Penultimate draft of April 2024—please cite/quote from the published version.

Knowledge of One's Own Credences

T. Parent (Nazarbayev University)
nontology@gmail.com

0. Introduction

This paper has two parts: Part I discusses a problem concerning subjective probabilities. Part II outlines a partial solution to the problem, mainly by defending a kind of "transparency" thesis concerning knowledge of one's own judgments.[1] Part I will primarily be of interest to those who take seriously the notion of subjective probability. (This need not imply that it is the *most important* notion, but rather just that it is worthwhile.) Part II will appeal primarily to those interested in knowledge about one's own mental states. But hopefully, those who are attracted to Part I will be drawn into Part II as well, given that latter bears on the problem from the former. Indeed, the main message will be that the theory of subjective probability can benefit substantially by adopting a transparency view of self-knowledge.

Part I

1. Hume's Regress

In the *Treatise of Human Nature*, Hume offers the following remarks:

> [E]very judgment, which we can form concerning probability…we ought always…to correct…by another judgment, deriv'd from the nature of the sceptical understanding… Here then arises a new species of probability to correct and regulate the first, and fix its just standard and proportion... [But] tho' it shou d be favourable to our preceding judgment, [it] must weaken still further our first evidence, and must itself be weaken'd by a fourth doubt of the same kind, and so on *in infinitum* ; till at last there remain nothing of the original probability, however great we may suppose it to have been, and however small the diminution by every new uncertainty…[E]ven the vastest quantity, which can enter into human imagination, must in this manner be reduc'd to nothing. (pp. 181–182).

[1] The transparency thesis is in the spirit of Evans (1982), and its defense is largely borrowed from chapter 8 of my (2017) book. However, the defense has been refined and made more digestible. And so, part II of this paper also serves to make more accessible the central material from chapter 8.

I shall discuss this argument in relation to subjective probabilities, a.k.a. credences or confidence levels.[2] We assess our confidence level in a proposition, but doubts arise about our assessment, and those doubts end up diminishing that confidence level. Suppose I begin with a 99% credence that evolution by natural selection is real, and I judge as much of myself. But suppose doubt causes me to be only 99% confident that I am 99% confident. Then, it seems my confidence in evolution ends up being less than 99%. For I am effectively hedging that I am 99% confident. So "the skeptical understanding" has started to erode my confidence in the way Hume suggested.[3]

This can be described in a more exact way by supposing that I also have a 1% credence in my first-order credence being 98% instead of 99%. (For simplicity's sake, assume all other possible second-order credences are 0.) Now a standard conditionalization rule says that, where C_t is the agent's credence function at a time t, for any p and q:

$$C_{t+1}(p) = [C_{t+1}(q) \times C_t(p/q)] + [C_{t+1}(\sim q) \times C_t(p/\sim q\ )]$$

Roughly, this suggests that my credence in evolution should be set to my credence in evolution being real when I have a 99% first-order credence *or* evolution being real when I do not have a 99% credence (which by assumption, means I have a 98% credence instead). So my credence in evolution should be set to $(.99 \times .99) + (.01 \times .98) = .9801 + .0098 = .9899$. Thus, if my first-order credence at t is .99, my second-order credences at t+1 push it down to .9899. Moreover, that the number can be diminished further—by means of third-order credences, fourth-order credences, and so on.

Two clarifications. First, when Hume suggests that the credence for p will be reduced to "nothing", this should not imply that the process will result in a credence of 0. After all,

[2] Some writers distinguish confidence levels from subjective probabilities and/or credences. Williamson (2000) also distinguishes credences from evidential probabilities (but see Eder 2023 for criticism). We shall not bother with these subtleties here.

[3] The issue can also be raised in terms of imprecise credences, where my confidence level is represented by an interval rather than a specific number. But for convenience, we shall work with precise credences.

this would yield certainty in ~p, and that is surely not what Hume had in mind. But fortunately, this is not a paper in Hume exegesis. For our purposes, we can just say that if Humean erosion of a credence continues to a concerning degree, then that is concerning. An erosion from .99 to .9899 per se may not be worrisome, but since the process can force the credence down much lower, we need to say something to prevent this.

