Updating for Externalists

J. Dmitri Gallow

Princeton University · April 29th, 2019

1 Externalism

Internalism
If e is your total evidence, then your evidence must tell you that e is your total evidence.

$$\Box(\mathbf{T}e \rightarrow \mathbf{ET}e)$$

Externalism
Your total evidence may be e without your evidence telling you that e is your total evidence.

$$\Diamond(\mathbf{T}e \wedge \neg\mathbf{ET}e)$$

- ▷ $\ulcorner\mathbf{E}e\urcorner \stackrel{\text{def}}{=}$ your evidence says (at least) e
- ▷ $\ulcorner\mathbf{T}e\urcorner \stackrel{\text{def}}{=}$ your evidence tells you e and no more (e is your *total* evidence)

1. Some reasons to be interested in externalism:
   (a) Externalism allows that your evidence may not tell you what evidence you possess. Given *evidentialism*, this means that your evidence may not tell you whether you are rational.
   (b) It has therefore played a starring role in debates about the rationality of epistemic *akrasia* and peer disagreement.
2. Prominent externalists have thought that:
   - ▷ epistemic akrasia *can* be rational; and
   - ▷ if you are rational and a disagreeing peer with the same evidence irrational, then you should not *conciliate*.
3. My goal: develop a general theory of how externalists should learn from their evidence.
   (a) This theory will be motivated by the thought that it is rational to aim at *accurate* beliefs.
   (b) In the end, this theory will lead the externalist to different positions on epistemic *akrasia* and peer disagreement.
4. Assume a Kripke semantics for **E** and **T**.[1] Assume that evidence is consistent.[2] Then, internalism is equivalent to the conjunction of *Positive Access* and *Negative Access*:
   - ▷ *Positive Access*: if your evidence tells you e, then your evidence must tell you that it tells you e: $\Box(\mathbf{E}e \rightarrow \mathbf{EE}e)$.
   - ▷ *Negative Access*: if your evidence *doesn't* tell you e, then your evidence must tell you that it *doesn't* tell you e: $\Box(\neg\mathbf{E}e \rightarrow \mathbf{E}\neg\mathbf{E}e)$.
5. A Williamsonian argument against *Positive Access*:[3] Suppose you catch a brief glimpse of an "irritatingly austere" clock. Then, if you accept (P1) and (P2), you must reject *Positive Access*.

[1] For any world w in a Kripke model, $\ulcorner\mathbf{E}e\urcorner$ is true at w iff $\ulcorner e\urcorner$ is true at all worlds accessible from w. And $\ulcorner\mathbf{T}e\urcorner$ is true at w iff $\ulcorner e\urcorner$ is true at *all and only* worlds accessible from w.

[2] That is: suppose that, if your evidence says that e, then it must not also say that $\neg e$, $\Box(\mathbf{E}e \rightarrow \neg\mathbf{E}\neg e)$. This means assuming that the accessibility relation is *serial*.

[3] *Cf.* Williamson (2000, 2011).

P1) The most your evidence tells you about the position of the clock hand that it lies in some interval $[a, b]$, with $a < b$.

P2) Your evidence tells you that: if the clock hand is located at b, then you won't get the evidence that it's located no further than b.

$$\mathbf{E}[H = b \rightarrow \neg\mathbf{E}(H \leqslant b)]$$

C) *Positive Access* is false.

$$\Diamond(\mathbf{E}e \wedge \neg\mathbf{E}\mathbf{E}e)$$

Proof. Assume (P1) and (P2). Then, from *Positive Access*, (A1), we derive a contradiction:

A1) $\mathbf{E}e \rightarrow \mathbf{E}\mathbf{E}e$

A2) By (P2) and contraposition: $\mathbf{E}[\mathbf{E}(H \leqslant b) \rightarrow H \neq b]$.

A3) By (A2) and the K-axiom: $\mathbf{E}\mathbf{E}(H \leqslant b) \rightarrow \mathbf{E}(H \neq b)$.

A4) By (P1): $\mathbf{E}(H \leqslant b)$.

A5) By (A4) and (A1): $\mathbf{E}\mathbf{E}(H \leqslant b)$.

A6) By (A3) and (A5): $\mathbf{E}(H \neq b)$.

But (A6) contradicts (P1), which told us that $H \in [a, b]$ was the *most* your evidence told you about the position of the clock hand. So, if (P1) and (P2) are true, *Positive Access* is false. □

6. Since internalism entails *Positive Access*, and externalism is the negation of internalism, this argument (if successful) establishes externalism.

7. I'll focus on a simplified model of Williamson's 'irritatingly austere' clock. The clock hand could be in one of four positions, and your total evidence will be that it is *not* at the position opposite its actual position. (See figure 1.)

