It is currently fashionable to talk about "synchronic conditionalization" – and more generally, synchronic or time-slice versions of norms that are normally understood as diachronic. But what is a synchronic version of conditionalization? Few authors address this question directly,[1] but one often sees this synchronic entity labeled as a "plan", "policy", or "disposition".

I want to look at these labels a little more carefully. I will argue that conditionalization is a *bad plan*. More precisely: the way we naturally assess plans makes conditionalization look bad. But being *disposed* to conditionalize is good. That is, the way we naturally assess dispositions makes conditionalization look good like a good one to have. So if we want to defend a synchronic analog to conditionalization, we should go with dispositions, not plans.

At the end of the paper I'll argue that our way of assessing plans has more in affinity with another important epistemic concept: deference.[2]

1. Warmup

It will be helpful to have a learning situation to keep in mind as an example. Consider this situation.

> **Pet.** Worn down by years of begging and a few months of good behavior, I'm choosing a low-maintenance pet for my son. My choices are a green lizard and red salamander.

Suppose it turns out that I got the lizard. What color is my pet?

I hope you said "green". That was an easy one. After all, you knew that only the lizard was green. So in full-belief terms, you knew "if green, then lizard" and you just learned "green". In credence terms, your prior credence that the pet is a lizard given that it's green, Pr(Lizard|Green), was very high (maybe even 1) and your prior credence that the pet is a salamander given that it's green, Pr(Salamander|Green) is very low (maybe even 0). Upon learning that the pet is green, your probability that it's a lizard should be high as well.

Now consider:

> **Pet 2 (extra salamander).** As before, but besides the green lizard, there are *two* salamanders – a spotted one and a red one.

Suppose it turns out I got a salamander. What color (or pattern) is my pet?

This is a little harder. Most likely, you aren't confident enough to give a single answer. If you are confident of one particular answer, it's probably because of other things you know about the situation. Maybe you know that the red salamander was more expensive, so "spotted" seems like a better answer; or maybe you know that red is my son's favorite color, so "red" seems

[1] See Zendejas Medina [2023] for one of the more detailed recent treatments of this issue. See also Carr [ms].

[2] The discussion here dovetails with some recent literature in epistemic utility theory: most notably, Carr [2017], Konek and Levinstein [2019], Schoenfield [2015] and [2017], and Zendejas Medina [2023]. Schoenfield [2015] argues that conditionalization is a good plan to *follow*, but a bad plan to *make*; I'll argue that it is neither. (I return to Schoenfield's argument at the end of the paper.) The main contribution of this paper is to unify these problems and present them in a more neutral framework, without presupposing any particular picture of accuracy or any particular way of formalizing arguments for conditionalization. There is also some overlap with Chapter 7 of my book manuscript in progress; however, this paper goes beyond the discussion in the book chapter.

better. In the absence of any of this extra information – that is, when given the rest of what you know, "red" and "spotted" seem like equally good answers – it seems reasonable that you'd say, "I don't know" or that you'd put about 50% probability on each of "red" and "spotted".

Let's take all of that as data. Our theory of learning should explain why these responses to learning are rational, or at the very least, why they seem rational. The next section will look at a popular way to formalize those judgments.

2. Conditionalization

In the scenarios above, it's plausible that we are trying to figure out a function from evidence to belief states, like this:

Input (evidence):	Lizard	**Output (belief state):**	Green
Input (evidence):	Salamander	**Output (belief state):**	50% spotted, 50% red

To fill in the outputs, it makes sense to ask yourself questions like: "How likely is Sophie's pet to be green, given that it's a lizard?" and "How likely is Sophie's pet to be spotted, given that it's a salamander?" To see if the answers are any good, we evaluate the accuracy of the *output*, for the state of the world where the *input* is true. In the original case, we judge that the answer "Green" is definitely accurate in the state of the world where the input "Lizard" is true. In the second case, we judge that the answer "Spotted" has about a 50% chance of being accurate when "Salamander" is true, as does the answer "Red". It seems plausible that you should assign degrees of belief in that same proportion.

This line of reasoning also gets us to a very popular rule in formal epistemology:

Conditionalization: $\Pr_E(H) = \Pr(H|E)$

This says that upon receiving evidence E, your credence (degree of belief, subjective probability) in H should equal your *prior conditional credence* in H, given E.

