It seems hard to explain how these examples could be acceptable without the similarity facts being as described.

[13]Pronouns in the scope of counterfactual attitude verbs do also exhibit "*de credito*" construals, e.g. 'Bill thinks a woman robbed him, and he wishes that she had never robbed anyone'. But in the case of (3), the "*de credito*" construal is essentially equivalent to the construal that we assume below.

[14]This argument is similar to that run by Levinson (2003) against Heim's comparative desirability semantics for 'want'.

To sum up, in this section we have seen that existing accounts of belief-relative wishing are not sufficiently sensitive to decision-theoretic considerations. This results in them not being able to handle examples that have the structure of insurance cases. In the next section, we'll consider the leading decision-theoretic semantics for 'want'. However, as we'll see, this account doesn't carry over to wish reports.

2 Decision-theoretic wanting

It is worth repeating the *Insurance Want* scenario:

> *Insurance Want*: Sue is deciding whether to take out house insurance. She estimates that the chances of her house burning down are $\frac{1}{1000}$. But the results would be calamitous: she'd lose her home which is valued at $\$1,000,000$. Comprehensive home insurance would cost her $100. Sue has a meeting with her insurance broker this afternoon, so she needs to decide what she wants to do.

(2) Sue wants/hopes to buy insurance.

The most popular approach to reports such as (2) is Levinson's (2003) decision-theoretic semantics, which I'll present here.[15] Levinson uses notions from *evidential decision theory* (Jeffrey, 1965). These concepts are fairly standard, but I will introduce them briefly. On this approach, states of belief are represented by *subjective probability functions*. These can be expressed as ordered pairs $\langle \mathrm{Dox}_{w,S}, \mathrm{C}_{w,S} \rangle$. As before, $\mathrm{Dox}_{w,S} \subseteq W$ represents the set of live doxastic possibilities for S in w. Levinson assumes for simplicity, as we will as well, that W is finite.[16] $\mathrm{C}_{w,S}$ is a function from $\mathcal{B}$, a Boolean algebra of subsets of W, to the unit interval. $\mathrm{C}_{w,S}$ represents S's credences over the live possibilities in w. Thus, $\mathrm{C}_{w,S}(\mathrm{Dox}_{w,S}) = 1$; and for disjoint p, $q \in \mathcal{B}$, $\mathrm{C}_{w,S}(p \cup q) = \mathrm{C}_{w,S}(p) + \mathrm{C}_{w,S}(q)$.

Evidential decision theory posits a revision operation on probability functions, namely *conditionalization*. A probability function $\mathrm{C}_{w,S}$ can be revised by *conditioning* on a proposition p consistent with $\mathrm{Dox}_{w,S}$. More explicitly, if $p \cap \mathrm{Dox}_{w,S} \neq \emptyset$, then $\mathrm{C}_{w,S}(q|p)$ is the probability function defined as

[15] Levinson was inspired by Goble's (1996) account of deontic modals which employs a notion of expected value. Levinson's semantics is also endorsed by Lassiter (2011, 2010, 2017) and Jerzak (2019).

[16] Using infinite spaces would require more sophisticated techniques, e.g. integration, but the essential points would remain unaffected.

follows: $\mathrm{C}_{w,S}(q|p) = \mathrm{C}_{w,S}(q \cap p)/\mathrm{C}_{w,S}(p)$. If $p \cap \mathrm{Dox}_{w,S} = \emptyset$, then this operation is undefined for $\mathrm{C}_{w,S}$ and p. Intuitively, $\mathrm{C}_{w,S}(w'|p)$ is S's credence (in w) that w' is the actual world, given that p is true.[17]

It is also assumed that a subject's desires generate an *evaluation function* $\mathrm{v}_{w,S}$, from W to the real numbers. Intuitively, $\mathrm{v}_{w,S}(w')$ measures how much utility S would get if w' was the actual world. A notion of *expected value* can then be defined:

Expected value:

$$EV_{w,S}(p) = \sum_{w' \in D_{w,S}} \mathrm{v}_{w,S}(w') \cdot \mathrm{C}_{w,S}(w'|p)$$

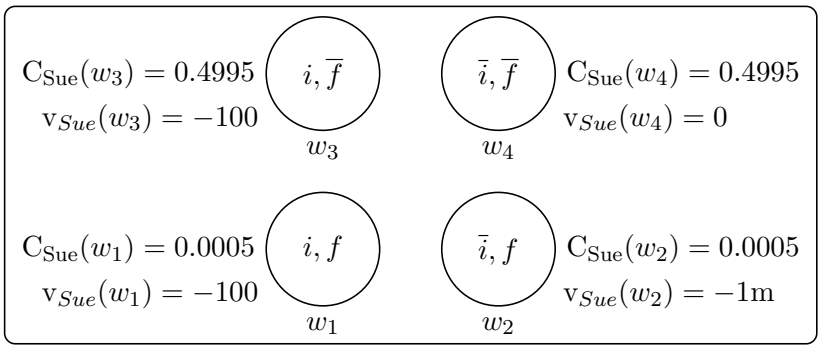
The expected value of a proposition measures the utility provided by the worlds in the proposition, weighted by the conditional probability of those worlds being actual.

Levinson's semantics then says that ⌜S wants p⌝ is true just in case S assigns a higher expected value to $[\![\mathrm{p}]\!]$ than $\overline{[\![\mathrm{p}]\!]}$ (note that $[\![\mathrm{p}]\!]$ now denotes a proposition):

Decision-theoretic semantics for *want*

$$[\![\text{S wants p}]\!]^w = \text{True iff } EV_{w,S}([\![\mathrm{p}]\!]) > EV_{w,S}(\overline{[\![\mathrm{p}]\!]})$$

To see how this account handles (2), let us represent Sue's belief state as follows (i is the proposition that Sue buys insurance, and f is the proposition that a fire occurs):[18]



[17] I'll write $\mathrm{C}(w)$ instead of $\mathrm{C}(\{w\})$.
[18] I assume that $\mathrm{C}_{\mathrm{Sue}}(i) = 0.5$.

Then $EV_{\text{Sue}}(i) = \text{v}_{\text{Sue}}(w_1) \cdot \text{C}_{\text{Sue}}(w_1|i) + \text{v}_{\text{Sue}}(w_3) \cdot \text{C}_{\text{Sue}}(w_3|i) = -100 \cdot 0.001 + -100 \cdot 0.999 = -100$. And $EV_{\text{Sue}}(\bar{i}) = \text{v}_{\text{Sue}}(w_2) \cdot \text{C}_{\text{Sue}}(w_2|\bar{i}) + \text{v}_{\text{Sue}}(w_4) \cdot \text{C}_{\text{Sue}}(w_4|\bar{i}) = -1,000,000 \cdot 0.001 + 0 \cdot 0.999 = -1,000$. Since $-100 > -1000$, Levinson's account predicts that (2) is true in *Insurance*.

