Forthcoming in *Disputatio, Special Issue: Selected Proceedings from the LanCog Summer Metaphysics Workshop 2024* (Editor: Ned Markosian)

Future-Directed Counterfactuals, Practical Reasoning, and the Open Future

Stephan Torre

Abstract

One stark difference between the past and the future lies in our ability to shape the future in a way in which we are unable to shape the past. This paper investigates what kind of beliefs about the future serve as premises in our reasoning about how to act. If we think about belief in terms of agents representing the world, we cannot lose sight of the fact that agents are part of, and shape, the same world they represent. Beliefs about the future appear to have a circularity about them: on the one hand, they serve as premises for deciding what we will do. On the other hand, what we decide we will do determines what we believe about the future. I argue that beliefs in future-directed counterfactuals play a central role in our practical reasoning and in how we conceptualize the actual future. I defend a robust distinction between future-directed counterfactuals and future-directed indicatives, and, *contra* Keith DeRose (2010), argue that it is future-directed counterfactuals that we use in deliberation about how to act. I argue that we construe the actual future in hypothetical terms and dependent on what we do now, in contrast to how we construe the actual past. This asymmetry in our belief content about the past versus the future fits well with an account of the open future in terms of counterfactual dependence.

1 A Circularity about Future-Directed Beliefs

Consider the following case:

TREE You are hiking in the mountains when melting snow frees a boulder in front of you. The boulder begins rolling downhill. Below, you see a tree

directly in the path of the now accelerating boulder. You form the belief that the boulder will hit the tree. Your belief is true: seconds later the boulder smashes into the tree.

What is the content of your belief that the boulder will hit the tree? The answer to this question seems relatively straightforward and unproblematic. We can maintain that the content of one's belief, when one believes that a future event *will* occur mirrors the content of one's belief that a past event *did* occur. If the content of a belief that the boulder did hit the tree is given by applying a past-tense *WAS* operator to a tenseless, temporal proposition *the boulder hits the tree*, the belief that the boulder *will* hit the tree is given by applying a future-tense *WILL* operator to the same temporal proposition: *WILL*(the boulder hits the tree).[1] Opinions vary on whether the resulting tensed propositions are analysable in terms of non-tensed notions.[2] Let us call a belief whose content is given by applying the *WILL* operator to a temporal proposition a "*WILL*-belief".

The picture becomes more complicated when we consider the role of future-directed beliefs in practical reasoning. Many have recently highlighted the important role that (outright) belief plays in practical reasoning: in believing a proposition outright, one is willing to rely on it, to use it as a premise in deciding what to do.[3] Let us state this principle as follows:

(PR) One believes p outright when one is willing to use p as a premise in practical reasoning.

For example, if you believe outright that it is currently pouring rain, you might use this as a premise, together with your desire not to get wet, to conclude that you should get your umbrella. (PR) is an attractive principle that explains some plausible normative features of outright belief and highlights an important difference between outright belief and (even very high) degrees of belief.[4]

[1] Here I understand a 'temporal proposition' to be a proposition that is true at some times and false at others.

[2] For example, I think such propositions can be analysed in terms of tenseless, B-theoretic relations of *being earlier than* and *being later than* the time at which the belief is held, however in this paper I remain neutral on whether such tenseless analyses are possible.

[3] This thesis is most prominently defended by Timothy Williamson. In his (2000), he writes, "Intuitively, one believes p outright when one is willing to use p as a premise in practical reasoning" (p.99). Others who have argued for a similar thesis include Wedgwood (2012), Schulz (2021), and Hawthorne (2004).

[4] Although you may have an extremely high degree of belief in the proposition that your lottery ticket will lose, you refrain from believing outright that your ticket will lose. The difference is explainable in terms of your hesitancy to use the proposition that the ticket will lose in a bit of practical reasoning for selling your ticket for a penny. See Hawthorne (2004) and Williamson (2020).

Suppose we accept (PR). How should we characterize the content of beliefs about the future in light of it? The question raises some complications. Belief, on the one hand, is typically thought of as having a world-to-mind direction of fit. A subject's belief aims to correctly represent how the world is. But we also cannot lose sight of the fact that agents are part of, and shape, the very same world they represent. Beliefs serve as premises in our practical reasoning, but practical reasoning also determines our beliefs. One way to come to believe that you will bring an umbrella is by now deciding to bring an umbrella. Beliefs about the future appear to have a circularity about them: on the one hand, they serve as premises for deciding what we will do. On the other hand, what we decide we will do determines what we believe about the future.

To further examine the question of how we should characterize the content of beliefs about the future given their role in practical reasoning, consider the following variant of the above case:

HIKERS You are hiking in the mountains when melting snow frees a boulder in front of you that begins rolling downward. Below, you see a group of picnicking hikers directly in the path of the now accelerating boulder. You form the belief that the boulder will hit the hikers.