2 Learning

8. I'll assume that you have opinions about how likely various propositions are, and that these opinions can be represented with a *credence* function, C, from propositions to real numbers between 0 and 1.[4]

[4] I assume throughout that your credence function C is a probability.

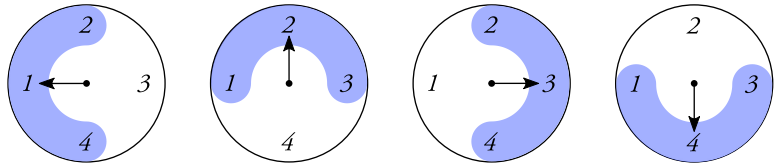


Figure 1: A simplified model of Williamson's clock. The clock hand could point at position 1, 2, 3, or 4. If it points at 1, your evidence will be that it's not at 3, and similarly for the other possible positions.

- $C(p)$ represents how likely you think the proposition p is.

9. I will also assume that you have *learning dispositions* to update your credences in light of the evidence.

- Let's represent these dispositions with a function, D, from evidence, e, to new credence functions, D_e
- D_e (the value of D, given the argument e) is the credence function you are disposed to adopt if your total evidence is e.
- You have the dispositions represented by D iff, for each e, you are disposed to manifest the *response* of adopting D_e in the *stimulus* condition $\mathbf{T}e$. And, let's suppose, you manifest this response at *all* possibilities in which $\mathbf{T}e$.

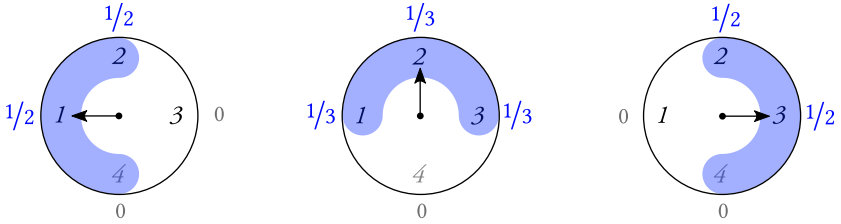
10. How should you be disposed to respond to your evidence? The orthodox Bayesian answer is: you should be disposed to *condition* on your total evidence.

> Conditionalization
> Be disposed to respond to the evidence e by adopting your current credence function, C, *conditioned on* e.
>
> $$D_e(p) = C(p \mid e) \qquad \text{(condi)}$$

11. I think the externalist sould reject condi, for at least two reasons:

(a) externalist conditionalizers must accept the rationality of deliberately *biased inquiry*

(b) the pursuit of accuracy will lead an externalist to violate condi



(a) **T**¬3 (b) **T**¬4 (c) **T**¬1

Figure 2: In figure 2a, the credences CONDI says you should be disposed to adopt upon learning that ¬3 (and no more). In figure 2b, the credences CONDI says you should be disposed to adopt upon learning that ¬4 (and no more). And, in figure 2c, the credences CONDI says you should be disposed to adopt upon learning that ¬1 (and no more).

2.1 Biased Inquiry

12. Suppose you're about to catch a glimpse of the clock, and a reliable confidant tells you that the clock hand is *not* at position 4. Then, you know that you won't learn that it's not at 2. So you'll either learn ¬3, ¬4, or ¬1. If you start off thinking that the clock hand is equally likely to be at positions 1, 2, and 3, then the learning dispositions recommended by CONDI are shown in figure 2. Notice that:

    (a) Your credence that 2 may rise, but definitely *won't* fall.

    (b) Moreover, your credence that 2 will only rise if 2 is *false*—and this is something you are capable of recognizing in advance of looking.

13. Salow (2018) twists the knife: if these are your learning dispositions, then it can be rational for you to be disposed to become more confident of any falsehood you wish.

    (a) Let p be the proposition that you are popular.

    (b) Have a friend who knows the truth about p place the clock hand at 2 iff p is true—else, flip a coin to decide whether to place it at 1 or 3.

    (c) Then, take a quick glimpse, and CONDI will say: it is rational for you to become more confident that p, so long as p is false.

    (d) There's no reason you can only do this once. Do it again, and again, and again, and you can get as confident that p as you wish—so long as p is false.

14. It's very difficult to see this as rational inquiry. Let's lay this down as a principle:

    > No Biased Inquiry
    > If you are disposed to raise your credence that p in response to some potential evidence, then you must also be disposed to lower your credence that p in response to some potential evidence.