By drawing on notions from evidential decision theory, Levinson's semantics is able to capture the way in which subjective probabilities impact the truth of non-counterfactual desire ascriptions. However, it should be clear that this account won't carry over to counterfactual wish reports. Consider (3) ('Sue wishes she had bought house insurance') in the *Insurance Wish* scenario once again. Since Sue knows that she didn't buy insurance, her subjective probability that she bought insurance is 0, i.e. $\text{C}_{\text{Sue}}(i) = 0$. Thus, one can't conditionalize Sue's subjective probability function on the proposition that she bought insurance. This means that the expected value of Sue buying insurance won't be defined. Consequently, (3) will either be false or undefined on a Levinson-style semantics for 'wish'. In short: this semantics fails with fairly simple wish reports, let alone more complex examples such as (1) ('Bill wishes that the person who robbed him had never robbed anyone').[19]

To sum up, in §1 we saw that existing accounts of wish reports that allow for *de credito* readings can't capture counterfactual insurance cases. This is because these theories do not permit subjective probabilities to play a role in determining what the subject desires. In the present section, we considered the leading decision-theoretic account of non-counterfactual desire reports. Although it is arguable that this account can capture want reports, this semantics doesn't carry over to wish reports, given its reliance on the notion of expected value. Intuitively, what's needed to capture wish reports is a notion similar to expected value, but does not rely on revising probability functions by conditionalization. I present such a concept in the next section, and use it to develop a decision-theoretic account of wishing.

3 A probabilistic semantics for 'wish'

In this section, I introduce the notion of revising a probability function by *imaging* on a proposition (§3.1). Using this idea, I provide a preliminary

[19] Our argument assumes (i) that Sue should assign a credence of 0 to i in the *Insurance Wish* scenario, and (ii) that if $\text{C}_{\text{Sue}}(i) = 0$, then any conditional expectation given i is undefined. Both of these assumptions are fairly controversial in the decision-theory literature; see for example (Hájek, 2003) and (Easwaran, 2019). However, lifting them doesn't substantially alter our conclusions. For instance, suppose that we allow $\text{C}_{\text{Sue}}(\cdot|i)$ to be well-defined even though $\text{C}_{\text{Sue}}(i) = 0$. Then Sue will be nearly certain that she's suffering from serious memory loss, conditional on her having purchased insurance. Since she very much doesn't want to have those kinds of neurological issues, (3) will plausibly still come out as false.

entry for 'wish' that can handle counterfactual insurance cases (§3.2). I then extend the imaging operation so that it is defined relative to paired propositions, which allows us to capture belief-relative readings of wish reports (§3.3). Finally, I discuss the presuppositions of wish reports, and present my final entry for 'wish' (§3.4).

3.1 Imaging

As discussed in §2, evidential decision theorists define expected value in terms of conditional probability. However, many argue that when it comes to providing a theory of rational choice, this is the wrong notion of value to use. More specifically, *causal* decision theorists maintain that defining expected value via conditionalization yields wrong results in so-called "Newcomb cases".[20] The basic idea in these cases is that how the agent acts provides them with information about what the world is like, even though their actions do not cause the world to be in that state. Consequently, if expected value is defined in terms of conditionalization, then in Newcomb cases, agents can be advised to perform actions that indicate that the world is a certain way, but are not maximally efficacious in bringing about desirable consequences. Causal decision theorists think that this is a problematic result.[21]

In causal decision theory, the relevant notion of value isn't defined in terms of conditional probabilities, but rather "counterfactual" or "subjunctive" probabilities (Joyce, 1999). This involves introducing a different type of revision operation for probability functions. There are several ways of specifying this revision operation, but the one that will be useful for our purposes is that of *imaging*.[22] Roughly put, imaging a probability function C by a proposition p, denoted C^p, shifts probability mass from worlds in the sample space where p is false, to the most similar worlds where p is true. To make this precise, let us begin by defining the following class of indicator functions:[23]

> **Indicator functions:**
> Given a world w, the indicator function for w, denoted $\hat{w}(\cdot)$, is the unique probability function such that for all propositions p: $\hat{w}(p) = 1$ if $w \in p$, and $\hat{w}(p) = 0$ otherwise.

[20] See §4 for an example of such a case.

[21] The literature on Newcomb problems and the evidential vs causal decision theory debate is enormous. Some classic texts include (Nozick, 1969; Gibbard & Harper, 1978; Lewis, 1981; Joyce, 1999). See, for example (Gallow, 2020) for more recent work.

[22] Imaging is discussed by (Gärdenfors, 1988; Joyce, 1999; Lewis, 1976, 1981; Sobel, 1994) among others.

[23] The exposition that follows is essentially from (Gärdenfors, 1988).

An indicator function for w can be understood as representing a maximally opinionated state of belief that w is the actual world. Given properties of indicator functions, it can be established that for any probability function C, and proposition p:

$$\mathrm{C}(p) = \sum_w \mathrm{C}(w) \cdot \hat{w}(p) \tag{DEC}$$

In other words, any probability function can be "decomposed" into a weighted sum of indicator functions.[24] The idea is that when we image a probability function C by the proposition p, we essentially replace the indicator functions $\hat{w}$ in (DEC) by the indicator functions of the worlds most similar to w. To this end, we will continue to assume that the similarity function obeys the constraints from §1.3. In addition, it will simplify our discussion if we assume that it obeys an additional constraint, namely *Uniqueness*:

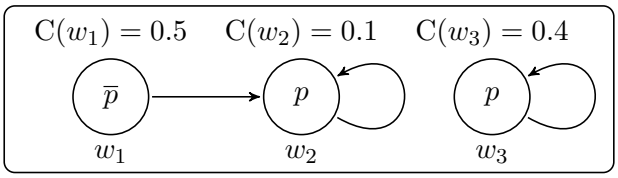
Uniqueness: $\mathrm{Sim}_w(p)$ contains at most one world

Given a world w and proposition $p \neq \emptyset$, let $\mathrm{sim}_{w,p}$ denote the unique world in $\mathrm{Sim}_w(p)$. Then imaging can be defined as follows:

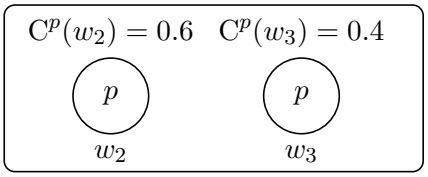
Imaging:
For any probability function C, and proposition $p \neq \emptyset$, the image of C by p, denoted $\mathrm{C}^p(\cdot)$, is the probability function defined as follows:

$$\mathrm{C}^p(q) = \sum_w \mathrm{C}(w) \cdot \widehat{\mathrm{sim}_{w,p}}(q)$$

To illustrate how imaging works, consider C and C^p displayed below:



[24] *Proof*: By the law of total probability, for any probability function C and proposition p, we have $\mathrm{C}(p) = \sum_w \mathrm{C}(w) \cdot \mathrm{C}(p|w)$ (it is assumed that $\mathrm{C}(w) > 0$, for all w). Now, for any probability function C, proposition p, and world w: either $\mathrm{C}(p \cap \{w\}) = \mathrm{C}(w)$ or $\mathrm{C}(p \cap \{w\}) = 0$. Now consider $\mathrm{C}(p|w) = \mathrm{C}(p \cap \{w\})/\mathrm{C}(w)$. If $w \in p$, then $\mathrm{C}(p|w) = \mathrm{C}(w)/\mathrm{C}(w) = 1$; and if $w \notin p$, then $\mathrm{C}(p|w) = 0$. Thus, $\mathrm{C}(\cdot|w) = \hat{w}(\cdot)$ and so $\mathrm{C}(p) = \sum_w \mathrm{C}(w) \cdot \hat{w}(p)$.



w_1 is a $\overline{p}$-world, while both w_2 and w_3 are p-worlds. The arrows in the first figure indicate the closest p-world from a given world. Thus, in imaging C by p, the probability mass assigned to w_1 gets shifted to w_2. So, for instance, $\mathrm{C}(\{w_2, w_3\}) = \mathrm{C}(w_1) \cdot \hat{w_1}(\{w_2, w_3\}) + \mathrm{C}(w_2) \cdot \hat{w_2}(\{w_2, w_3\}) + \mathrm{C}(w_3) \cdot \hat{w_3}(\{w_2, w_3\}) = 0 + 0.1 + 0.4 = 0.5$. But $\mathrm{C}^p(\{w_2, w_3\}) = \mathrm{C}(w_1) \cdot \hat{w_2}(\{w_2, w_3\}) + \mathrm{C}(w_2) \cdot \hat{w_2}(\{w_2, w_3\}) + \mathrm{C}(w_3) \cdot \hat{w_3}(\{w_2, w_3\}) = 0.5 + 0.1 + 0.4 = 1$.