Suppose we assume, plausibly, that your beliefs and desires at this point provide an argument for the conclusion that you should yell out and warn the hikers. As a result, you do yell out "Watch out! Boulder!", the hikers hear your warning, and move out of the way just before the boulder flattens their picnic. What future-directed belief served as a premise in the bit of practical reasoning which led you to yell out and warn the hikers? It cannot be the *WILL*-belief analogous to the belief formed in TREE, that *WILL*(the boulder hits the hikers). If you treat *WILL*(the boulder hits the hikers) as a premise in your practical reasoning, you cannot derive a conclusion consistent with the boulder not hitting the hikers. Using *WILL*(the boulder hits the hikers) might serve as a premise for an argument with the conclusion that you ought to summon the mountain rescue service, or you ought to run and check the casualties, but not for an argument with the conclusion that you ought to perform an action that results in the boulder not hitting the hikers.

So what future-directed beliefs serve as premises in the practical reasoning employed in HIKERS? A lesson in how to think about the content of future-directed beliefs that figure into practical reasoning can be learned from recent literature on mental simulation in cognitive science. In some studies, such as

Battaglia, et. al (2013), subjects are asked to predict whether a configuration of blocks will fall, and, if so, in what direction. The results are compared with a computer-based 'intuitive physics engine', similar to the sort of engine used in many computer games, that produces an output based on an input of initial object states and force dynamics. The study found a strong correlation between the human judgments and the judgments of the computer-based model, hypothesizing that a cognitive process like this model plays a role in human judgments about what will happen.

However, in cases in which the subject can intervene, the model must be complicated. Rather than using the actual, present object-states and force dynamics as input, simulating, and generating an output, *multiple* simulations are run using inputs of *hypothetical* actions on the part of the subject, plus present object-states and force dynamics.[5] It is not merely a matter of how the actual, present situation *will* evolve, but rather how the present situation, together with various hypothetical actions on the part of the agent, *would* evolve. To apply this to HIKERS, the subject runs multiple simulations such as the following:

1. Run simulation with input: present object states, force dynamics plus action: do nothing.

2. Run simulation with input: present object states, force dynamics plus action: yell out warning.

3. Run simulation with input: present object states, force dynamics plus action: jump in path of boulder.

Each simulation will generate its own output, for example, crush hikers, save hikers, get crushed, respectively.

Let us return to our earlier questions: What is the content of the subject's future-directed beliefs when she runs these simulations? What serves as premises in her practical reasoning? Our belief in, and knowledge of, many counterfactuals is closely tied to mental simulation. This connection is explored at length by Timothy Williamson in various places (2007, 2009, 2016). One way we gain knowledge of counterfactuals is by imaginatively simulating how a hypothetical situation would evolve. Such simulations employ the imagination, but the

[5] An example of a study that models mental simulation involving agent intervention is Allen, Smith and Tenenbaum (2020).

imagination is informed and disciplined by background knowledge and past experience. The simulation exercises above may lead the subject performing them to endorse the following counterfactuals:

(1) If I were to do nothing, the hikers would be crushed.

(2) If I were to yell out a warning, the hikers would be saved.

(3) If I were to jump in the path of the boulder, I would be crushed.

The subject's simulations involve imagining a scenario in which the antecedent holds, and deriving the consequent by simulating how the scenario would realistically evolve under such conditions. This imaginative exercise can lead to the subject believing the above counterfactuals, and, assuming the method by which they are derived is sufficiently reliable, the subject can know them.

The future-directed beliefs that we form in a scenario like HIKERS are beliefs in future-directed counterfactuals. We acquire these beliefs by simulating how the scenario would evolve given various hypothetical actions. It is these future-directed counterfactuals, derived through mental simulation, that we employ as premises in our practical reasoning in cases like HIKERS. Many of the beliefs we form about the future that figure into our practical reasoning are derived by simulating how the future would evolve under various hypothetical circumstances. This applies not only to beliefs about the future of our immediate physical surroundings as in HIKERS, but also more remote and less physics-based scenarios, such as how a friend would react to cancelling a lunch date at the last minute, or how a future election would turn out. The simulations we run in such cases are arguably more complex than the simulations in HIKERS, and involve different kinds of cognitive models. For example, in addition to an intuitive physics engine, we also have something like an intuitive psychology engine that simulates mental states and behaviors of others in various hypothetical situations, such as cancelling a lunch date at the last minute.[6]

2 Future-Directed Conditionals

In the previous section I proposed that future-directed counterfactuals serve as the content of our beliefs about the future and the premises in our practical reasoning in scenarios like HIKERS. In this section, I wish to further bolster this

[6] The view that folk psychology involves mental simulation is defended by Nichols and Stich (2003) and Goldman (2006), among others.

proposal by focussing on future-directed conditionals more generally. The aim of the section is to defend a robust distinction between future-directed counterfactual conditionals and future-directed indicative conditionals, a distinction that has been challenged in the conditionals literature, and argue that it is future-directed counterfactuals that serve as premises in our practical reasoning.