I won't try to fully defend conditionalization here. For now I have two simpler aims. First, I want to give a general motivation for the rule: it looks like a good way to formalize the intuitive judgments about the cases I gave you. Second, I want to motivate a particular way of thinking about how that rule is justified, exemplified by the questions I listed above. This mode of evaluation is in fact very close to that used by Greaves and Wallace [2006], as well as Zendejas Medina's more recent ([2023]) defense: though these authors' approaches differ in certain ways, both could be fairly characterized as ways of formalizing the question, "how likely is it that Sophie's pet is green, given that it's a lizard?"

3. Synchronic conditionalization

Now that we have motivated conditionalization, let's return to the idea of *synchronic* conditionalization. What is the synchronic analogue – or, if you like diachronic norms, the synchronic *component* – of conditionalization? Many people say: **a plan**.[3]

[3] Easwaran [2013], for example, defends a synchronic norm called "Plan Conditionalization". Zendejas Medina [2023] defends conditionalization as a "conditional plan". (His view ends up being quite close to what I will defend, though I'll argue that it is not best thought of as a plan). Briggs and Pettigrew [2020] discuss the idea of a "credal strategy", which they define as "a two-stage plan" that specifies both initial credences and credences to adopt after learning. Teller's [1973] argument also the term "plan". Greaves and Wallace [2006] use the term "policy." While

The view is pretty intuitive. You might express it like this: "If Sophie's pet is a lizard, it's *very likely* (guaranteed) that her pet is green. So in that scenario, I plan to *believe that her pet is green*." (I've expressed this in full-belief terms because those are more natural, but it would work just as well to say, "… I plan to be *highly confident*," "I plan to be *certain*", or "I plan to adopt *credence 1* that her pet is green.")

You have a plan for what to believe about my pet's color if I get a lizard. *Lizard* is the state of the world where that plan is carried out, and believing *Green* (or having high confidence in *Green*) is the output of the plan. So this plan looks like a good way to represent the results of your evaluation: which answers will be most accurate in different states of the world that could obtain.

The best part is that plans are synchronic: regardless of what happens later, you can have that plan *right now*. Want to turn it into the synchronic part of a diachronic plan? Easy. Just add a norm that tells us to follow through on our plans.

This reasoning looks pretty good. But I will argue that it's mistaken.

My objection will rely on a couple of assumptions about how plans are most naturally evaluated. (I think these assumptions are true, but they lead to the conclusion that conditionalization is a bad plan.) The first is very general:

First-person: Plans are most naturally evaluated prospectively and first-personally.

Arguably, the most important or central mode of evaluation for a plan is from the perspective of someone making the plan – that is, someone looking ahead to the future and predicting where different courses of action might lead her. That's not to say that we can't evaluate plans in other ways (retrospectively or third-personally); of course, we can do that too. But since the purpose of plans is to guide one's own choices, a good plan had better look good from that perspective.

The second assumption is as follows:

Execution: A plan is good insofar as it yields good results when it's executed.

Since a plan is made to be carried out, it's natural to focus on the plan's results. Because of **First-person**, we can interpret **Execution** as referring to the results we can expect the plan to have, prospectively, from the first person point of view.

As before, there are other things that might make a plan good or bad. For example, we might criticize a plan for being too ambitious. But even in that case, **Execution** still seems to capture an important aspect of evaluation. Compare the following two plans. Plan A: if I win the London Marathon, I plan to give half of the award money to charity and save the other half for my kids' education. Plan B: if I win the London Marathon, I plan to take the award money in cash and throw it all down a sewer grate. Clearly, neither plan is feasible for me, and both are kind of silly to make. But the first one is still better; **Execution** can explain why.

Finally: since the kind of evaluation we are interested in is epistemic, the kind of *goodness* we are interested in is accuracy.

---

most of these authors seem to use "plan" as a mathematical label without committing to much of a substantial theory of plans, some do take it seriously. See Schoenfield [2015] for an extended defense of plans as a way to understand rationality. See Carr [2019] for related discussion (though put in terms of decision theory) of how to interpret many of the thought experiments and illustrations used by authors in the epistemic utility theory literature.

> **Accuracy First:** In an epistemic context, a plan yields good results insofar as it yields an accurate belief state.

Again, given **First-person**, we should interpret this assumption as referring to the results we can expect the plan to have from the first-person point of view. We can have good plans that (unluckily) happen to go badly, as when we receive misleading evidence.