Hopefully the reader now has a decent grip on the imaging operation. Let us now develop a semantics for wish reports that uses imaged probability functions.

3.2 Counterfactual insurance

Using imaged probability functions, a notion similar to expected value from §2 can be defined:

Expected utility:

$$EU_{w,S}(p) = \sum_{w'} \mathrm{v}_{w,S}(w') \cdot \mathrm{C}^p_{w,S}(w')$$

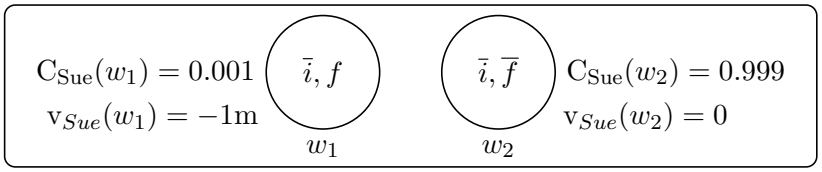
To distinguish this notion from expected value, I'll follow the literature and call it *expected utility* (Gibbard & Harper, 1978).

The basic idea is to use the above notion of expected utility, rather than expected value, in our semantics for wish reports. Let us assume for present purposes that the complements in wish reports express propositions (and not paired propositions). Then the thought is that ⌜S wishes p⌝ is true just in case the expected utility of $[\![\mathrm{p}]\!]$, for S, is greater than the expected utility of $\overline{[\![\mathrm{p}]\!]}$.

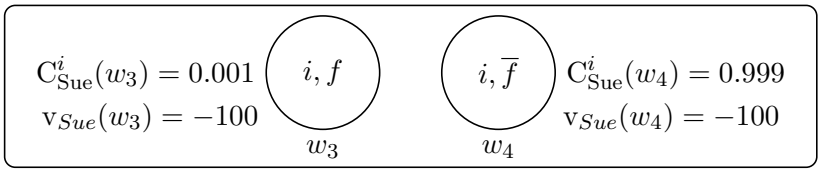
Expected Utility Semantics for *wish*

$$[\![\text{S wishes p}]\!]^w = \text{True iff } EU_{w,S}([\![\mathrm{p}]\!]) > EU_{w,S}(\overline{[\![\mathrm{p}]\!]})$$

Let us consider the truth-conditions this account generates for (3) ('Sue wishes she had bought house insurance') in the *Insurance Wish* scenario. Sue's belief state can be represented as follows:



Since Sue knows that she didn't buy insurance, imaging C_{Sue} by the proposition that she bought insurance (i) effectively just shifts the probability of each world in Dox_{Sue} to its nearest i-world. This can be represented as follows:



Then $EU_{Sue}(i) = v_{Sue}(w_3) \cdot C^i_{Sue}(w_3) + v_{Sue}(w_4) \cdot C^i_{Sue}(w_4) = -100 \cdot 0.001 + -100 \cdot 0.999 = -100$. Since Sue knows that she didn't buy insurance, and given Strong Centering, $EU_{Sue}(\bar{i}) = EU_{Sue}(\top)$, where $\top$ is the tautology. $EU_{Sue}(\top) = v_{Sue}(w_1) \cdot C^\top_{Sue}(w_1) + v_{Sue}(w_2) \cdot C^\top_{Sue}(w_2) = -1,000,000 \cdot 0.001 + 0 \cdot 0.999 = -1000$. Since $-100 > -1000$, (3) is predicted to be true, as required.

3.3 Belief-relative desire again

The semantics presented in the previous section constitutes a step in the right direction. However, as discussed in §1, our account of wish reports also needs to be able capture belief-relative readings of these ascriptions. The aim of this subsection is develop the account further so that it satisfies this desideratum. I'll first bring out an important feature of *de credito* readings that constrains the space of viable theories of wishing (§3.3.1). Then I'll present my positive proposal that extends the definition of imaging (§3.3.2).

3.3.1 An important feature of belief-relativity

Recall that we are trying to capture reports such as (1) in contexts such as the following:

> *Burgled Bill*: Bill wakes up to find a trail of muddy footprints leading to his study. Fearing the worst, he runs to his study to check on his safe. He discovers the safe door open, and the

safe emptied of its contents. His valuable collection of silverware is nowhere to be found. Given all of the evidence, Bill is quite certain that he's been burgled. As it happens, Bill wasn't robbed. His wife removed the silverware from the safe so that it could be cleaned; and the muddy footprints belonged to Bill—he made them unknowingly the night before.

(1) Bill wishes that the person who robbed him had never robbed anyone.

Also recall that existing approaches maintain that the meaning of the complement in (1) is a paired proposition. More specifically:

Meaning of complement in (1)**:**
$c^*(w') = \{w'' \mid$ there exists a unique person who robbed Bill in w', and this individual never robs anyone in $w''\}$

Since imaging has only been defined relative to sets of worlds, and not sets of pairs of worlds, i.e. paired propositions, the account presented in §3.2 can't capture (1). To make our approach to wish reports adequate, we must incorporate paired propositions. What I want to bring out here is an important constraint on such an attempt. This is most easily seen by considering the following extension of the *Burgled Bill* story:

Thieving Pair: Bill also can't find his laptop. For various reasons he now believes that he was robbed by *two* people, rather than just one. He thinks that the culprits are Joe and Steve, but he isn't sure who took what. Bill's laptop is old and is insured for much more than it's worth, so unlike the situation with the silverware, he's happy about his laptop being gone. But Bill's laptop isn't actually gone, it's just hidden under a pile of books.

(6) a. Bill wishes that the person who stole his silverware had never robbed anyone.
b. Bill wishes that the person who stole his laptop had never robbed anyone.

On its *de credito* reading, (6a) is true. By contrast, (6b) is false. This is significant because relative to Bill's beliefs, the paired propositions expressed by the complements in (6a) and (6b) output *the same range of propositions.* More precisely, the complements in (6a) and (6b) denote the following paired propositions:

Meaning of the complement in (6a)**:**

$s^*(w') = \{w'' \mid$ there exists a unique person who stole Bill's silverware in w', and this individual never robs anyone in $w''\}$

Meaning of the complement in (6b)**:**

$l^*(w') = \{w'' \mid$ there exists a unique person who stole Bill's laptop in w', and this individual never robs anyone in $w''\}$

We can suppose that Bill's belief set is comprised of two worlds: w_1, a world where Steve steals Bill's silverware and Joe steals Bill's laptop; and w_2, a world where Joe steals Bill's silverware and Steve steals Bill's laptop. Then $s^*(\{w_1, w_2\}) = \{$the proposition that Joe never robs anyone, the proposition that Steve never robs anyone$\} = l^*(\{w_1, w_2\})$. Put another way, the image (in the set-theoretic sense of "image") of Bill's belief set under s^* is identical to the image of Bill's belief set under l^*. Nevertheless, the semantic values of (6a) and (6b) diverge: the former is true but the latter is false.