Identifying what distinguishes indicative conditionals from counterfactual conditionals is about as controversial as it gets, and there is even less consensus when it comes to future-directed variants of each.[7] First, it can be difficult to tease apart future-directed indicatives from future-directed counterfactuals. A sentence like:

(4) If it rains tomorrow, the picnic will be cancelled.

may well be taken to express an indicative conditional, yet it seems to express the same thing as the conditional sentence expressed in the subjunctive:

(5) If it were to rain tomorrow, the picnic would be cancelled.[8]

At the very least, they don't strike us as clearly distinct in the way in which certain past-directed versions of each do, such as the well-known pair introduced by Ernest Adams (1970):

(6) If Oswald didn't kill Kennedy, someone else did.

(7) If Oswald hadn't killed Kennedy, someone else would have.

(6) and (7) plausibly have distinct meanings and many endorse (6) but deny (7). The conditional expressed by (6) is taken to be an example of an indicative, whereas the conditional expressed by (7) is taken to be an example of a counterfactual. However, a corresponding future-directed pair such as:

(8) If Putin doesn't kill Zelenskyy, someone else will.

(9) If Putin were not to kill Zelenskyy, someone else would.

seem to stand or fall together: one who accepts (8) likely also accepts (9) and *vice versa*.[9] In an effort not to pre-judge any of the relevant issues, let us call

[7] DeRose (2010) refers to the semantics of indicative conditionals as "a swamp" and Frank Jackson warns us that "almost everything about indicative conditionals is controversial" (1998).

[8] DeRose (2010) calls this latter conditional a "were'ed up" version of the former conditional and argues that were'ed up future-directed conditionals express indicative conditionals.

[9] Although I think there are situations where they can come apart. Suppose a reliable oracle tells me that someone will kill Zelenskyy. If I think it is likely that Putin is the one who kills him, I may endorse (8) but not (9).

sentences like (4) and (8) *if-will* sentences and sentences like (5) and (9) *if-would* sentences. This asymmetry between past-directed pairs like (6) and (7), on the one hand, and future-directed pairs like (8) and (9), on the other hand, has led some to deny that what is expressed by *if-will* sentences like (4) and (8) differs from what is expressed by *if-would* sentences like (5) and (9): the meaning of the conditional expressed by both of these two types of sentences is the same. Victor Dudman argues that counterfactual sentences like (7) are just a past-tense form of a future-tensed indicative like *If Oswald doesn't kill Kennedy, someone else will* and since counterfactuals are just a past-tense form of future-tensed indicatives, there is no future-tensed indicative / counterfactual distinction.[10] Keith DeRose (2010) has argued that (a) *if-will* sentences have the same meaning as their "were'ed up" counterparts[11], (b) what is expressed in both cases is an indicative conditional, and (c) it is indicative conditionals that we use in deliberating about what we will do. In what follows, I will argue that all three of DeRose's claims are false: *if-will* sentences do not always express the same thing as their were'ed up counterparts, there is a robust distinction between future-directed indicatives and future-directed counterfactuals, and it is future-directed counterfactuals, rather than indicatives, that typically figure into our practical reasoning.

Of course, pointing to cases where an *if-will* sentence seems to express the same thing as an *if-would* sentence is a bad argument for the conclusion that there is no substantive distinction between future-directed indicatives and future-directed counterfactuals. After all, there are past-directed pairs that are similarly difficult to distinguish, such as,

(10) If *Homo sapiens* interbred with *Homo neanderthalensis*, there are traces in modern human DNA.

(11) If *Homo sapiens* had interbred with *Homo neanderthalensis*, there would be traces in modern human DNA.

And this does nothing to undermine the existence of a robust indicative / counterfactual distinction evidenced by pairs like (6) and (7). It is not difficult to

[10] That counterfactuals just are past-tensed versions of indicatives goes by the name "the relocation thesis" and has been defended most prominently by Dudman (1983, 1984) and is discussed by Bennett (2003, 13-15, 350-355).

[11] DeRose (2010) notes, however, that 'were'ed up counterparts have "a couple of additional components to their meaning": they can call attention to the possibility that the antecdent is (or will be) false and they cannot be felicitously asserted on deliberationally-useless grounds.

come up with future-directed pairs of conditionals that demonstrate a clear difference between indicatives and counterfactuals, like Adams' past-directed pair. Here is a variant of one provided by Adam Morton (2004): Staff Sergeant James is a bomb disposal expert sent to investigate a patch of dirt on the side of the road that likely contains a landmine. Given his expertise, he would never step on a patch of dirt unless he was certain it was safe. His colleague, Sanborn, watches from a safe distance, knowing full well that James would never step anywhere he didn't know was safe. Sanborn says to his colleague standing next to him:

(12) If Sgt James steps on that patch of dirt, it won't explode.