Accuracy First is (to my mind) intuitive, and similar principles have been widely adopted in the literature – often by the very same people who support synchronic conditionalization.[4] You could argue for it from a couple of different angles. Some might support it on the basis of axiological commitments that they take to be bedrock: for instance, the view that truth is of the utmost epistemic value. So a good plan is one that (you expect) promotes this foundational value.

Another route might appeal to immodesty, or the idea that we should take our own doxastic states to be the best available options in terms of accuracy. If we are rationally required to be immodest, one might argue that we are implicitly committed to something along the lines of **Accuracy First**. If we are required to see ourselves as maximizing accuracy, then on pain of inconsistency, we are also required to make our epistemic plans in a way that accords with that self-evaluation. There is something odd about evaluating our plans along one dimension of value, and our current states – which are, one would hope, the results of those plans – along another.[5]

So we evaluate our plans according to how well their results pan out, and we evaluate these results in terms of accuracy; all of this is framed from within the subject's own point of view. So far so good; isn't that exactly what we were doing a minute ago, when we asked, "How likely is it that Sophie's pet is green, given that it's a lizard?" "How likely is it that Sophie's pet is green" evaluates the accuracy of the belief (or high credence) that the pet is green, and "it's a lizard" gives us the circumstances under which we're planning to believe that the pet is green. Right?

Not exactly. I will argue that this question does not correctly capture the way we evaluate plans, at least if we take **Execution** seriously.

4. Problems with Plans

4.1 First problem: Learning

Here's a fact about your plan to believe that my pet is green. It doesn't get carried out in just any old world where *it turns out that I get the lizard*: rather, you need to *learn* that I get the lizard! Your learning is necessary to set the plan into motion.[6]

But this changes the state of the world for which we are evaluating the plan. Instead of asking:

---

[4] For example: Pettigrew, Easwaran, and Greaves and Wallace.

[5] I will leave this second line of argument rough, for now, because I am going to argue that it's not quite right. But I think the basic idea is compelling.

[6] See Schoenfield [2017] for previous discussion of this problem. Schoenfield argues that Conditionalization* is what's supported by Greaves and Wallace [2006]'s defense of conditionalization, based on a natural interpretation of their formal argument. See also Zendejas Medina [2023].

"How likely is it that Sophie's pet is green, given that it's a lizard?"

We are now asking:

"How likely is it that Sophie's pet is green, given that *I learn it's a lizard*?"

This lets us evaluate the belief that my pet is green, for the state of the world in which that plan is carried out – that is, a state of the world in which you learn that my pet is a lizard.

But changing the question already seems to justify a different norm. Instead of Conditionalization, we will have:

Conditionalization*: $Pr_E(H) = Pr(H|\text{Learn } E)$

Conditionalization* is different from Conditionalization, because E and Learn E are two different propositions. And while it probably doesn't make a difference in this particular case, we can imagine cases in which the fact that I am learning E makes a difference to the probability of H. So when we take learning into account, we get a different formula for what new credence in H to adopt, and will sometimes result in you adopting a different credence in H.

4.2 Second problem: Uptake

Now notice something else. Your plan is only carried out if you *come to believe that my pet is green*.[7] (After all, that's what it is to carry out the plan!)

This means that we know *even more* about the state of the world relevant to evaluating this plan! Going by **Execution**, instead of asking:

"How likely is it that Sophie's pet is green, given that I learn it's a lizard?"

we should really ask:

"How likely is it that Sophie's pet is green, given that I learn it's a lizard *and come to believe that it's green*?"

In other words, we should follow:

Conditionalization:** $Pr_E(H) = Pr(H|\text{Learn } E \text{ and come to adopt } Pr_E(H))$

By this reasoning, if we think that something is a good plan, it had better turn out to yield good results *on the assumption that it's carried out.* If, under the assumption that Plan A has been carried out, it turns out that Plan B have been better, this is a mark against Plan A.

The state of the world which is relevant for evaluating our plans – what follows "given that" – is now very different from what we started with. And conditionalization** will arguably give very different recommendations from conditionalization in cases where your belief state about H makes a difference to H's probability.

Let's work through an example where this consideration arguably makes a difference:

[7] See Carr [2019] for further discussion of this problem. Like Schoenfield, Carr argues that the problem arises from a plausible interpretation of the formal framework used by Greaves and Wallace, and others.

> **Pet 3 (self-fulfilling salamander).** I am at the pet store, choosing between a green lizard, red salamander and a spotted salamander. I like them about equally, but I can only choose one. In addition, I really like when you're right about things, and I have an uncanny ability to make my choices match your beliefs. If you believe that my pet is red, I'll pick the red salamander. And if you believe that my pet is spotted, I'll pick the spotted salamander. If you remain neutral, I'll flip a coin.