What this means is that wish reports do not supervene on the image (again, in the set-theoretic sense of "image") of the subject's belief set under the relevant paired proposition. This fact places substantive constraints on our theory of wishing. It rules out straightforward attempts to incorporate paired propositions into the decision theoretic semantics from §3.2. For instance, one might have thought that paired propositions could simply be incorporated by running the expected utility check point-wise over the worlds in the subject's belief set: where $[\![\text{p}]\!]$ denotes a paired proposition, $\ulcorner$S wishes p$\urcorner$ is true iff $\forall w' \in \text{Dox} : EU([\![\text{p}]\!](w')) > EU(\overline{[\![\text{p}]\!](w')})$. But our observation above shows that this entry won't do. For example, it predicts that (6a) is true iff (6b) is, which is a bad result. In short, it's not just the range of a paired proposition which matters; the functional relationships encoded by paired propositions also play a role in shaping the truth-conditions of wish reports. Our positive account needs to reflect this.

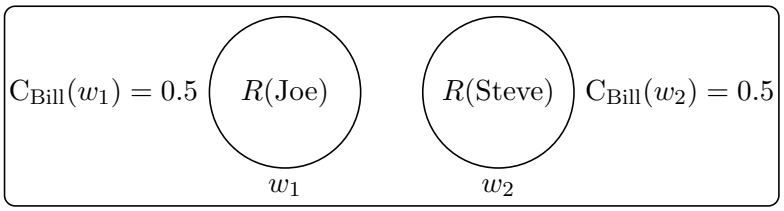
3.3.2 Imaging by paired propositions

My proposal is based on a revision operation on probability functions that is defined directly in terms of paired propositions. I call this operation *imaging** to distinguish it from imaging:

Imaging*:
For any probability function C, and paired proposition p^* such that $p^*(w) \neq \emptyset$ for all worlds w, the image of C by p^*, denoted $\text{C}^{p^*}(\cdot)$, is the probability function defined as follows:

$$\mathrm{C}^{p^*}(q) = \sum_w \mathrm{C}(w) \cdot \widehat{\mathrm{sim}_{w,p^*(w)}}(q)$$

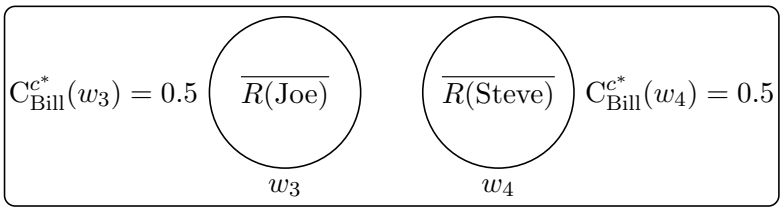
In words: imaging* a probability function by a paired proposition p^* shifts probability mass from a world w to the nearest $p^*(w)$-world to w. Imaging* is a conservative extension of imaging in the following sense: whenever p^* denotes a constant function, i.e. $p^*(w) = p^*(w')$, for all worlds w, w', then $\mathrm{C}^{p^*}(\cdot) = \mathrm{C}^{p^*(w)}(\cdot)$, for any world w. The difference between these notions comes out when p^* is not a constant function, e.g. when p^* is the meaning of the complement clause in a belief-relative wish ascription. To illustrate, suppose that Bill's credence is evenly split between two worlds: w_1, where Joe robs Bill; and w_2, where Steve robs Bill.



Consider once again the paired proposition expressed by the complement of (1) ('Bill wishes that the person who robbed him had never robbed anyone'):

> **Meaning of complement in** (1)**:**
> $c^*(w') = \{w'' \mid$ there exists a unique person who robbed Bill in w', and this individual never robs anyone in $w''\}$

Now suppose that we image* Bill's credences by c^*. This requires that for each of Bill's belief worlds w, we shift his credence from w to the closest world to w where the person who robbed Bill *at* w never robbed anyone. More specifically, we shift 0.5 of Bill's credence from w_1 to the closest world to w_1 where *Joe* never robs anyone; and we shift 0.5 of Bill's credence from w_2 to the closest world to w_2 where *Steve* never robs anyone. This can be represented as follows, where w_3 is the closest world to w_1 where Joe never robs anyone, and w_4 is the closest world to w_2 where Steve never robs anyone:



We can now define a variant of the notion of expected utility that takes paired propositions as arguments:

Expected utility*:

$$EU^*_{w,S}(p^*) = \sum_{w'} \mathrm{v}_{w,S}(w') \cdot \mathrm{C}^{p^*}_{w,S}(w')$$

Then a semantics for wish reports that incorporates paired propositions can be given (note that $[\![\mathrm{p}]\!]$ denotes a paired proposition):[25]

Semantics for *wish* (*to be revised*)

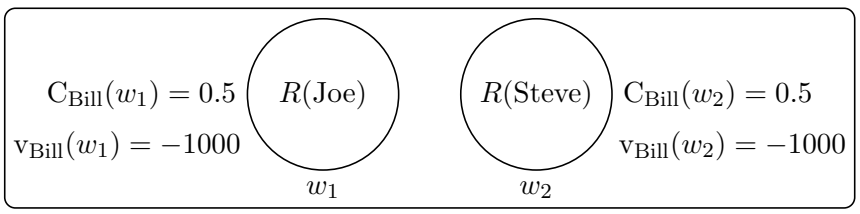
$$[\![\mathrm{S\ wishes\ p}]\!]^w = \mathrm{True\ iff}\ EU^*_{w,S}([\![\mathrm{p}]\!]) > EU_{w,S}(\top)$$

This entry captures the essence of my proposal. For one thing, it handles all of the examples handled by the account presented in §3.2, namely (3) ('Sue wishes she had bought house insurance') in the *Insurance Wish* scenario. I assume that the meaning of the complement in (3) is the following paired proposition (see the discussion in §1.4):

Meaning of the complement in (3)**:**
$s^*(w') = \{w'' \mid \text{Sue buys insurance in } w''\}$

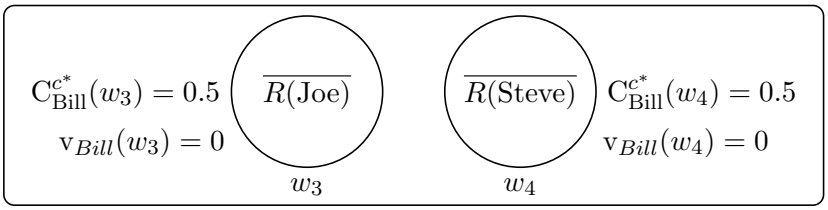
This is a constant function from any world to the set of worlds in which Sue buys insurance. Hence, as remarked above, imaging* Sue's probability function by this paired proposition is equivalent to imaging her probability function by the proposition, i.e. set of worlds, in which Sue buys insurance. Thus, $EU^*_{\mathrm{Sue}}(s^*) = EU_{\mathrm{Sue}}(i)$. So, since the account presented in §3.2 makes (3) true, the semantics developed here will as well.