And given James' expertise, Sanborn's utterance is plausibly true and/or felicitously assertable.[12] But given the high likelihood that there is a landmine buried in the dirt, the following conditional utterance is likely judged true in Sanborn's context:

(13) If Sgt James were to step on the patch of dirt, it would explode.

(12) and (13) are plausibly taken to express an indicative and a counterfactual respectively. (12) expresses something about how we would revise our beliefs upon receiving the unexpected news that Sgt James steps on the patch of dirt. (13) is not about how we would revise our beliefs in light of unexpected news, but rather makes a claim about the causal dependence of distinct events.

Here is another such case; this time my own: Emily is an alternate for her local basketball team. She's not planning to attend tonight's game because, as far as she knows, all the regular players are able to play. Her roommate asks her if she is attending the game tonight. She replies:

(14) No. If I were to attend the game tonight, I would not be needed.

At the same time recognizing that there's always a small chance that one of the regular players won't be able to play at the last minute, Emily can also felicitously assert:

(15) If I attend the game tonight, I will be needed.

Pairs such as (12) - (13) and (14) - (15) convincingly show that there is a substantive indicative / counterfactual distinction in the future-directed, as well as the past-directed, case.

[12] I remain neutral on the controversial question of whether indicative conditionals have truth-conditions or merely assertability-conditions.

As many have noted, indicative conditionals express something epistemic in the sense that they concern how we would revise our beliefs in light of new, potentially unexpected, information, such as learning that Oswald didn't in fact kill Kennedy. Counterfactuals, in contrast, oftentimes express dependency relations, frequently (but not always) causal, between distinct events. My diagnosis for why future-directed indicatives are more difficult to tease apart from future-directed counterfactuals than their past-directed counterparts is that oftentimes what we expect to learn in the future is tied up with the causal dependence of the future on the present and earlier future events.[13] What we expect to learn about the future is frequently based on reasoning about how we take the future to causally depend on earlier events. Consider the above pair:

(4) If it rains tomorrow, the picnic will be cancelled.

(5) If it were to rain tomorrow, the picnic would be cancelled.

If I come to learn tomorrow that it is raining, I will come to believe that the picnic will be cancelled and this is due to the fact that the cancellation of the picnic is causally dependent on tomorrow's rain. But examples like the ones involving Sgt James and Emily demonstrate that we can indeed distinguish between in fact learning something unexpected in the future, and how the future would evolve under hypothetical circumstances.

Having defended the distinction between future-directed indicatives and future-directed counterfactuals, we can ask the question: which of these most plausibly figures into our practical reasoning? There are compelling reasons for claiming that it is future-directed counterfactuals. There is a version of (12) *If Sgt James steps on that patch of dirt, it won't explode,* perhaps stated in the first-person, that Sgt James himself may well endorse: he may well think that the only circumstance in which he steps on the patch of dirt is one where he has ascertained its safety. But given his current predicament and his high credence that there is a landmine buried under the patch of dirt, this conditional does not figure into his practical reasoning. His next steps are instead guided by the counterfactual *If I were to step on the patch of dirt, it would explode.* This counterfactual, together with his desire not to get blown up, explains why he chooses the steps he takes. Similarly, it seems that in deciding whether to drive to the game or

[13] Schulz (2019) provides a similar diagnosis: "...forward-looking indicative conditionals are usually about the future and knowledge of the future can often only be gained indirectly by combining knowledge about the present or the past with knowledge about causal dependencies...Such causal connections will normally also support the corresponding counterfactual" (p.1209).

stay home, Emily ought to employ the counterfactual (14) *If I were to show up to the game tonight, I would not be needed* in her practical reasoning rather than the indicative (15) *If I attend the game tonight, I will be needed.*