Second clarification: the issue as described thus far would not pertain to an ideally rational person. Suppose an ideal agent starts with a 90% credence in the hypothesis of evolution (henceforth, "h_e"). Once higher-order doubt is introduced, she would instead distribute second-order credences in something like the following manner: she has .9 credence in having .9 credence in h_e; she has .05 credence in having .85 credence in h_e, she has .05 credence in having credence .95 in h_e—and she has 0 credence in having any other credence. (For convenience, assume that this agent can use only intervals of .05 for different credences.) Then, per conditionalization:

$$C_{t+1}(h_e) = (.9 \times .9) + (.05 \times .85) + (.05 \times .95) = .81 + .0425 + .0475 = \mathbf{.9}$$

So in this case, the first-order credence at t+1 is 90%, which is exactly what it was before. No erosion here.

The explanation is that our ideal agent's uncertainty weighs equally the possibility that she is overestimating *and* the possibility that she is underestimating the first-order credence. In contrast, Hume treats higher-order uncertainty as concerning only the possibility of overestimating. But uncertainty *per se* suggests that the agent could also be misjudging in the other direction. Once that is respected, second-order credences do not result in a lower first-order credence.

Nonetheless, a reasonable human agent might consider only the possibility of overestimating, at least in some instances.[4] The hypothesis of evolution may be a case in point. Suppose I am as confident as is psychologically possible in h_e without being certain. (For what it's worth, this is not altogether implausible.) Even so, suppose I am not certain whether my confidence is that high. I know that my credence is not 1; I know I am not *absolutely certain* in h_e like I am that $7 + 5 = 12$.[5] Yet there remains doubt on whether I am otherwise as confident as can be. Then, given that certainty is ruled out, my doubt must concern *only* the possibility that I am overestimating my confidence.

However, this is what gives rise to Humean erosion. It is crucial, however, that it can make sense to say that I am "as confident as possible without being certain". Normally, such talk does not make sense in describing an ideal agent, for normally, an ideal agent has a continuum of credence values approaching 1. On the other hand, it is quite plausible that human beings are psychologically limited in how fine grained our credences can be.[6] For convenience, continue to assume that .05 is the interval between different credences (though .05 is chosen arbitrarily—we could make our arguments below using any other credence intervals). Then, in describing myself as as confident as possible without being certain, assume that this describes me as having a credence of 95%.

---

[4] Many have regarded Hume's regress argument as a bad argument; for recent discussion, see Atkinson & Peijnenburg (2020). Yet as far as I can tell, it has gone unnoticed that in some cases, doubt about overestimating a credence can be reasonable, even when there is no doubt about underestimating. (Atkinson & Peijnenburg bypass this by instead considering an agent who doubts that her credence is *at least* 95%. But this seems not to be an issue about credences as such, but rather beliefs about *the lower bound* of a credence.)

[5] I will assume throughout that I can know with certainty that a lower-order credence is uncertain. I take to be defensible: In the spirit of Descartes, if I can doubt whether I am certain, then certainly I am not certain.

[6] Mike Titelbaum (in conversation) qualifies this by noting cases where people differentiate between extremely fine-grained probabilities. If A buys one Powerball ticket, and B buys ten, we can perceive that B is more likely to win. That is so, even if the chances of winning are 1 in 300,000,000. See also Titelbaum (2024). However, Titelbaum grants that our sensitivity to different probabilities surely must have limits, and that is all that I require. (Actually, the lottery case might suggest that A can differentiate between *arbitrarily* fine-grained probabilities, given that the chances of winning could be made arbitrarily small. This suggests that the example may not quite show what it may seem to show. But I cannot delve into this here.)

Again, I am not certain about my credence in h_e. But suppose I remain optimistic and I have .95 credence in a 95% first-order credence in h_e. In consequence, and to maintain probabilistic coherence, suppose I also have a .05 credence in a 90% first-order credence. (Assume all other second-order credences are 0.)