Suppose you learn that my pet is a salamander. What should you believe about its color? Above, we said it seemed rational to be about 50/50 – that is, Pr(Red) = .5 and Pr(Spotted) = .5. But now plug this recommendation into the question we just introduced:

> "How likely is it that my pet is red, given that I learn it's a salamander and come to have 50% probability that it's red?"

According to **Pet 3**, the probability that it's red, given that you adopt Pr(Red) = .5, is 50%. Now plug in an alternative plan, one that says that if you learn it's a salamander, you should just outright *believe* that it's red.

> "How likely is it that Sophie's pet is red, given that I learn it's a salamander and come to believe that it's red?"

How likely is that? Well, it's guaranteed!

Which response is more accurate? There is something fitting about assigning .5 credence to a proposition whose probability is 50%, and as we said before, this does seem like the *rational* response. But if accuracy is about closeness to the actual truth value, surely you would be *more accurate* if you took the second response, outright believing that it's red: an outright belief in a true proposition is more accurate than a middling probability could ever be.[8]

Let's bring this back to conditionalization**. We got to conditionalization** by arguing that a good plan is one that yields good results in the state of the world in which it's executed – and then incorporating the information that the plan has been executed. We are comparing two plans:

> Plan A recommends that, in Pet 3, you adopt Pr(Red) = .5.
> Plan B recommends that, in Pet 3, you adopt Pr(Red) = 1.

If the reasoning above is right, Plan B is better than Plan A. While Plan A yields only middling accuracy when carried out, Plan B yields perfect accuracy.

To be clear about the problem here: the problem is not that some plans look better than others. This is what we expected! (Think about Plan C, which recommends that, in **Pet 3**, you adopt full confidence that my pet is a fish. Plan C is guaranteed to be very inaccurate, and this method of evaluation rules it out, just as we intended.) Rather, I think this example illustrates two different problems. The first is that this method of evaluation – starting with the assumptions

[8] This line of thought seems to be borne out by the most popular ways of thinking about accuracy: someone who fully believes (or is very highly confident of) that my pet is red will be perfectly accurate, whereas someone who has Pr(Red) = .5 will be less than perfectly accurate.

First-person, **Execution**, and **Accuracy First**, leads us away from conditionalization. The second problem is that this method of evaluation seems to give the *wrong* answer, ruling out a response that looks rational and recommending one that looks irrational. Something has gone wrong.

4.3 Third problem: Rational certainty

By now there are enough problems, and it seems likely that the culprit is our mode of evaluation rather than conditionalization itself. But let's add just one more, both for completeness and for insurance (in case some readers are not convinced by the previous problems). Here is the problem. Because we are trying to evaluate these plans as *rational* responses to our evidence, it seems we can specify just one more thing that when we do follow the rational plans, we are responding rationally.[9]

This specification changes our central question again. We are now asking something like:

> "How likely is it that Sophie's pet is red, given that I learn it's a salamander and come to believe that it's red, and *it's rational to believe that it's red*?"

Or, **Conditionalization***:** $Pr_E(H) = Pr(H|$Learn E and come to adopt $Pr_E(H)$, and $Pr_E(H)$ is rational$)$

Does Conditionalization*** give different recommendations from the previous rules we have considered? It looks like it could. Suppose we have some evidence *against* $Pr_E(H)$ being the rational response to E. Conditionalization*** looks like it will tell us to always ignore such evidence. That's because we are only supposed to evaluate our response in a state of the world in which $Pr_E(H)$ *is* rational – so we can't account for the possibility that it might not be.[10]

4.4 Is Execution the problem?

I have argued for three ways in which focusing on the evaluation of plans, given the natural assumptions I started with, leads us away from conditionalization. All three have relied on our knowledge about the way the world will be in cases where the plan is executed. So one might naturally wonder: can we just give up **Execution**?

Schoenfield [2015] makes a suggestion along these lines. Her motivation is the fact that we sometimes predictably fail to follow through on our plans. Because of this, she argues, we sometimes make plans that take our predictable failures into account: we'll plan to leave at 10, for example, knowing that this will get us out the door at 10:30. Schoenfield argues that we see the same phenomenon in the epistemic case: sometimes the best plan to *make* can come apart from the best plan to *follow*. (Schoenfield's argument assumes that "best" is evaluated in terms of accuracy, in line with **Accuracy First**.)