15. What Salow shows is this: if cases like Williamson's clock are possible, then

    - ▷ Externalism;
    - ▷ Conditionalization; and
    - ▷ No Biased Inqiury

    are inconsistent. Salow (2018) recommends rejecting externalism. Perhaps that's the right lesson. But I think there's a plausible version of externalism left standing which accepts No Biased Inquiry while rejecting Conditionalization.

16. One final observation: the reasons we have to accept No Biased Inquiry also give us reason to accept:

    > Reflection
    > You shouldn't expect your new credence that p to be higher or lower than your current credence that p.
    >
    > $$\sum_e D_e(p) \cdot C(\mathbf{U}e) = C(p)$$

    (a) Here, I use $\ulcorner \mathbf{U}e \urcorner$ for $\ulcorner$ you have *updated* your credences to $D_e \urcorner$.

2.2 Accuracy

17. Take some measure of the *accuracy* of a credence function C at a world w, $\mathscr{A}(C, w)$. I'll assume that $\mathscr{A}$ is 'well-behaved'—where this is a technical term

which means that $\mathscr{A}$ is *strictly proper*, *additive*, and *extensional* (the terms are explained in this[5] footnote.) All accuracy measures which have been taken seriously in the literature are well-behaved, in this sense.

(a) Then, we may ask: which learning dispositions do you *expect* to be most accurate?[6]

(b) I will make the normative assumption that learning dispositions may be evaluated in terms of their expected accuracy, and that they are rational if they *maximize* expected accuracy.

18. If internalism is correct, then the learning dispositions which maximize expected accuracy are the ones that conform to CONDI.[7]

19. On the other hand, if *externalism* is correct, then the learning dispositions which maximize expected accuracy are the ones that conform to CONDI*.[8]

> CONDITIONALIZATION*
> Be disposed to respond to the total evidence e by adopting your current credence function, C, conditioned on $\mathbf{T}e$.
>
> $$D_e(p) = C(p \mid \mathbf{T}e) \qquad \text{(CONDI}^*\text{)}$$

20. Externalists should also want to accept *certainty* externalism.

> CERTAINTY EXTERNALISM
> Your total evidence may be e without it being rational for you to be certain that your total evidence is e.

(a) The reasons we have for endorsing externalism (*e.g.*, the Williamsonian argument) also apply *mutatis mutandis* to certainty externalism.

[5] $\mathscr{A}$ is *strictly proper* iff, for every probabilistic credence function P, the unique credence function C which maximizes $\sum_w P(w) \cdot \mathscr{A}(C, w)$ is P itself. $\mathscr{A}$ is *additive* iff it is of the form $\mathscr{A}(C, w) = \sum_p \mathscr{A}(C(p), p, w)$, for some function $\mathscr{A}(x, p, w)$, of the accuracy of a credence x in proposition p in world w. It is *extensional* iff there are functions $\mathscr{A}_1$ and $\mathscr{A}_0$ such that $\mathscr{A}(x, p, w)$ is $\mathscr{A}_1(x)$ if $w \in p$ and $\mathscr{A}(x, p, w)$ is $\mathscr{A}_0(x)$ if $w \notin p$.

[6] This expectation is given by: $\sum_e \sum_{w \in \mathbf{T}e} C(w) \cdot \mathscr{A}(D_e, w)$.

[7] This is shown in Theorem 2 of GREAVES & WALLACE (2006). Here, I am also assuming that evidence is *factive*—so that, if $\mathbf{E}e$, then e must be true.

[8] This is shown in SCHOENFIELD (2017).

	1($\mathbf{T}\neg 3$)	2($\mathbf{T}\neg 4$)	3($\mathbf{T}\neg 1$)	4($\mathbf{T}\neg 2$)
$\mathbf{U}\neg 3$	8/40	1/40	○	1/40
$\mathbf{U}\neg 4$	1/40	8/40	1/40	○
$\mathbf{U}\neg 1$	○	1/40	8/40	1/40
$\mathbf{U}\neg 2$	1/40	○	1/40	8/40
	1/4	1/4	1/4	1/4

Table 1: A credence distribution for our simplified model of Williamson's clock in which we allow that your learning dispositions may *misfire*. (Recall: $\ulcorner \mathbf{U}e \urcorner$ says that you have *updated* your credences to D_e.)

21. However, CONDI* is inconsistent with certainty externalism. It does not allow you to be less than certain of what your total evidence is.

22. So: if the externalist adopts learning dispositions which maximize expected accuracy, then they cannot be uncertain about what their evidence says.

▷ Since an externalist should want to allow that you can be rationally uncertain about what your evidence says—since they should want to be a *certainty* externalist—this could be seen as an argument against externalism.