Suppose now that 'h_0' and 'h_1' stand in, respectively, for 'I have a 95% confidence level in h_e', and 'I have an 90% confidence level in h_e'. Then (where references to times are suppressed to reduce clutter):

$$C(h_e) = (C(h_0) \times C(h_e/h_0)) + (C(h_1) \times C(h_e/h_1))$$

$$C_{t+1}(h_e) = (.95 \times .95) + (.05 \times .9) = .9025 + .045 = \mathbf{.9475}$$

It looks like my second-order uncertainty has eroded my first-order credence—but if my psychologically-real credences are at .05 intervals, then there is no such thing as having a credence of .9475. Thus, our hypothesis about actual human credences would suggest we round the number up to .95, which is the first-order credence we started with.

Yet this does not give much comfort, for higher-order doubts can iterate, as Hume indicated. In particular, notice above that $C(h_1) = .95$ was treated as a *given*; it was effectively assumed that I have a third-order credence of 1 in $C(h_0) = .95$. But suppose I doubt that, as is reasonable. Then, here too, my doubt will concern only the possibility of overestimation. Underestimation is again irrelevant, not because I doubt that I am as confident as possible (short of certainty). It is rather because I doubt that I have a third-order credence of 1. And if a third-order credence of 1 is incorrect, then it *must* be an overestimation!

In fact, on reflection I am certain that my third-order credence is not 1. If I legitimately doubt my ability to assess a lower-order credence, then the same doubts apply here. Nonetheless, suppose I optimistically assign a third-order credence of .95 in $C(h_0) = .95$. Though in fact, to ensure probabilistic coherence, .95 should be the credence of the *conjunction* of $C(h_0) = .95$ and $C(h_1) = .05$; after all, it should turn out that, necessarily, both

conjuncts are true or both are false. So let h_2 be the hypothesis that $C(h_0) = .95$ and $C(h_1) = .05$, and assume that $C(h_2) = .95$.

Also, for probabilistic coherence, I should assign credences that sum to .05 to one or more alternative hypotheses to h_2. But assuming I have credence intervals of .05, and $C(h_2) = .95$, this only allows for one alternative hypothesis. So, let h_3 be the hypothesis that $C(h_0) = .9$ and $C(h_1) = .1$, and assume that $C(h_3) = .05$.

It deserves emphasis that these new third-order credences are introduced *only* because of the doubt that $C(C(h_0) =.95)=1$ is an overestimation. Accordingly, these new credences end up hedging my second-order credences in a way that diminishes them. And diminished second-order credences will, in turn, diminish my first-order credence.[7]

In detail, consider again the claim from the earlier calculation that:

$$C(h_e) = (C(h_0) \times C(h_e/h_0)) + (C(h_1) \times C(h_e/h_1))$$

We now apply conditionalization to $C(h_0)$ and to $C(h_1)$ as they occur in this claim, in order to incorporate our new hypotheses about second-order credences. (N.B., for concision's sake, the formula below exploits that in each of h_2 and h_3, the relevant conjunction has the same credence as its individual conjuncts. Again, this is because in each case, necessarily, the two conjuncts are both true or the two conjuncts are both false.)

[7] Weirdly, no erosion need result if my higher-order credences are *less* favorable to my initial credence in evolution. For example, suppose that $C(h_0) = .9$ rather than $C(h_0) = .95$. Then, I would be able to consider the hypothesis that $C(h_0) = .9$ is an overestimation *and* the hypothesis that $C(h_0) = .9$ is an underestimation. That sort of "balance of doubt" prevents erosion. So the skeptical case is limited to situations where my higher-order credences downplay the possibility of underestimating, e.g., when my higher-order credences are repeatedly at .95 (and certainty is ruled out). This means that as things are currently formulated, erosion will not be a threat in a number of situations. Still, it strikes me as paradoxical that there can be one sort of case where Humean erosion results, where doubts about overestimation (and not underestimation) are reasonable. Moreover, given that actual humans have an upper and a lower bound on credences between 0 and 1, other cases can be constructed where the "overestimating hypotheses" disjointly have a greater probability than the "underestimating hypotheses"—and reasonably so. (I have also recently discovered a version of the problem that extends to more kinds of cases; however, I will have to wait until another occasion to present it.)