Here is an example discussed by Schoenfield, as well as elsewhere in the literature on higher-order evidence.

---

[9] For now let's just focus on propositional justification: so when I say "responding rationally", I mean, "responding in line with the rational requirements".

[10] Schoenfield [2015] discusses this problem as well, arguing that the best plan to follow would necessitate rational omniscience. Although she does not draw this conclusion here, I think the argument she gives could extend to the other problems I raise in this section.

Hypoxia: You're flying a small unpressurized plane at high altitude, and you're wondering if you have enough fuel to make it to Paris. You do some calculations and determine that you can make it. (In fact, your evidence does support this conclusion.) Then you look at the altimeter and realize that you're flying so high that you're at risk of hypoxia, which would impair your judgment and ability to calculate.

In this case, let's define the following propositions:

Enough Fuel: You have enough fuel to make it to Paris.

Fuel Evidence: The total evidence that directly bears on whether your fuel is sufficient, including the readings of various dials and gauges, and various calculations.

Too High: Your current altitude puts you at risk of hypoxia.

Your total evidence right now is Fuel Evidence & Too High. As we have stipulated, Fuel Evidence supports Enough Fuel; if Fuel Evidence were your total evidence, you should believe (or be highly confident of) Enough Fuel. But what should you believe when your evidence also includes Too High?

Let's return to the form of question we began with at the outset of this section:

"How likely is Enough Fuel, given Fuel Evidence and Too High?"

Presumably, the answer to the question is: Enough Fuel is highly likely, or even guaranteed to be true under these circumstances. As many authors have pointed out, your possible hypoxic state has no bearing on whether you have enough fuel, and you can expect this a priori. So Conditionalization would yield accurate results if followed:

Conditionalization: $Cr_{\text{Fuel Evidence + Too High}}(\text{Enough Fuel}) = Cr(\text{Enough Fuel}|\text{Fuel Evidence + Too High})$

Since your conditional credence in Enough Fuel, given Fuel Evidence + Too High, is very high, conditionalization recommends coming to have very high credence in Enough Fuel. And arguably, so do the modified rules I have discussed: presumably your conditional credence in Enough Fuel given *everything* we know about the situation (Fuel Evidence, Too High, Learn Fuel Evidence + Too High, Believe Enough Fuel, Rationally Believe Enough Fuel…) is also very high.

But many authors have argued that this is the wrong result. Instead, something like the following seems more rational:

Always Suspend: Whenever you learn Too High, suspend judgment about the results of any calculations performed in the air.

Schoenfield argues that we can defend rules like **Always Suspend** by adopting a new way of evaluating plans. Instead of looking at which plan is best to *follow*, we should look at which plan is best to *make*. So instead of:

Execution: A plan is good insofar as it yields good results when it's actually executed.

We should adopt something like:

Attempt: A plan is good insofar as it yields good results when it's attempted.

Schoenfield's thought is that, because we don't always succeed in carrying out our plans, these modes of evaluation might diverge. Suppose the best plan to follow involves leaving the house at 10:15. For a chronically late person, however, it might be best to *make* the plan to leave at 10: that would result in leaving on time, whereas making the plan to leave at 10:15 would result in leaving too late.

In epistemic cases like **Hypoxia**, Schoenfield argues, the same thing can happen. If we learn a proposition like Too High, it will often turn out to be the case that we are actually hypoxic. If we try to conditionalize (or if we try to follow conditionalization*, **, or ***) we might well end up misjudging our evidence (for example, by misreading a dial or a gauge) or making a mistake in reasoning, ending up with a belief state that is not the one supported by conditionalization. Depending on the details, our conclusion might be very inaccurate: we might confidently judge that we have enough fuel when we do not, for instance.

Why does **Always Suspend** fare better? I take it the thought is that hypoxia, and the other sources of higher-order evidence commonly discussed in the literature, can cause *subtle* mistakes but not obvious ones. Of the evidence we have, we are much more likely to misjudge Fuel Evidence than we are to misjudge Too High. And as far as doing what the plan recommends, we're much more likely to make a mistake conditionalizing on our total evidence than we are simply suspending judgment. So while *trying* to follow the plan to conditionalize can yield vastly different results because of our impairment, *trying* to follow Always Suspend will yield just one result: suspension of judgment.