23. I have a suggestion for how an externalist can respond:

(a) By assuming that you adopt the new credences D_e in *every* possibility in which your total evidence is e, we presupposed that you take your dispositions to respond to evidence to be *perfect*.

(b) The externalist should deny this—they should believe that your learning dispositions may 'misfire'. (That is, they should say that you foresee the possibility of responding as if your evidence were $f \neq e$, when in fact your evidence is e.)[9]

24. If you foresee the possibility of your learning dispositions misfiring, then we should enrich our model of Williamson's clock to include these possibilities. (See table 1.)

[9] *Cf.* SCHOENFIELD (2015) and STEEL (2018).

	1(**T**¬3)	2(**T**¬4)	3(**T**¬1)	4(**T**¬2)
U¬3	8/100	8/100	0	0
U¬4	1/100	64/100	1/100	0
U¬1	0	8/100	8/100	0
U¬2	1/100	0	1/100	0
	1/10	8/10	1/10	0

Table 2: The result of updating the distribution from table 1 on the evidence ¬4 with EXCONDI.

25. If our measure of accuracy is well-behaved, then the (potentially misfiring) learning dispositions with maximal expected accuracy are those which conform to EXCONDI:[10]

> EXTERNALIST CONDITIONALIZATION
> Be disposed to respond to the evidence e by changing your credence in $\mathbf{T}f$ to your current credence in $\mathbf{T}f$, conditional on $\mathbf{U}e$, $C(\mathbf{T}f \mid \mathbf{U}e)$, and holding fixed your credence in each proposition conditional on $\mathbf{T}f$ (for each f which might be your evidence).

$$D_e(p) = \sum_f C(p \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e) \qquad \text{(EXCONDI)}$$

   (a) The result of updating the credence distribution in table 1 on ¬4 with EXCONDI is shown in table 2.

26. Unlike CONDI*, EXCONDI allows uncertainty about what your total evidence is. So it is consistent with certainty externalism.

   (a) For instance, in our model of Williamson's clock, after you've learned that the clock hand isn't at position 4, you will think it's 20% likely that your total evidence was ¬3 or ¬1 instead.

   (b) So: EXCONDI says that you should be uncertain about what your total evidence is.

   (c) So: EXCONDI is a more externalist-friendly norm than CONDI*.

27. Also: EXCONDI entails the principle of Reflection.[11]

   (a) So: an ex-conditionalizer will not be capable of engaging in intentionally biased inquiry.

3 Applications

3.1 Epistemic Akrasia

28. (a) Suppose that you have evidence, e, which supports believing it will rain.

   (b) Then, you get some *new* evidence, e^*, which supports believing that your belief in rain was likely irrational.

   (c) Lasonen-Aarnio (2015, forthcoming): your total new evidence, $e \cap e^*$, supports believing it will rain *and* that it's irrational to believe that it will rain.

29. (a) Elga contends that some forms of epistemic *akrasia* are irrational. He defends a principle which says (roughly): reason to think that D_f are the rational credences is reason to move your credences towards D_f.

   (b) More carefully:

> NEW RATIONAL REFLECTION
> Conditional on D_f being the rational credences for you to hold, your credences should agree with D_f, once D_f is informed that it is rational.

$$D_e(p \mid D_f \text{ is rational }) = D_f(p \mid D_f \text{ is rational })$$

[10] I measure the expected accuracy of potentially misfiring learning dispositions with: $\sum_e \sum_{w \in \mathbf{T}e} C(w) \cdot \sum_f C(\mathbf{U}f \mid \mathbf{T}e) \cdot \mathscr{A}(D_f, w)$.

[11] *Proof*: According to EXCONDI, you should have:

$$\begin{aligned}
\sum_e D_e(p) \cdot C(\mathbf{U}e) &= \sum_e \sum_f C(p \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e) \cdot C(\mathbf{U}e) \\
&= \sum_f C(p \mid \mathbf{T}f) \sum_e C(\mathbf{T}f \mid \mathbf{U}e) \cdot C(\mathbf{U}e) \\
&= \sum_f C(p \mid \mathbf{T}f) \cdot C(\mathbf{T}f) \\
&= C(p)
\end{aligned}$$

$$D_e(p \mid \mathbf{T}f) = D_f(p \mid \mathbf{T}f)$$

30. Lasonen-Aarnio rejects New Rational Reflection.

    ▷ She additionally provides a counterexample (though this counterexample presupposes CONDI).

31. EXCONDI sides with Elga: an ex-conditionalizer will always satisfy New Rational Reflection.[12]

3.2 Peer Disagreement

32. (a) Suppose you and an epistemic peer (let them be an identical clone of you) both catch the same glimpse of the clock, and both of you receive the evidence that it's not at position 4.