$$C(h_e) = [(C(h_2) \times C(h_0/h_2)) \times C(h_e/h_0)] + [(C(h_2) \times C(h_1/h_2)) \times C(h_e/h_1)] +$$

$$[(C(h_3) \times C(h_0/h_3)) \times C(h_e/h_0)] + [(C(h_3) \times C(h_1/h_3)) \times C(h_e/h_1)]$$

Plugging in the numbers:

$$C(h_e) = [(.95 \times .95) \times .95] + [(.95 \times .05) \times .9)] + [(.05 \times .9) \times .95] + [(.05 \times .1) \times .9)]$$

$$C(h_e) = [.9025 \times .95] + [.0475 \times .9] + [.045 \times .95] + [.005 \times .9]$$

$$C(h_e) = .857375 + .04275 + .04275 + .0045 = \mathbf{.94375}$$

This still rounds up to .95 but the raw value is less than before, and we can see the writing on the wall. We may observe that, in this latest calculation, $C(h_2) = .95$ is treated as if it had credence 1, and that is certainly not true. We can thus expand this and other terms, via the conditionalization rule, in the way that we did for $C(h_0)$ and $C_t(h_1)$, as a means of hedging this and other third-order credences using fourth-order credences. Thus, in lieu of certainty, we can assign 95% credence to $C(h_2) = .95$ and only 5% to some alternative credence. But then we notice that this newest 95% credence is being treated as if it had credence 1—and so we can continue the 95–5 split between a credence and some alternative on up the hierarchy.

In this way, we arrive at a first-order credence that falls below .95, below .9, below .85… In fact, we need not envision the doubts arising *ad infinitum*; arguably, that would be a bit of psychological fantasy. But if the doubts iterate enough to force my credence in evolution below .75, that is jarring enough, I assure you.

The significance of the problem is that it supports a kind of skepticism. Hume was wrong if he thought the problem reduces a credence to "nothing". But apparently, higher-order doubt can force us to set our credences much lower than we think is appropriate. So at least, what's at stake is the possibility of believing in a well-confirmed scientific hypothesis with more than a modicum of confidence, in light of higher-order doubt.

2. The Persistence of the Problem

Because of this sort of problem, statisticians like Savage (1972, p. 58) have proposed simply eliminating second-order and other higher-order credences from consideration. Yet this is not workable. Besides being intuitively compelling, second-order credences seem non-optional in certain instances. This is because they are *implied* by credences in objective probabilities, at least in some cases. That is so, assuming something like David Lewis's (1980) "Principal Principle".

In what follows, let P be a function that maps propositions to objective probabilities only, in contrast to C_t. Also, let 'p' be a variable that ranges only over propositions that concern a single event. (Thus, $P(p)$ will be the objective probability of an event token, not a frequency or a propensity for a type of event). The Principal Principle can then be approximated as:

> (PP) Suppose that at time t all your evidence E_t is "admissible" and consistent with $P(p)=n$. [For short: "E_t is prototypical".] Suppose also that given E_t you know that $P(p)=n$ at t. Then, if you are being rational, $C_t(p)=n$.

Lewis (1980, p. 272) confesses that he lacks a definition of "admissible" evidence; however, Jenann Ismael (2015) provides an expedient gloss of the overall import of (PP). The Principal Principle basically says that "if one knows what the [objective] chance of p is then one should (barring magical information from the future) adopt it as one's credence" (p. 197).[8] Ismael (2008; 2015) nonetheless criticizes (PP) and offers a different principle in its place. (See also Hall 1994 and Thau 1994 for a third alternative.) But without going into details, my remarks below will not depend the outcome of that debate. The essential points could be made in terms of these alternatives to (PP), *mutatis mutandis*.

[8] In conversation, Mike Titelbaum observes that some cases of "inadmissible" evidence are not actually odd or magical. I think he is right about this, but we need not hash this out for present purposes.