This is an interesting argument which leads us to a very different way of evaluating plans. But notice that it does not solve our main problem – quite the opposite. If we take something like **Attempt** to be our primary mode of evaluating plans, we will not salvage conditionalization (according to Schoenfield's argument), but in fact move farther from it. So while you might have objections to **Execution**, it won't help to replace it with **Attempt**.

4.5 Summing up

Let's review. I started off by observing that conditionalization is, intuitively, the right way to respond to new information. I then asked what "synchronic conditionalization" might look like. Is it a plan? If so, it seems like we should be able to argue that a plan to conditionalize is the best epistemic plan. And it looked like there would be an easy way to do that, starting with **First-person**, **Execution** and **Accuracy First**.

What I have argued, however, is that conditionalization is not the best epistemic plan. I first argued that if we focus on **Execution**, we will reach the conclusion that rather than conditionalizing on what we have learned, we should conditionalize on something stronger. On

the other hand, replacing **Execution** with **Attempt** doesn't help. Instead, this line of reasoning ends up with something weaker than conditionalization.

What should we conclude from this? One option would be to get rid of conditionalization altogether, and adopt whatever turns out to be the best plan (Conditionalization*** if you liked **Execution**, perhaps, or something else if you liked **Attempt**). A second option, which I prefer, is to get rid of planning as a way to support conditionalization.

5. Where does this leave us?

So what is conditionalization good for, if anything? What is its synchronic analog? And supposing we do find another way to defend conditionalization, where does the discussion of planning get us? Is there anything we can learn from that way of thinking? I will try to answer those questions in this final section.

5.1 A new defense of conditionalization

In the preceding discussion I argued that **Execution** was the root of the problem. But **First-person** is equally to blame. Putting those two assumptions together gives us a way to describe more generally where planning goes wrong. We evaluate our plans based on what we can predict will be the case when they come about. But when the time comes to follow through on our plans, we might not realize that we are in the situation we have planned for. We might *know less*, when the time comes, than we predicted ex ante. So what makes sense to do in the moment might be different from what we had planned for, or even could plan for.

The fact that conditionalization is a bad plan, then, doesn't mean that it's a bad response in the moment. In order to say what's good about conditionalization, we need a mode of evaluation that eliminates this first-personal, prospective outlook. And we need a synchronic analog to conditionalization which is naturally evaluated in this way.

I think the answer can be found by going back to the defenses of synchronic conditionalization. While many authors discuss plans, some also discuss *dispositions*. I will argue that the *disposition* to conditionalize is a defensible one, and that we can make sense of why it looks rational.

Why is this? While plans seem most paradigmatically evaluated from a first-personal, prospective point of view, dispositions afford more neutrality. They are more naturally evaluated from a third-personal, theoretical perspective, which gives us more leeway: we can define what makes for a good disposition according to our theoretical goals. To argue that conditionalization is a good disposition, then, we need to define those theoretical goals.

Let's go back to the original case, **Pet**, in which I was choosing between a green lizard and a red salamander. Compare two people in this case: Connie, who is disposed to conditionalize, and Donnie, who is not. Both Connie and Donnie have high conditional credence in my pet being green, given that it's a lizard. (Given the question: "How likely is Sophie's pet to be green, given that it's a lizard?" both will say "highly likely!" or "guaranteed!") But when they learn that my pet is a lizard, Connie and Donnie are disposed to adopt different belief states. Connie is disposed to become highly confident that my pet is green. But Donnie is disposed to become highly confident that my pet is blue.

Connie is doing better than Donnie. Why is that? Most obviously, Connie is disposed to adopt a belief state that is, by her own lights, accurate given what she's learned. Donnie is disposed to adopt a belief state that is, by his own lights, inaccurate. There is something coherent in Connie's disposition, but not Donnie's.

A harder question is: why should we think that Connie is doing better than Super-Connie, who is disposed to update by conditionalization***? After all, we (as theorists) know all of the extra bits of information about the state of the world in which these dispositions for learning are activated. We know that Super-Connie will be more accurate.

This is a more restricted version of a familiar question: what's the difference between rationality and omniscience? Conditionalization allows us to draw that line in a natural place. We only expect people to update their beliefs in response to what they have learned. Learning E is what triggers the disposition to change one's beliefs, and they should be changed in a way that coheres with one's earlier beliefs. One should not incorporate more than what one has learned, like Super-Connie. And one should not update in a way that looks inaccurate by one's own lights, like Donnie.[11]

5.2 A new application for the planning argument

I have spent a lot of time in this paper drawing out ways in which the planning interpretation goes wrong. This revealed some lessons about how to argue for conditionalization: since conditionalization involves reacting only to what we learn, we need to find a mode of evaluation that does not invite us to build in extra information. But beyond that lesson, is there anything we can learn from the line of argument we followed to Conditionalization***?