As for (1), we can assume that Bill's preferences/belief state looks as follows:



[25] Officially, the condition on the right-hand side is $EU^*_{w,S}([\![\mathrm{p}]\!]) > EU^*_{w,S}(\overline{[\![\mathrm{p}]\!]})$, where $\overline{[\![\mathrm{p}]\!]}$ is the set-theoretic complement of the paired proposition $[\![\mathrm{p}]\!]$ with respect to $W \times W$. But it is plausible that for the counterfactual notion of wishing at issue, $w \notin [\![\mathrm{p}]\!](w)$, for all $w \in \mathrm{Dox}_{w,S}$ (see §3.4). That is, $w \in \overline{[\![\mathrm{p}]\!]}(w)$, for all $w \in \mathrm{Dox}_{w,S}$. Thus, given Strong Centering, $\mathrm{C}^{\overline{[\![\mathrm{p}]\!]}}_{w,S}(\cdot) = \mathrm{C}^{\top}_{w,S}(\cdot)$.

Imaging on the paired proposition c^* yields the following:



$EU_{\mathrm{Bill}}(c^*) = \mathrm{v}_{Bill}(w_3) \cdot \mathrm{C}^{c^*}_{\mathrm{Bill}}(w_3) + \mathrm{v}_{Bill}(w_4) \cdot \mathrm{C}^{c^*}_{\mathrm{Bill}}(w_4) = 0 \cdot 0.5 + 0 \cdot 0.5 = 0$. On the other hand, $EU_{\mathrm{Bill}}(\top) = \mathrm{v}_{Bill}(w_1) \cdot \mathrm{C}^{\top}_{\mathrm{Bill}}(w_1) + \mathrm{v}_{Bill}(w_2) \cdot \mathrm{C}^{\top}_{\mathrm{Bill}}(w_2) = -1000 \cdot 0.5 + -1000 \cdot 0.5 = -1000$. Since $0 > -1000$, (1) is predicted to be true in the *Burgled Bill* scenario on its *de credito* reading.

At this point, it is worth pausing to consider a concern raised by a reviewer. The reviewer suggests that the most similar world to w_1 where Joe never robs anyone will be a world where someone still robs Bill. More specifically, the most similar world to w_1 where Joe never robs anyone is w_2, the world where Steve robs Bill. Similarly, the closest to w_2 where Steve never robs anyone is w_1, the world where Joe robs Bill. In this case, the expected utility* of c^* will be equal to the expected utility of the tautology, and our semantics incorrectly predicts that (1) should be false.

In response, it is important to distinguish between two different notions of similarity. These notions link up with the distinction between indicative and subjunctive supposition. To bring this out, consider Adams's (1970) famous pair of conditionals:

(7) a. If Oswald didn't kill Kennedy, someone else did.
b. If Oswald hadn't killed Kennedy, someone else would have.

These two sentences clearly differ in meaning. (7a) will be judged true by anyone who believes that Kennedy was assassinated. By contrast, (7b) is plausibly false, since it suggests that Kennedy's assassination was inevitable. This difference in meaning can be traced to distinct types of supposition that are involved in the evaluation of each sentence, and thus distinct notions of similarity. When we evaluate (7a), we engage in *indicative* supposition which holds fixed our belief that Kennedy was assassinated. On this conception of similarity, the most similar worlds where Oswald didn't kill Kennedy are ones where someone else did. By contrast, when we assess (7b) we engage in *subjunctive* supposition which does not hold fixed our belief that Kennedy was assassinated. On this notion of similarity, the most similar worlds where Oswald didn't kill Kennedy are *not* ones where Kennedy was killed anyway.

Now, if the proposition that Joe never robs anyone is taken as an *indicative* supposition, then it is plausible that the most similar world to w_1 where Joe

never robs anyone is w_2, on the *indicative* notion of similarity. But it does not follow that the most similar world to w_1 where Joe never robs anyone is w_2, on the *subjunctive* notion of similarity. Indeed, we have good evidence that the most "subjunctively" similar world to w_1 where Joe never robs anyone is a world where Bill isn't robbed. For (8) is true in the *Burgled Bill* scenario:

(8) Bill thinks that if the person who robbed him never robbed anyone, then he would never have been robbed.

It is plausible that the notion of supposition relevant for wishing is subjunctive supposition, and thus that the relevant notion of similarity is subjunctive similarity. After all, although there is no tension between (9a) and (9b), there is a clear tension between (9a) and (9c):

(9) a. Mary wishes that Oswald hadn't killed Kennedy.
b. Mary thinks that if Oswald didn't kill Kennedy, someone else did.
c. Mary thinks that if Oswald hadn't killed Kennedy, someone else would have.

If Mary wishes that Oswald hadn't killed Kennedy, then this will be because she thinks that if Oswald hadn't killed Kennedy, Kennedy wouldn't have been assassinated. But this contradicts (9c).[26]

If this is correct, then for the purposes of evaluating (1), the most similar world to w_1 where Joe never robs anyone will be w_3, and the most similar world to w_2 where Steve never robs anyone will be w_4, just as I have suggested.

Finally, the present proposal can distinguish between (6a) ('Bill wishes that the person who stole his silverware had never robbed anyone') and (6b) ('Bill wishes that the person who stole his laptop had never robbed anyone') in the *Thieving Pair* scenario. Recall that the meaning of these clauses can be expressed as follows:

Meaning of the complement in (6a)**:**

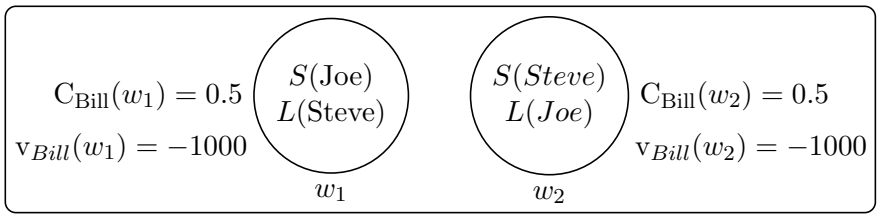
$s^*(w') = \{w'' \mid$ there exists a unique person who stole Bill's silverware in w', and this individual never robs anyone in $w''\}$

[26] Also note that if we stipulate that (8) is false in the *Burgled Bill* scenario, then (1) starts to sound false as well.