Another reason in favour of taking the conditionals that figure into our practical reasoning to be counterfactuals rather than indicatives, is that we have use for, and rely on, future-directed conditional premises even when we are certain or near certain that their antecedent is false. This is a feature that distinguishes counterfactuals from indicatives. In distinguishing between indicatives and counterfactuals, Jonathan Bennett notes that whereas counterfactuals are "zero-tolerant", indicatives are "zero-intolerant". Bennett (2003, p.55-57) says that we have "use for" counterfactuals even when we assign zero credence to their antecedent (zero-tolerant), but we do not have "use for" indicatives when we assign zero credence to their antecedent (zero-intolerant). A natural way of making sense of Bennett's notion of whether we have a "use for" a given conditional is whether we would employ it in our reasoning. To have a use for a conditional is for it to be relevant in one's theoretical and practical reasoning. Consider again the case involving Sgt James. Having ascertained that there is a live landmine buried under the patch of dirt, he knows with certainty that he will not step on it. Nonetheless, he still has a use for *If I were to step on it, it would explode.* In fact, this is how he might explain his decision not to step on it, were someone to ask or were he to justify his actions. Another case: Jim has an exam tomorrow and decides to spend the evening studying for it. His friend, Hal, asks him why he plans to stay in and study, rather than join him at the pub. Jim says "If I don't study, I will not pass the exam". This seems like a good justification and an informative, non-trivial claim. But it is difficult to make sense of the fact that the utterance provides good justification and is non-trivial if it is interpreted as expressing an indicative conditional. After all, Jim knows (let us suppose) that he will study and so assigns a low (perhaps zero) credence to the antecedent. Despite this, he still has a use for the conditional, it serves as his justification to Hal for his staying home. This suggests it is a zero-tolerant counterfactual, rather than a zero-intolerant indicative, that Jim bases his decision on.[14] So the fact that we see many future-directed conditionals with antecedents we know to be false as worthy of endorsement, assertion,

[14]Consider a past-directed case. Jim studied last night and took the exam this morning. He has little use for the indicative "If I didn't study, I didn't pass" since he knows he did study. But he endorses and does have use for the counterfactual "If I hadn't studied, I wouldn't have passed".

and reliance in our theoretical and practical reasoning, suggests that they are zero-tolerant counterfactuals rather than zero-intolerant indicatives.

This example involving Jim illustrates that, although there is a distinction between future-directed counterfactuals and future-directed indicatives, this distinction does not always line up with the distinction between *if-will* and *if-would* sentences. The *if-will* sentence expressed by Jim: "If I don't study, I will not pass the exam" is plausibly interpreted as a future-directed counterfactual rather than a future-directed indicative. The context is such that, given his decision to stay home and hence his low (perhaps zero) credence in the antecedent, taking Jim to be expressing an indicative would be uncharitable. In some contexts, whether an *if-will* expresses an indicative or a counterfactual may not matter much, such as in a typical utterance of "If it rains tomorrow, the picnic will be cancelled." However, in other contexts where one's knowledge of the future and causal dependency between future events come apart, *if-wills* may plausibly be interpreted as expressing future-directed counterfactuals, as in an utterance of "I will bring my umbrella. If I don't bring it, I will get soaked". Interpreting the *if-will* conditional as expressing an indicative would be to attribute an odd and sudden change of mind to the speaker, substantially favoring the counterfactual interpretation.

One *prima facie* worry with taking the conditionals that are employed in practical reasoning in cases like HIKERS to be future-directed counterfactuals is that in deciding what one will do, one considers what will actually be the case. Don't counterfactuals concern *non-actual* possibilities? Don't they involve considering how alternative, non-actual histories would go? This worry is misplaced. As many have noted, counterfactuals are poorly named.[15] Although frequently we take the antecedent of a counterfactual to be 'counter to fact' or non-actual, there are many instances of counterfactuals that do not involve this commitment. Such cases are given by Alan Ross Anderson (1951). Suppose a doctor is examining an ill patient Jones and attempting to diagnosis him based on his symptoms. She notes:

(16) If Jones had taken arsenic, he would have shown just exactly those

[15]The name 'subjunctive conditional' is no better since many conditionals that are plausibly interpreted as what I am calling 'counterfactuals' here are not expressed in the subjunctive, such as Jim's utterance "If I don't study, I will not pass". I am not happy about using the name 'counterfactual' for the sort of future-directed conditionals that I believe figure into practical reasoning since it is crucial to my view that believing and asserting such a conditional need not, and frquently does not, involve commitment to the antecdent being counter to fact/false. I reluctantly use the term to be consistent with its use in the conditionals literature.

symptoms which he does in fact show (Anderson, 37).

Or consider an environmental historian of the Americas noting:

(17) If there had been intensive agriculture in the Pre-Columbian Americas, the natural environment would have been impacted in specific ways. That is exactly what we find in many watersheds. ("Counterfactuals" entry, SEP)

In these instances of past-directed counterfactuals, we do not take the antecedent to be false or non-actual, but, on the contrary, we may use such counterfactuals in abductive arguments for the *truth* of antecedent: that Jones in fact took arsenic or that there in fact has been intensive agriculture in the Pre-Columbian Americas. What I take to be essential to counterfactuals is that we treat the antecedent as *hypothetical* without commitment to it being actual. This is, of course, compatible with, but does not require, the presupposition that the antecedent is non-actual. In considering counterfactuals, we consider a hypothetical situation while either being committed to, or agnostic about, whether it is non-actual. Oftentimes, context will determine whether the presupposition is in play.