I want to end by suggesting that there is something we can learn. In particular, the reasoning we went through in thinking about planning, while not well-suited to a defense of conditionalization, is quite similar to what we ought to take into account when thinking about a rule for rational *deference*.

We defer to sources that we take to be experts. I often defer to my phone, for example, about today's weather. Supposing I *always* deferred to my phone, I would follow a rule like this:

Phone Deference: Cr(Rain|My phone says the chance of rain = n) = n

Of course, I don't always defer to my phone about the weather. Suppose I look at my phone *and* look out the window: in that case, my credence in Rain will often end up much higher or lower than the chance reported by my phone. In that scenario, I know more than my phone – or at least, I know things that screen off the evidence my phone has available.

[11] This argument very similar to with Zendejas Medina [2023]'s view, though he puts his view in terms of "conditional plans." He also writes that we *prefer* the conditional plan to conditionalize. I disagree: I think in many cases we prefer, or should prefer, that we update by conditionalization***. (See Zhang and Meehan [2025] for a more complicated argument that we should sometimes prefer not to conditionalize.) Zendejas Medina also writes that a plan is only "admissible" if "its condition is such that, after learning, a rational agent will implement the antecedently best actionable plan with that condition." (p. 6) This requirement would rule out conditionalization***, since no rational agent will actually update by conditionalization***. I do not want to rely on Zendejas Medina's admissibility requirement here: since I am most interested in synchronic analogues to conditionalization, I do not want to use a diachronic norm to argue for or against different views. And while I agree with many of Zendejas Medina's suggestions, I think they are more naturally seen as pertaining to dispositions. See p. 11, where Zendejas Medina introduces the example of poker player: here he writes in terms of plans, but the evaluation of the subject is all done third-personally with the aim of explaining why she acted in the way she did. I think it is much more apt to describe this scenario in terms of which dispositions we expect to see in a rational agent, rather than in terms of which plans the agent is required to carry out.

I might also defer to my future self. Since I expect my future self to be better informed, I should also expect her beliefs to be more accurate than my beliefs are now. Supposing I *always* deferred to my future self, I would follow:

Reflection:[12] $Cr(P|\text{my future credence in } P = n) = n$

But it doesn't always make sense to defer to my future self, either. Sometimes I expect to forget something I now know,[13] or to make rational mistakes.[14] I might, in the future, be unaware of my beliefs, or unaware of what I have learned.[15] I might, in the future, come to think that I'm not an expert about some topic – though I know now that I will be an expert.[16]

In all of these cases, there is an information gap between me and my future self. And unlike the normal case, where epistemic disparities favor my future self, all of these are examples where my current self is in a better epistemic position. I either know *more* than my future self, or else I am currently, for some other reason, more capable than my future self of forming accurate beliefs.

In response to these problems, Briggs [2009] argues for a modified version of Reflection, according to which I should defer to my future beliefs corrected for any anticipated errors. Briggs' principle is below. Where D_r is the difference between the credence you expect your future self to have, and the credence you would have if you conditionalized:

Distorted Reflection: $Cr_0(A|Cr_1(A) = r) = r - D_r$

In effect, this principle says that the problem with Reflection is that sometimes we don't conditionalize. So instead of deferring to what you expect your future self to actually believe, you should defer to the results of conditionalizing on whatever evidence you will have learned – what your future self *would* have believed, had she conditionalized.

Briggs' principle accounts for certain kinds of errors, such as forgetting things we now know, or adopting irrational beliefs under the influence of a drug. Both of these are examples of non-conditionalizing changes in belief, where we expect, ex ante, that conditionalizing will be most accurate. But if my arguments here are right, we can also sometimes expect, ex ante, that certain *departures* from conditionalization will be more accurate than conditionalizing would be.

It seems to me that the prospective, self-conscious mode of assessment we entered into to evaluate plans is very similar to how we ought to approach deference, including deference to our future self. Deferring to our future self, like planning, is forward-looking and first-personal. It therefore builds in much of the same information that I argued was built into planning. But with that further information in mind, we can anticipate certain ways in which conditionalizing will be inaccurate. For example, suppose we anticipate being in a situation in which we believe P, but don't know that we believe P. Deferring to our future self should not mean simply adopting whatever belief state our future self would have adopted, had we conditionalized; we should also adjust that deference to incorporate whatever extra information we know will be true under those circumstances.