    (b) You correctly respond to the evidence, and think it's 80% likely that it's at position 2.

    (c) You then discover that your peer thinks it is 80% likely to be at position 1.

    (d) The *Right Reasons* view says that this does not give you a reason to revise your opinion.

    (e) In contrast, *Conciliationism* says that it does.

33. (a) In this case, at least, EXCONDI sides with the conciliationist.

    (b) Let's suppose that your peer's learning dispositions are just as likely to misfire as yours, and that whether/how your learning dispositions misfire is independent of whether/how your peer's do.

    (c) Then: if you begin with the credences shown in table 2, and update on the fact that your peer updated on ¬3, then you'll end up with the credences shown in table 3.[13]

    (d) So: according to EXCONDI, in this case at least, you should see the disagreement of your peer as a reason to revise your views about the position of the clock hand.[14]

[12] If you update with EXCONDI, then both $D_e(p \mid \mathbf{T}f)$ and $D_f(p \mid \mathbf{T}f)$ will be equal to $C(p \mid \mathbf{T}f)$.

[13] I'm supposing that you're certain to correctly learn what your peer's credences are—so there's no possibility of your learning dispositions misfiring, and EXCONDI and CONDI agree.

[14] For a similar justification of conciliationism, see SCHOENFIELD (2018) and STEEL (2018).

	1(**T**¬3)	2(**T**¬4)	3(**T**¬1)	4(**T**¬2)
U¬3	8/20	1/20	0	0
U¬4	1/20	8/20	0	0
U¬1	0	1/20	0	0
U¬2	1/20	0	0	0
	1/2	1/2	0	0

Table 3: The result of updating the distribution from table 2 on the evidence that your peer updated on ¬3.

4 In Summation

34. (a) My goal was to say something about how an externalist should be disposed to revise their opinions in light of their evidence.

    (b) I began by assuming that learning dispositions which maximize expected accuracy are rational.

    (c) However, if your learning dispositions are perfect, then maximizing expected accuracy won't allow you to be uncertain about what your evidence says.

    (d) But externalists think that it's not always rational to be certain what you should be certain of.

    (e) I suggested that the externalist permit a kind of rational modesty—you may foresee the possibility of your learning dispositions 'misfiring', and you mistaking your evidence.

    (f) If we allow this kind of modesty, then the learning dispositions which maximize expected accuracy will be the ones conforming to *Externalist conditionalization*.

35. Externalist conditionalization will always satisfy the principle of *Reflection*, so it will not permit intentionally biased inquiry.

36. Externalist conditionalization entails Elga's enkratic requirement *New Rational Reflection*.

37. In some cases of peer disagreement, externalist conditionalization will counsel conciliation.

References

Elga, Adam. 2013. "The puzzle of the unmarked clock and the new rational reflection principle." *Philosophical Studies*, vol. 164: 127–139. [5]

Greaves, Hilary & David Wallace. 2006. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind*, vol. 115 (495): 607–632. [4]

Hild, Matthias. 1998. "Auto-Epistemology and Updating." *Philosophical Studies*, vol. 92: 321–361.

Lasonen-Aarnio, Maria. 2014. "Higher Order Evidence and the Limits of Defeat." *Philosophy and Phenomenological Research*, vol. 88 (2): 314–345.

—. 2015. "New Rational Reflection and Internalism about Rationality." In *Oxford Studies in Epistemology*, Tamar Szabó Gendler & John Hawthorne, editors, vol. 5, chap. 5. Oxford University Press, Oxford. [5]

—. forthcoming. "Enkrasia or Evidentialism? Learning to Love Mismatch." *Philosophical Studies*. [5]

Salow, Bernhard. 2018. "The Externalist's Guide to Fishing for Compliments." *Mind*, vol. 127 (507): 691–728. [3]

Schoenfield, Miriam. 2015. "Bridging Rationality and Accuracy." *Journal of Philosophy*, vol. 112 (12): 633–657. [4]

—. 2017. "Conditionalization does not (in general) Maximize Expected Accuracy." *Mind*, vol. 126 (504): 1155–1187. [4]

—. 2018. "An Accuracy Based Approach to Higher Order Evidence." *Philosophy and Phenomenological Research*, vol. 96 (3): 690–715. [6]

Steel, Robert. 2018. "Anticipating Failure and Avoiding It." *Philosophers' Imprint*, vol. 18 (13). [4], [6]

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press, Oxford. [1]

—. 2011. "Improbable Knowing." In *Evidentialism and its Discontents*, T. Dougherty, editor. Oxford University Press, Oxford. [1]