I suggest that our beliefs about the future, in a case like HIKERS, involve beliefs in various counterfactuals, but believing these counterfactuals does not involve commitment to the antecedent being non-actual. In believing various future-directed counterfactuals in cases like HIKERS, we are not entertaining beliefs about futures that we take to be non-actual. Instead, believing these counterfactuals involves various beliefs about what the future would be given various hypothetical suppositions involving the present state of the world combined with possible actions on the part of the subject. Belief in each counterfactual involves belief in a future that would come about were certain actions performed. Prior to deciding, these counterfactuals represent hypothetical futures, and the agent brings about the actual future by deciding what action to perform. The way in which we represent the future is nicely described by Jenann Ismael:

> When one is looking into the future, one represents one's own choices in hypothetical form, sees a range of actions that directly or indirectly depend on them, and makes the choice by imaginatively tracing out their downstream consequences and comparing the results. The choice itself is the product of this imaginative exploration and

> the decision-maker is right to treat it in that context in hypothetical form (Ismael 2013, 163).

This quote captures the hypothetical stance we adopt towards our range of actions and their downstream consequences. It also echoes the mental simulation account of how these hypotheticals are arrived at by "imaginatively tracing out their downstream consequences" as a process of "imaginative exploration". It is precisely this imaginative exploration that allows us to arrive at the future-directed counterfactuals that we employ in our practical reasoning.

3 The Actual Future and OF-Counterfactuals

I wish to further explore the idea that our beliefs about the future involve believing counterfactuals of a certain sort by considering how our conception of actuality interacts with past-directed vs future-directed counterfactuals. Our attitudes towards the actual future differ in interesting ways from our attitudes towards the actual past. I think this asymmetry supports the idea that we construe the actual future, at least in part, in hypothetical terms and dependent on what we do now, in contrast to how we construe the actual past. I suggest that this asymmetry in how we construe the actual future vs the actual past is central to how we mentally represent the future as open in contrast to the past.

Suppose that last night you turned down a friend's invitation to go to the pub and you stayed home instead. Consider the following counterfactual:

(18) If I had gone to the pub last night, I would have actually gone to the pub last night.

Although 'actually' has many senses, I am interested here in the rigidifying reading of 'actually' according to which it refers back to the world of the context of evaluation as opposed to the counterfactual world at which the antecedent holds, as in the example "If Max ate less, he would be thinner than he actually is."[16] On this reading, (18) seems clearly false: after all, you did not actually go to the pub last night. This is borne out by the Lewis-Stalnaker analysis of counterfactuals: in the closest world in which you go to the pub last night, it is false back here in the actual world that you went to the pub last night.

[16] The example is Lewis's from his "Anselm and Actuality" in which he distinguishes the rigidifying reading from a non-rigid reading. 'Actually' actually has several different uses, including acting as a discourse marker.

Now suppose a friend has invited you to the pub tonight and you are genuinely undecided about whether to go or whether to stay home. You consider the following counterfactuals:

(19) If I were to go to the pub tonight, I would actually go to the pub tonight.

(20) If I were to stay home tonight, I would actually stay home tonight.

It seems plausible to maintain that (19) and (20) are both trivially true. We lack the intuition that one of them is false in the way we do with past-directed counterfactuals like (18). At the very least, there seems to be an asymmetry in how we evaluate past-directed counterfactuals containing a rigidifying 'actually' in the consequent versus how we evaluate future-directed counterfactuals containing a rigidifying 'actually' in the consequent. Similarly, at the time of your deliberation, it seems acceptable to judge both of the following counterfactuals as true:

(21) If I were to decide to go to the pub, I would actually go to the pub.

(22) If I were to decide to stay home, I would actually stay home.

In contrast, in the case in which you turned down a friend's invitation to go to the pub last night and you stayed home, the following is plausibly false:

(23) If I had decided to go to the pub last night, I would have actually gone to the pub last night.

Again, you did not actually go to the pub last night, so at the closest world in which you decided to go, it is false back here in the actual world that you went to the pub last night. Given our role in shaping our (actual) future, what the actual future will be in many cases depends on what we do now in a way in which our actual past does not, and this dependence of the actual future on what we do now is reflected in the relevant counterfactuals. Obviously, in many cases, how the actual future will be does not depend on what I do now, and so counterfactuals like:

(24) If I were to decide to go to the pub this evening, it would actually rain in Melbourne tomorrow.

are plausibly false if it fails to actually rain in Melbourne tomorrow, regardless of whether I decide to go to the pub. However, there are some future-directed

counterfactuals, like (21) and (22) that state a dependence of the actual future on what we do now. I call these 'open future counterfactuals' or 'OF-counterfactuals', for short.