We are forced into second-order credences by (PP) as follows. Start with your credence regarding an objective probability. Suppose you determine based on your evidence that the objective probability of rain today is 95%. But suppose you are not certain that this is the correct objective probability; instead, your credence in that number is 95%. Then, it is a case where $C_t(P(\text{Rain})=.95)=.95$. Yet, assuming that you are being rational and that E_t is prototypical, (PP) implies that the objective probability reflects your first-order subjective probability. In which case, your 95% confidence in the objective probability means you are similarly 95% confident in the first-order subjective probability. That is, under the relevant assumptions, (PP) suggests in this case that $C_t(C_t(\text{Rain})=.95)=.95$, which is a second-order credence. And once we admit that there can be credences about our own credences, our credences become vulnerable to Humean erosion.

Now in an obvious way, this is not fair to Lewis. This is most apparent if we formulate (PP) in a more exacting manner:

(PP*) Suppose E_t is prototypical. Then, if you are being rational, $C_t(p/P(p)=n\&E_t)=n$.

This reveals that n ends up being a conditional credence for p, viz., the credence for p *given* that the objective probability of p is n (and given E_t). Moreover, it is common to assign credence 1 to what is given. So if we assume that (PP) applies only if the credence assigned to $P(p)=n$ is 1, then (PP) will not apply above, where your credence in the objective probability was only .95. Whence, (PP) need not force us into higher-order credences.

This is a fair point in the context of pure Bayesian theory, which is what concerned Lewis. But when it comes to applied Bayesianism, the idealization here is hard to tolerate. Our credence in objective probabilities is almost never 1. Or, if we return to the formulation at (PP), the point would be that the antecedent condition where $P(p)=n$ is known is virtually never a condition where $P(p)=n$ is known *with certainty.* So it seems that first-order credences will remain hedged in the applied realm, and Humean erosion remains a concern.

Before moving toward a solution, however, one further observation will prove helpful. I submit that (PP) remains plausible if 'know' is replaced with 'judge' (where a judgment is understood as an occurrent belief):[9]

(PPj) Suppose that E_t is prototypical. Suppose also that given E_t you judge that $P(p)=n$. Then, if you are being rational, $C_t(p)=n$.

This version of (PP) will be important below. Also, (PPj) is better suited for applied Bayesiansim. It is clear that a judgment of an objective probability need not be certain. What is more, evidence-based judgment (unlike knowledge) is not always true. This makes (PPj) even easier to apply: to recognize that (PPj) applies in a given circumstance, it is not necessary to establish that your assessment of an objective probability is actually true. It is enough to realize that given E_t you *judge* that $P(p)=n$. Admittedly, it is not entirely straightforward to decide what you judge at a specific time, much less whether your judgment is based on a prototypical E_t. But those decisions were necessary anyway for recognizing whether you *know* given E_t that $P(p)=n$. So in removing the truth requirement from (PP), there is still one less obstacle to identifying when (PPj) applies.

Part II

3. A Transparency-Theoretic (Partial) Solution

To circumvent Humean erosion, this section develops the hypothesis that in some recognizable conditions, you *must* be judging $P(p)=n$ at t (where 'must' expresses the right type of modality). If this correctly describes some conditions, then assuming you recognize that (PPj) also applies in those conditions, you can then recognize on those occasions that $C_t(p)=n$ must be true. Whence, in such cases, you will have some reason to say that

[9] Andreotta (this volume) argues that judgments are not the same as occurrent beliefs. He may be right about that, but if preferred, one may replace my talk of judgments with 'occurrent beliefs' without loss to my arguments. (Unfortunately, talk of occurrent belief is also difficult to sort out; see Bartlett 2018. But alas, I cannot work this out here.)

$C_t(C_t(p)=n)=1$. And if you can reasonably say *that*, you will have avoided eroding your first-order credence.

Here are two caveats. First, since the solution will operate only in certain types of circumstances, it will at best be a partial solution. There will remain other circumstances which lead to Humean erosion. Second, the solution assumes you can recognize when (PPj) applies to judgments of objective probability. Again, that requires an ability to recognize that that your evidence is prototypical, that your judgment is based on your evidence, plus an ability to recognize that you are being rational at the time. And such recognitions can hardly be taken for granted. However, the ability to recognize such things is not beyond the pale.[10] At least, assuming as much is not as ruthless as instead simply taking for granted that $C_t(C_t(p)=n)=1$, which thus far seems to have been the attitude vis-à-vis Hume.