---

[12] See van Fraassen [1984].
[13] See Talbott [1991].
[14] See Christensen [1991].
[15] See Williamson [2014], Elga [2013], and Lasonen-Aarnio [2015], among others.
[16] See Elga [2013].

This observation suggests a new use for the planning argument: while evaluating plans does not get us an argument for conditionalization, it might help us think about deference. Deference is a context in which we should want to incorporate whatever knowledge we have, from the prospective point of view, even if the expert to which we are deferring does not have that same information.

Conclusion

I have argued that we should not plan to conditionalize. Relatedly: if there is a defensible synchronic version of conditionalization, it is not a plan to conditionalize. This is because, as I have argued, plans are most naturally evaluated in a manner which builds in extra information – information which conditionalization itself disregards.

So is synchronic conditionalization defensible? If it is, I have argued that we will have better luck defending the *disposition* to conditionalize, rather than the *plan*, since dispositions need not look good from the subject's own prospective point of view. I have not spelled out the details of how this should go, but hopefully this discussion has suggested some limiting factors on what a suitable mode of evaluation should look like.

I have also argued that the way we evaluate plans is similar to, and might inform our thinking about, deference. But this means that deference and conditionalization come apart.

Both of these lines of thought also point to a conflict between conditionalization and a certain kind of internalism. Those who think that rationality is, centrally, something we should be able to prospectively endorse – for example, something we should plan for, or something we should defer to – may have reason to worry about conditionalization. And those who think we should conditionalize may have reason to worry about this flavor of internalism.

References

Briggs, R. A. and Pettigrew, Richard. [2020] "An Accuracy-Dominance Argument for Conditionalization." *Noûs* 54(1):162-181.

Briggs, R. A. [2009] "Distorted Reflection" *Philosophical Review* 118 (1):59-85.

Brown, Peter M. [1976] "Conditionalization and Expected Utility". *Philosophy of Science* 43(3): 425-19.

Carr, Jennifer. [2017] "Epistemic Utility Theory and the Aim of Belief." *Philosophy and Phenomenological Research*. Vol XCV No. 3. pp. 511–534.

Christensen, David. [1991] "Clever Bookies and Coherent Beliefs." *The Philosophical Review* 100 (2):229-247.

Easwaran, Kenny. [2013] “Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection.” *Philosophy of Science* 80 (1):119-142.

Elga, Adam. [2013] “The Puzzle of the Unmarked Clock and the New Rational Reflection Principle.” *Philosophical Studies* 164 (1): 127-139.

Greaves, Hilary. [2013] “Epistemic Decision Theory.” *Mind* 122(488):915-925.

Greaves, Hilary and Wallace, David. [2006] “Justifying Conditionalization: Conditionalization Maximizes Expected Accuracy.” *Mind* 115 (459): 607-632.

Konek, Jason and Levinstein, Benjamin. [2019] “The Foundations of Epistemic Decision Theory.” *Mind* 128(509): 69-107.

Lasonen-Aarnio, Maria. [2015] “New Rational Reflection and Internalism About Rationality.” *Oxford Studies in Epistemology* 5.

Lewis, David. [1999] “Why Conditionalize?” In *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Pettigrew, Richard. [2020] “What is conditionalization, and why should we do it?” *Philosophical Studies* 177:3427-3463.

– [2016] *Accuracy and the Laws of Credence.* Oxford: Oxford University Press.

Schoenfield, Miriam. [2017] “Conditionalization does not (in general) maximize expected accuracy.” *Mind* 126(504): 1155-1187.

– [2015] “Bridging Rationality and Accuracy”. *The Journal of Philosophy* 112(12): 633-657.

Talbott, William. [1991] “Two Principles of Bayesian Epistemology.” *Philosophical Studies* 62 (2):135-150.

van Fraassen, Bas. [1984] “Belief and the Will”, *Journal of Philosophy*, 8: 235–256.

Williamson, Timothy. [2014] “Very Improbable Knowing”. *Erkenntnis* 79 (5):971-999.

Zendejas Medina, Pablo. [2023] “Just As Planned: Bayesianism, Externalism, and Plan Coherence.” *Philosophers' Imprint* 23.

Meehan, Alexander & Zhang, Snow. [2025] “Bayes Is Back”. *Philosophical Review* 134 (3):285-350.