Let us return to HIKERS. When you observe the boulder beginning to roll down the mountain, you consider various ways in which the actual future depends on what you do now. If you were to do nothing, the hikers would be crushed in the actual future. If you were to yell a warning, the hikers would be saved in the actual future. These observations motivate a modification of the above proposal where instead of taking the counterfactuals (1)-(3) as the content of the subject's belief, we take as the content of the subject's belief the following OF-counterfactuals:

(25) If I were to do nothing, the hikers would be actually crushed.

(26) If I were to yell out a warning, the hikers would actually be saved.

(27) If I were to jump in the path of the boulder, I would actually be crushed.

I propose that it is these OF-counterfactuals that we believe outright, and that figure into our practical reasoning. Furthermore, assuming the simulation that you used to derive the counterfactual is sufficiently accurate, the OF-counterfactuals that you come to believe may well be true at the time of deliberation and, assuming the process by which you arrived at them is sufficiently reliable, they can be known. Prior to deciding, these counterfactuals represent the hypothetical futures that would become actual were you to perform the action described in the antecedent.

This raises some delicate issues about the truth-value of certain counterfactuals before and after deliberation. On the current proposal

(21) If I were to decide to go to the pub, I would actually go to the pub.

and

(22) If I were to decide to stay home, I would actually stay home.

are both true on, say, Friday evening when I am undecided about whether to go to the pub. However, after staying home for on Friday evening, the following counterfactual:

(23) If I had decided to go to the pub last night, I would have actually gone to the pub last night.

is false Saturday morning. Is this shift in truth-value problematic? I'm not convinced it is. It may well be that the *sentence* (21) expresses a truth whereas the *sentence* (23) expresses a falsehood, but it doesn't follow that there is a unique counterfactual expressed by both sentences that changes in truth-value between the earlier and the later time. The difference in tense may mean that different counterfactuals are expressed by (21) on Friday and (23) on Saturday, and it may be that the referent of 'actually' differs between the earlier and the later context.

4 OF-Counterfactuals and the Open Future

One of our strongest, most fundamental intuitions about the nature of time is that the past is fixed, settled, and closed, whereas the future is unsettled and open. What exactly does this asymmetry in openness amount to? How is it best understood? There are various ways in which philosophers have attempted to characterize the asymmetry in openness between the past and future: in terms of an asymmetry in knowledge of past and future, an asymmetry in truth-value: bivalence fails for future contingents in a way in which it does not for past contingents, in terms of an ontological asymmetry: the future involves numerous branches whereas the past involves a singular trunk, or our world is a growing block where past events are ensconced in spacetime and future events have yet to be. I do not find any of these attempts at characterizing the asymmetry in openness to be plausible for reasons that I have given elsewhere.[17]

The account of the asymmetry in openness between past and future that I find most attractive is that briefly offered by David Lewis in "Counterfactual Dependence and Time's Arrow" (1979). Lewis claims "The literal truth is just that the future depends counterfactually on the present. It depends, partly, on what we do now...In short, I suggest that the mysterious asymmetry between open future and fixed past is nothing else than the asymmetry in counterfactual dependence" (Lewis, 1979, p.38). The future state of my office floor depends on what happens presently, such as whether I presently knock over my coffee mug, in a way in which the past state of my office floor does not. The future state of the news cycle depends on whether a political scandal breaks out today in a way in which the past state of the news cycle does not. The future state of the surface of Mars depends on whether there are Martian meteor showers today in

[17] See Torre (2012) and (2023).

a way in which the past state of the surface of Mars does not.

Lewis's characterization of the asymmetry in openness between the past and the future is a metaphysical characterization. It is a fact about our world: that future states of it counterfactually depend on the present in a way in which past states of it do not. As an account of openness, it is broader than, and is not rooted in, human agency or human representation, as the above example involving Mars demonstrates. It makes no attempt to characterize the asymmetry in our representations or attitudes about the world, but rather an asymmetry in our world itself.

What I have offered in this paper is an account of the asymmetry in our doxastic attitudes about the past versus the future. When it comes to beliefs about past events, the content is plausibly characterized by applying a past-tense *WAS* operator to a tenseless description of a state or event that our belief concerns, as in *WAS*(the boulder hits the tree). If we were mere observers of our world, in the business of representing past, present and future states of the world we inhabit, a similar account could be given for beliefs about the future. The content could plausibly be characterized by applying a future tense *WILL* operator to a tenseless, temporal proposition, for example, *WILL*(the boulder hits the tree). But, we are not merely in the business of representing past, present and future states of the world we inhabit. We play a causal role in bringing about the very future states and events that we seek to represent. This leads to the circularity about future belief described in Section 1. To accommodate this feature and to explain how our representations of the future figure into our practical reasoning, we must recognise an asymmetry in our doxastic attitudes about the past versus the future. Whereas beliefs about the past can be characterized in terms of *WAS*-beliefs, beliefs about the future substantially involve representing how hypothetical futures depend on what we do now. And this dependency of the actual future on what we do now is represented in terms of future-directed counterfactuals.