Again, the objective of this section is to make defensible the starting point of the solution, namely, that in some recognizable circumstances (to be specified), you must be judging that $P(p)=n$. The goal is thus to claim a kind of self-knowing ability—an ability to recognize one of your own judgments under specified conditions. We shall defend such self-knowledge by, in effect, elaborating on a remark by Bernard Williams, to wit, that often "I am confronted with my belief as what I would spontaneously assert" (Williams 2002, p. 76). I also take as inspiration Sellars who proposes:

> [W]e can know what we think, in the primary sense, by literally hearing ourselves think. When we hear ourselves say (in a candid frame of mind) "I've just missed my bus," we are literally hearing the thought occur to us that we have just missed the bus. And in hearing this, we would be thinking the higher-order thought: The thought has just occurred to me that I have missed my bus. And this higher-order thought would be an auditory perceptual response to one's actual thinking-out-loud "I've just missed the bus." (1975, §§27–29)

[10] Indeed, discerning whether a conclusion is based on a set of premises is not as difficult as one might think. Demircioğlu (2021) argues quite convincingly that reflective endorsement of an argument is sufficient for that argument to express the reasons-for-which you believe the conclusion. That is so, even if your judgment of the conclusion was previously caused by something nonrational. (This also comports well with the "ratification account" of reflective reasoning, detailed in Parent 2017, chapter 11.)

Both Williams and Sellars are suggesting that given your candid utterance of a sentence "*p*" at time *t*, you can recognize by means of such an utterance that you judge that *p* at *t*.[11] However, I propose something stronger: you can recognize your judgments if you are guided (perhaps implicitly) by a strengthened version of the Sellars-Williams thesis. Roughly:

> (SW^+) If a speaker utters a sentence "*p*" at *t* in a reflex-like manner, then *as a matter of psychological law*, she judges that *p* at *t*.

If something like (SW^+) is correct, then when you produce the relevant kind of utterance, you *must* be judging that *p*, where 'must' expresses necessity in all possible worlds in which a certain psychological law holds. (Specifying this law is delayed until later.)

In consequence, when you recognize a "spontaneous" uttering of "$P(p)=n$", (SW^+) indicates that there is a psychological law which *guarantees* that you are judging that $P(p)=n$. And, assuming the antecedent conditions are met, (PPj) then *necessitates* that $C_t(p)=n$. Recognizing all this, you then have some reason to assign a second-order credence of 1 to $C_t(p)=n$.[12]

Much more elaboration is required, but it is worth noting first that the view can be classified as a "transparency" view of self-knowledge, according to which one identifies one's own beliefs not by means of a metacognitive introspective act, but by instead making a judgment about the world. Recall Gareth Evans on the matter:

> [I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me 'Do you think there is going to be a third world war?' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' (1982: 225)

---

[11] I am also greatly influence here by Bar-On's Neo-Expressivist account of self-knowing; see especially Bar-On (2004).

[12] Take heed that (PPj) does *not* suggest that $C_t(C_t(p)=n)=1$ implies $C_t(P(p)=n)=1$. (PPj) is a conditional, not a biconditional. This is as it should be: Second-order certainty about my first-order credence should not imply certainty about an objective probability.

The solution to be defended is consonant with this, insofar as one's first-order credence is known not by higher-order "inner perception" or the like, but rather by expressing a judgment about an *objective* probability concerning rain, evolution, or what have you. The verbal expression then provides a means of recognizing this judgment—and assuming (PPj) applies to the case, you can thereby *infer* your first-order credence on that basis (cf. Byrne's 2005 inferentialist transparency view).[13] If the strengthened Sellars-Williams thesis is correct, moreover, the basis of this inference (your probability judgment) is necessitated by the verbal expression, as a matter of psychological law. And this is a reason, albeit not an unassailable reason, to assign credence 1 to the first-order credence inferred. (I admit that the non-assailable feature can revive worries about Humean erosion; I comment on this further in the concluding section.)[14]

3.1 Anscombe's Distinction

So again, the aim is to defend the strengthened Sellars-Williams thesis—viz., given the right sort of spontaneous utterance at *t*, it is psychologically necessary that I judge that *p* at *t*. In making this idea more precise, we shall draw upon a distinction made in the opening of Anscombe (1957/1963). This is the distinction between the intention *to* perform an action versus the intention *in* performing the action. I can intend *to* go to the gym in the sense that I