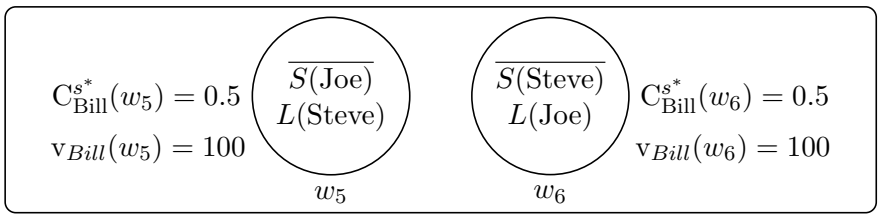
Meaning of the complement in (6b)**:**

$l^*(w') = \{w'' \mid$ there exists a unique person who stole Bill's laptop in w', and this individual never robs anyone in $w''\}$

And this is what Bill's belief state/preferences look like:



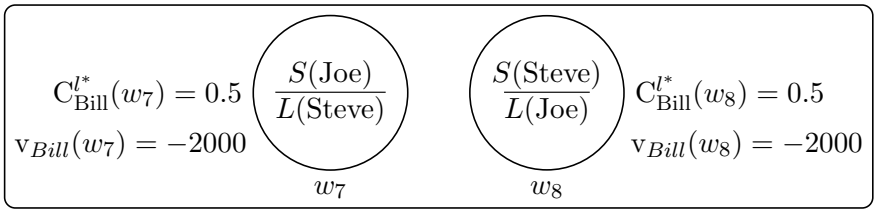
Imaging* $\mathrm{C}_{\mathrm{Bill}}$ by s^* requires that for each of Bill's belief worlds w, we move probability mass from w to the closest world to w where the person who stole Bill's silverware *at* w never robs anyone. For instance, that we move probability mass from w_1 to the closest world to w_1 in which Joe never robs anyone. Importantly, Steve will *still* steal Bill's laptop at this closest world (this is world w_5 below).[27] More generally, Bill's revised belief state/preferences will look as follows:



It is straightforward to check that the present semantics thus makes (6a) true.

By contrast, imaging* $\mathrm{C}_{\mathrm{Bill}}$ by l^* requires that we move probability mass in a different way. For instance, that we move probability mass from w_1 to the closest world to w_1 where the person who stole Bill's laptop at w_1, i.e. Steve, never robs Bill. But Joe will *still* steal Bill's silverware at this closest world (this is world w_7 below). More generally, Bill's revised belief state/preferences will look as follows:

[27]Note that if Bill came to believe that neither thief would have robbed him if one of them didn't, then it becomes much harder to detect any semantic difference between (6a) and (6b). This is exactly what our account predicts.



It is straightforward to check that the present semantics makes (6b) false.

Overall, the semantics developed above satisfies our central desideratum: it provides an entry for 'wish' that can capture both insurance cases as well as *de credito* readings of desire reports.[28] A remaining issue concerns the presuppositions of these ascriptions. I turn to this next.

3.4 The presuppositions of wish reports

Recall the *Insurance Want* scenario from earlier:

> *Insurance Want*: Sue is deciding whether to take out house insurance. She estimates that the chances of her house burning down are $\frac{1}{1000}$. But the results would be calamitous: she'd lose her home which is valued at $\$1,000,000$. Comprehensive home

[28] For simplicity, we have assumed that the similarity function obeys Uniqueness, i.e. that for any proposition p and world w, there is a unique closest p-world to w. But if we follow Gärdenfors (1982, 1988), we can develop our account in a way that doesn't require this assumption. If we drop Uniqueness, then we can define imaging* in two steps. First, we define it relative to indicator functions:

> **Imaging* for indicator functions:**
> Given an indicator function $\hat{w}$ and a paired proposition p^*, the image* of $\hat{w}$ by p^* is a probability function $\hat{w}^{p^*}(\cdot)$ such that $\hat{w}^{p^*}(p^*(w)) = 1$, and for all worlds w': $\hat{w}^{p^*}(w') > 0$ iff $w' \in \text{Sim}_w(p^*(w))$.

We then define imaging* for arbitrary probability functions:

> **Imaging***:
> For any probability function C, paired proposition p^*, and world w, the image of C by p^*, denoted $\text{C}^{p^*}(\cdot)$, is the probability function defined as follows:
>
> $$(i)\ \text{C}^{p^*}(w) = \sum_{w'} \text{C}(w') \cdot \hat{w'}^{p^*}(w)$$
> $$(ii)\ \text{C}^{p^*}(q) = \sum_{w} \text{C}^{p^*}(w) \cdot \hat{w}(q)$$

Clause (i) tells us that the probability of a world w assigned by an imaged* function is equal to the probability mass shifted to w as weighted by imaged* indicator functions. Clause (ii) follows from (DEC) (§3.1). It is easily verified that the present definition of imaging* collapses into the previous notion if we assume Uniqueness.

insurance would cost her $100. Sue has a meeting with her insurance broker this afternoon, so she needs to decide what she wants to do.

(2) Sue wants/hopes to buy insurance.

(3) # Sue wishes she had bought house insurance.

As discussed above, (2) is acceptable here, and plausibly true. By contrast, (3) is unacceptable; the report suggests that the opportunity to buy insurance has passed Sue by, which conflicts with the details of the case. The problem is that our account of wish reports doesn't predict any infelicity here. Indeed, it is fairly straightforward to check that our account predicts that (3) should be true, granted the assumptions of §2. In short, our proposal makes counterfactual desire reports too similar to non-counterfactual desire reports.

Theorists usually try to explain the contrast between (2) and (3) by saying that ⌜S wishes p⌝ presupposes that S believes $\overline{[\![\text{p}]\!]}$. This stipulation might be adequate in a setting where the complement to 'wish' denotes a proposition, i.e. set of worlds, but it doesn't carry over to the two-dimensional setting. This is because subjects don't stand in the belief relation to paired propositions. In a two-dimensional framework, the requirement of counterfactuality needs to be expressed in a more sophisticated way.

Where p^* is a paired proposition, and Δ is a set of worlds, let us say that p^* is *counterfactual over* Δ just in case $w \notin p^*(w)$, for all $w \in \Delta$. I suggest that wish reports carry the presupposition that the paired proposition expressed by the complement is counterfactual over the subject's belief set. Given that the complement in (3) expresses the paired proposition below, this explains why (3) is unacceptable in the *Insurance Want* scenario. This is because it is compatible with what Sue believes that she buys insurance. Hence, for some $w \in \text{Dox}_{\text{Sue}}$, $w \in s^*(w)$, and so s^* isn't counterfactual over Sue's belief set.

Meaning of the complement in (3)**:**
$s^*(w') = \{w'' \mid \text{Sue buys insurance in } w''\}$

The final entry for 'wish' is then the following:[29]

[29]Our account allows subjects to wish for outcomes that aren't best by their lights. For instance, suppose neither horse A nor horse B won the race. And suppose that if horse A had won, you would have received $100; and if horse B had won, you would have received $50. Then 'You wish that horse B had won the race' is predicted to be true, since horse B winning (and thus receiving $50) is better for you than the status quo (you receiving $0). Yet the report is unacceptable. More generally, it seems that subjects can only wish for what is best. See Blumberg & Hawthorne forthcominga for further discussion of this effect, and for a way to tweak decision-theoretic analyses of desire in order to capture it.

Semantics for *wish*
$[\![\text{S wishes p}]\!]^w$ is defined only if $[\![\text{p}]\!]$ is counterfactual over $\text{Dox}_{w,S}$.

If defined, $[\![\text{S wishes p}]\!]^w = \text{True}$ iff $EU^*_{w,S}([\![\text{p}]\!]) > EU_{w,S}(\top)$

4 Desire reflection and backtracking

I'll conclude by bringing out an ambiguity in wish reports, and briefly discussing what this means for some principles governing rational choice. Consider the following Newcomb case adapted from Joyce (1999, 146-147):

> *Deposit*: Suppose there is a brilliant (and very rich) psychologist who knows Bill so well that she can predict his choices with a high degree of accuracy. One Monday as Bill is on the way to the banks she stops him, holds out a thousand-dollar bill, and says: "You may take this if you like, but I must warn you that there is a catch. This past Friday I made a prediction about what your decision would be. I deposited $1,000,000 into your bank account on that day if I thought you would refuse my offer, but I deposited nothing if I thought you would accept". Bill accepts the psychologist's offer and takes the thousand-dollar bill. He then checks his bank account, and, as expected, finds that there has been no $1, 000, 000 deposit.