The asymmetry in belief content fits well with metaphysical theory of openness proposed by Lewis. According to Lewis's theory of openness in terms of counterfactual dependence, the future state of the hikers depends on what you do now. Fortunately for the hikers, this fact about the world is mirrored in your beliefs about how their future depends on what you do now.[18]

[18]I would like to thank Jesper Kallestrup, Ned Markosian, Moritz Schulz, Glen Scislowski, Francisca Silva, Paula Sweeney, Brandt van der Gaast, as well as the audience at the *Perspectives on Time* conference in Geneva, the audience of the *Lancog Summer Metaphysics*

References

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

[2] Kelsey R. Allen, Kevin A. Smith, and Joshua B. Tenenbaum. Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 117(47):29302–29310, 2020. Place: US Publisher: National Academy of Sciences.

[3] Alan Ross Anderson. A Note on Subjunctive and Counterfactual Conditionals. *Analysis*, 12(2):35–38, December 1951.

[4] Peter W. Battaglia, Jessica B. Hamrick, and Joshua B. Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332, November 2013. Publisher: National Academy of Sciences Section: Biological Sciences.

[5] Jonathan Bennett. *A Philosophical Guide to Conditionals*. Clarendon Press, 2003.

[6] K. DeRose. The Conditionals of Deliberation. *Mind*, 119(473):1–42, 2010. Publisher: Oxford University Press.

[7] V. H. Dudman. Tense and Time in English Verb Clusters of the Primary Pattern. *Australian Journal of Linguistics*, 3(1), 1983.

[8] V. H. Dudman. Conditional interpretations of if sentences. *Australian Journal of Linguistics*, December 1984. Publisher: Taylor & Francis Group.

[9] Alvin I. Goldman. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Philosophy of Mind Series. Oxford University Press, New York, 2006.

[10] John Hawthorne. *Knowledge and Lotteries*. Oxford University Press, 2004. Publication Title: Knowledge and Lotteries.

[11] Jenann Ismael. Decision and the Open Future. In *The Future of the Philosophy of Time*. Routledge, 2013. Num Pages: 20.

Workshop 2024, and an anonymous referee for this journal for helpful comments on earlier versions of this paper.

[12] Frank Jackson. Indicative conditionals. In *Routledge Encyclopedia of Philosophy*. Routledge, London, 1 edition, 2016.

[13] David Lewis. Anselm and Actuality. *NoÃ»s*, 4(2):175–188, 1970.

[14] David Lewis. Counterfactual Dependence and Time's Arrow. *NoÃ»s*, 13(4):455–476, 1979.

[15] Adam Morton. Indicative versus Subjunctive in Future Conditionals. *Analysis*, 64(4):289–293, 2004. Publisher: [Analysis Committee, Oxford University Press].

[16] Shaun Nichols and Stephen P. Stich. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford Cognitive Science Series. Oxford University Press, Oxford, 2003.

[17] Moritz Schulz. The Past Tense View of Counterfactuals Revisited. *Ergo, an Open Access Journal of Philosophy*, 6, 2019.

[18] Moritz Schulz. Practical reasoning and degrees of outright belief. *Synthese*, 199(3):8069–8090, December 2021.

[19] W. Starr. Counterfactuals. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2022 edition, 2022.

[20] Stephan Torre. The Open Future. *Philosophy Compass*, 6(5):360–373, 2011. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1747-9991.2011.00395.x.

[21] Stephan Torre. The Growing Block, the Open Future and Future Truths. *Disputatio*, 13(63):423–432, December 2021.

[22] Ralph Wedgwood. Outright Belief. *Dialectica*, 66(3):309–329, 2012. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1746-8361.2012.01305.x.

[23] Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, Oxford, 2000.

[24] Timothy Williamson. *The Philosophy of Philosophy*. Wiley-Blackwell, Malden, MA, 2007.

[25] Timothy Williamson. Knowledge of Counterfactuals. *Royal Institute of Philosophy Supplements*, 64:45–64, July 2009.

[26] Timothy Williamson. Knowing by Imagining. In Amy Kind and Peter Kung, editors, *Knowledge Through Imagination*, pages 113–123. Oxford University Press, March 2016.

[27] Timothy Williamson. Knowledge, Credence, and the Strength of Belief (version of 31 May, 2020, to appear in Amy Flowerree and Baron Reed (eds.), Expansive Epistemology: Norms, Action, and the Social World, London: Routledge). 2020.