Consider (10):

(10) Bill wishes that he had refused the thousand-dollar bill.

Does (10) express a reasonable desire of Bill's? On the one hand, one could argue 'no', since Bill's refusing the money made no difference to whether a deposit had been made. It only would have resulted in Bill being $1000 poorer. This lines up with the acceptability of the following conditional:

(11) If Bill had refused the thousand-dollar bill, then there would still have been no deposit into his account, and he'd have been $1000 poorer.

On the other hand, one could argue 'yes', since Bill's refusing the money means that the brilliant psychologist would have predicted this, and deposited $1, 000, 000 into Bill's account. This sort of reasoning is captured by the conditional (12):

(12) If Bill had refused the thousand-dollar bill, then this would have been predicted by the psychologist, and she would have deposited the money into Bill's account.

So, there are (at least) two ways of taking wish reports. Our account can make sense of this ambiguity. In contexts where (10) expresses a reasonable desire of Bill's, the similarity relation relevant for calculating expected utility patterns with the "forward-looking" conditional (11), so that the most similar worlds where Bill refuses the money are ones where there is still no deposit. And in contexts where (10) does not express a reasonable desire of Bill's, the similarity relation relevant for calculating expected utility patterns with the "backtracking" conditional (12), so that the most similar worlds where Bill refuses the money are ones where a deposit has been made.

The availability of a backtracking interpretation of (10) marks something of a divergence from causal decision theory, since on this approach to rational choice, the relevant similarity relation is *always* spelled out in terms of a forward-looking, interventionist notion. According to causal decision theory, there is no context where it is rational for Bill to refuse the thousand-dollar bill if this makes no causal difference to whether the money is deposited in the bank. Thus, although the theory of wishing and causal decision theory both appeal to a notion of similarity, they differ in that the former's conception is less constrained than the latter's.

This difference has significance for how we should understand so-called "desire reflection principles" (Arntzenius, 2008). These principles demand that the expectation value of future desirability must equal current desirability. Put another way, one should not be such that one can foresee that one's future desires will differ from one's current ones in such a way that one will later regret earlier decisions. It has been argued that causal decision theory satisfies desire reflection.[30] But what our observations above suggest is that this isn't necessarily the case when desire is spelled out in terms of wishing: there are contexts where Bill can reasonably wish that he had refused to take the thousand-dollar bill, e.g. contexts where (12) is true, even though causal decision theory dictates that Bill should always take the offer. This indicates that causal decision theorists need to take some care care in how they formulate these principles. More pointedly: just as causal decision theorists maintain that backtracking counterfactuals fail to be probative with respect to what a rational agent ought to choose in a given situation, they should also set to one side backtracking desires when interpreting desire reflection principles.

[30]See, e.g. (Arntzenius, 2008). Arntzenius ultimately rejects causal decision theory, but for reasons unrelated to desire reflection.

Appendix

Here I outline one approach to representing paired propositions at logical form, namely Blumberg's (2018) account. Blumberg builds on the framework that posits syntactically realized *world pronouns*. World pronouns were introduced as an alternative to scopal accounts of the *de dicto*/*de re* distinction. They were designed to get the relevant readings without movement, and solve other problems as well.[31] Each predicate at LF is assigned an index—w_i, where i is a natural number—that indicates the world relative to which the predicate is to be evaluated. For instance, $teacher_{w_7}$ indicates that the predicate 'teacher' should be evaluated at world w_7. Thus, when we evaluate $teacher_{w_7}$ we will get the set of teachers at w_7.

To get a feel for how the system works, here is how the *de dicto* reading ((13b)) and the *de re* reading ((13a)) of (13) would be represented:

(13) Bill believes the person who robbed him dances.

a. λw_1 Bill believes$_{w_1}$ [λw_2 [the person-who-robbed-Bill$_{w_1}$ dance$_{w_2}$]]

b. λw_1 Bill believes$_{w_1}$ [λw_2 [the person-who-robbed-Bill$_{w_2}$ dance$_{w_2}$]]

In this system, pronoun binders appear at the top of sentences, and a sentence, ϕ, is true at a world, w, just in case $[\![\phi]\!](w) = 1$. Let us assume, as is standard, that at each world w, a subject S is assigned a *belief set*, denoted $\text{Dox}_{w,S}$, which contains all of the worlds compatible with everything that S believes at w. When (13a) is evaluated at the actual world, $w_@$, only 'believes' is evaluated at $w_@$ (giving us Bill's belief set at $w_@$)—'person who robbed Bill' (and 'dance') is evaluated relative to Bill's belief worlds (i.e. those worlds in Bill's belief set at $w_@$). This gives us the *de dicto* reading. By contrast, when (13b) is evaluated at $w_@$, *both* 'believes' and 'person who robbed Bill' are evaluated there. The extension of 'person who robbed Bil' at the actual world is the set of individuals that actually robbed Bill, giving us the *de re* reading.

Blumberg extends the world pronoun approach by allowing *two* pronoun binders to appear at the top of embedded clauses rather than just one. This allows these clauses to represent paired propositions. For instance, on the *de credito* reading of (1) ('Bill wishes that the person who robbed him had never robbed anyone'), the LF of the complement looks as follows:

(14) $\lambda w_1 \lambda w_2$ [the person-who-robbed-Bill$_{w_1}$ never-rob-anyone$_{w_2}$]

[31] See (von Fintel & Heim, 2011, 102-110) for an introduction to the world pronouns approach, and (Keshet, 2008) for a more detailed discussion.

The *de dicto*, *de re*, and *de credito* readings of (1) ('Bill wishes that the person who robbed him had never robbed anyone') are then represented as follows, respectively:

(15) a. λw_1 Bill wishes$_{w_1}$ [λw_2 λw_3 [the person-who-robbed-Bill$_{w_3}$ never-rob-anyone$_{w_3}$]

b. λw_1 Bill wishes$_{w_1}$ [λw_2 λw_3 [the person-who-robbed-Bill$_{w_1}$ never-rob-anyone$_{w_3}$]

c. λw_1 Bill wishes$_{w_1}$ [λw_2 λw_3 [the person-who-robbed-Bill$_{w_2}$ never-rob-anyone$_{w_3}$]

Constraints on world pronoun binding (Percus, 2000; Keshet, 2008) prevent the system from overgenerating readings. For instance, logical forms such as (16) are ruled out by Percus's "Generalization X":

(16) λw_1 Bill wishes$_{w_1}$ [λw_2 λw_3 [the person-who-robbed-Bill$_{w_3}$ never-rob-anyone$_{w_2}$]

> **Generalization X** (Percus, 2000):
> The world variable that the main verb selects for must be coindexed with the nearest λ above it.

Clearly, this blocks (16), since the main verb in the complement is 'never-rob-Bill', and this verb isn't coindexed with the nearest λ above it.

References

Adams, Ernest W. 1970. Subjunctive and Indicative Conditionals. *Foundations of Language*, **6**(1), 89–94.

Anand, Pranav, & Hacquard, Valentine. 2013. Epistemics and attitudes. *Semantics and Pragmatics*, **6**(8), 1–59.

Arntzenius, Frank. 2008. No Regrets, Or: Edith Piaf Revamps Decision Theory. *Erkenntnis*, **68**(2), 277–297.

Blumberg, Kyle. 2018. Counterfactual Attitudes and the Relational Analysis. *Mind*, **127**(506), 521–546.

Blumberg, Kyle. forthcoming. A Problem For The Ideal Worlds Account of Desire. *Analysis*.

Blumberg, Kyle, & Hawthorne, John. forthcominga. A New Hope. *Journal of Philosophy*.

Blumberg, Kyle, & Hawthorne, John. forthcomingb. Wanting What's Not Best. *Philosophical Studies*.

Blumberg, Kyle, & Holguín, Ben. 2019. Embedded Attitudes. *Journal of Semantics*, **36**(3), 377–406.

Condoravdi, Cleo, & Lauer, Sven. 2016. Anankastic conditionals are just conditionals. *Semantics and Pragmatics*, **9**(8), 1–69.

Crnič, Luka. 2011. *Getting even*. Ph.D. thesis, MIT.

Drucker, Daniel. 2017. Policy Externalism. *Philosophy and Phenomenological Research*, **94**(3), 1–25.

Easwaran, Kenny. 2019. Conditional Probabilities. *Pages 131–198 of:* Pettigrew, Richard, & Weisberg, Jonathan (eds), *The Open Handbook of Formal Epistemology.* PhilPapers Foundation.

Fodor, Janet Dean. 1979. *The Linguistic Description of Opaque Contexts.* Garland Pub.

Gallow, J. Dmitri. 2020. The Causal Decision Theorist's Guide to Managing the News. *Journal of Philosophy*, **117**(3), 117–149.

Gärdenfors, Peter. 1982. Imaging and Conditionalization. *Journal of Philosophy*, **79**(12), 747–760.

Gärdenfors, Peter. 1988. *Knowledge in flux: Modeling the dynamics of epistemic states.* Cambridge, MA, US: The MIT Press.

Gibbard, Allan, & Harper, William. 1978. Counterfactuals and Two Kinds of Expected Utility. *Pages 125–162 of:* Hooker, A., Leach, J. J., & McClennen, E. F. (eds), *Foundations and Applications of Decision Theory.* D. Reidel.

Goble, Lou. 1996. Utilitarian Deontic Logic. *Philosophical Studies*, **82**(3), 317–357.

Graff Fara, Delia. 2013. Specifying Desires. *Noûs*, **47**(2), 250–272.

Grano, Thomas. 2017. The Logic of Intention Reports. *Journal of Semantics*, **34**(4), 587–632.

Hájek, Alan. 2003. What Conditional Probability Could Not Be. *Synthese*, **137**(3), 273–323.

Heim, Irene. 1992. Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics*, **9**(3), 183–221.

Hintikka, Jaakko. 1962. *Knowledge and Belief.* Ithaca: Cornell University Press.

Iatridou, Sabine. 2000. The Grammatical Ingredients of Counterfactuality. *Linguistic Inquiry*, **31**(2), 231–270.

Jeffrey, Richard C. 1965. *The Logic of Decision.* University of Chicago Press.

Jerzak, Ethan. 2019. Two Ways to Want? *Journal of Philosophy*, **116**(2), 65–98.

Joyce, James M. 1999. *The Foundations of Causal Decision Theory.* Cambridge University Press.

Keshet, Ezra. 2008. *Good Intensions: Paving Two Roads to a Theory of the De re/De dicto Distinction.* Ph.D. thesis, Massachusetts Institute of Technology.

Lassiter, Daniel. 2010. Gradable epistemic modals, probability, and scale structure. *In:* Li, Nan, & Lutz, David (eds), *Semantics and Linguistic Theory (SALT) 20.* CLC Publications.

Lassiter, Daniel. 2011. *Measurement and Modality: The Scalar Basis of Modal Semantics.* Ph.D. thesis, New York University.

Lassiter, Daniel. 2017. *Graded Modality: Qualitative and Quantitative Perspectives.* Oxford: Oxford University Press.

Levinson, Dmitry. 2003. Probabilistic Model-theoretic Semantics for 'want'. *Semantics and Linguistic Theory*, **13**(0), 222–239.

Lewis, David. 1976. Probabilities of Conditionals and Conditional Probabilities. *Philosophical Review*, **85**(3), 297–315.

Lewis, David. 1981. Causal Decision Theory. *Australasian Journal of Philosophy*, **59**(1), 5–30.

Liefke, K. forthcoming. Modelling selectional super-flexibility. *Proceedings of SALT XXXI.*

Liefke, Kristina, & Werning, Markus. 2021. *New Frontiers in Artificial Intelligence.* Springer. Chap. Experiential Imagination and the Inside/Outside-Distinction.

Mackay, John. 2019. Modal interpretation of tense in subjunctive conditionals. *Semantics and Pragmatics.*

Maier, Emar. 2015. Parasitic Attitudes. *Linguistics and Philosophy*, **38**(3), 205–236.

Maier, Emar. 2016. Referential Dependencies Between Conflicting Attitudes. *Journal of Philosophical Logic*, 1–27.

Maier, Emar. 2017. Fictional Names in Psychologistic Semantics. *Theoretical Linguistics*, **43**(1-2), 1–46.

Ninan, Dilip. 2008. *Imagination, Content, and the Self.* Ph.D. thesis, Massachusetts Institute of Technology.

Ninan, Dilip. 2016. Imagination and the Self. *In:* Kind, Amy (ed), *Routledge Handbook of the Philosophy of Imagination*. Routledge: London.

Nozick, Robert. 1969. Newcomb's Problem and Two Principles of Choice. *Pages 114–146 of:* Rescher, Nicholas (ed), *Essays in Honor of Carl G. Hempel*. Reidel.

Pearson, Hazel. 2018. Counterfactual de se. *Semantics and Pragmatics*.

Percus, Orin. 2000. Constraints on Some Other Variables in Syntax. *Natural Language Semantics*, **8**(3), 173–229.

Phillips-Brown, Milo. 2018. I want to, but... *Proceedings of Sinn und Bedeutung 21 preprints*.

Phillips-Brown, Milo. Forthcoming. What does decision theory have to do with wanting? *Mind*.

Quine, W. V. 1956. Quantifiers and Propositional Attitudes. *Journal of Philosophy*, **53**(5), 177–187.

Rubinstein, Aynat. 2012. *Roots of Modality*. Ph.D. thesis, University of Massachusetts Amherst.

Schulz, K. 2014. Fake Tense in conditional sentences: a modal approach. *Natural Language Semantics*, **22**(2), 117–144.

Sobel, Jordan Howard. 1994. *Taking Chances: Essays on Rational Choice*. Cambridge University Press.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge University Press.

Villalta, Elisabeth. 2008. Mood and Gradability: An Investigation of the Subjunctive Mood in Spanish. *Linguistics and Philosophy*, **31**(4), 467–522.

von Fintel, Kai. 1999. NPI Licensing, Strawson Entailment, and Context Dependency. *Journal of Semantics*, **16**(2), 97–148.

von Fintel, Kai, & Heim, Irene. 2011. *Intensional Semantics*. MIT unpublished class notes.

Wrenn, Chase. 2010. A Puzzle About Desire. *Erkenntnis*, **73**(2), 185–209.

Yanovich, Igor. 2011. The problem of counterfactual de re attitudes. *Semantics and Linguistic Theory*, **21**, 56–75.