[13] The view from Parent (2017) was distinguished from Byrne's inferentialism, for there I claimed that the expression of a first-order judgment *automatically* has a second-order judgment as a conversational implicature. Thus, knowledge of the first-order judgment, in the form of a second-order judgment, was not due to an inferential *act*. I still think this is correct as an account of the "recognition" of the first-order judgment referred to above (although space prevents me from detailing this here). Even so, it is a recognition of your *objective* probability judgment only. To recognize your first-order *credence*, something more is needed. The addition offered above is that you can *infer* the first-order credence, by means of (PPj), on the basis of the objective probability judgment. This addition thus puts the present account closer to Byrne than the (2017) account.

[14] Tang (2016) argues that a transparency account fails for self-knowledge of credences. But some of his remarks indicate only that we must be careful in how we understand objective probability. Other objections consist in problem cases for transparency (some of which feature non-prototypical evidence). But these cases can be conceded while upholding the transparency view for other cases. Nevertheless, Tang suggests this would create pressure to explain why transparency works only some of the time. Yet he allows that this pressure arises only if one expects a uniform epistemology regarding one's own doxastic attitudes (see pp. 155–156, 164). I have no such expectation; indeed, I find it quite unjustified; cf. Parent (2017, pp. 40–41); also see Parent (2019).

have a future-directed plan to visit a certain locale. But this does not yet describe my intention *in* going to the gym. For my intention *in* going to the gym is *to realize a telos*—to maintain a certain level of fitness, to relieve stress perhaps, and so forth. Anscombe also speaks of an intention-*in* as an intention-*with*: I go to the gym *with* the intention of achieving those goals. I could intend to go to the gym with a different *telos* in mind, e.g., meeting potential romantic partners. But in that case, I would still have the same intention-to, viz., to go to the gym. (Anscombe also makes further distinctions with "intentional action", but these are not needed for our purposes.)

Intentions-in come apart from intentions-to in other ways. My preferred example is where a baseball is suddenly hurled with great speed toward your head. This normally triggers a reflex of raising your arms to protect your head. And since the event happened so quickly, it is clear you never had a plan to raise your arms. That is, there was no intention *to* raise your arms—raising your arms was merely an automatic response to a threatening stimulus. Nevertheless, there was an intention *in* raising your arms; it was to protect your head. So an action can have an intention-in even when there was no intention-to.

Importantly, we sometimes utter sentences without an intention to utter the sentence. This is especially clear when I "blurt out" something that embarrasses someone. I may have had no prior intention to utter anything, much less utter something that causes discomfort. Notwithstanding, there was a communicative intention in the utterance, even though the utterance occurred with no forethought (much less an intention-to).

Question: why does the utterance occur with an intention to communicate that *p*, if there is no intention to communicate that *p*? As mentioned, an intention in an action is achieving a *telos*, and in the baseball example, it is clear that the *telos* is to protect your head. And what fixes the *telos* of the act probably has something to do with its being an adaptive behavior. It has something to do with the evolutionary advantage in protecting the head—this

is key to explaining why the behavior exists. That is so, even if protective behavior often fails. If the behavior succeeds enough to make the behavior adaptive, then the behavior will proliferate, and the adaptive feature will explain the proliferation.

Similarly, what fixes the *telos* of uttering "*p*" is whatever makes that behavior evolutionarily advantageous. And what makes it advantageous is its role in causing audiences to believe that *p*. Here too, the behavior may often fail in this regard. But if the behavior succeeds enough to make it adaptive, then its successes will explain why the behavior proliferates. (My debt to Millikan 1984; 2005 here is obvious, although following Hutto & Myin 2014, I do not use adaptionist explanations to account for the *content* of an utterance *per se*. Rather, the adaptionist explanations are used to explain the uttering behavior directly, without invoking any semantic middleman.[15])

In any event, the blurting-out utterances can be described as "reflex-like"; for the action is like the defensive maneuver in the baseball example. Both are actions associated with an intention-in but not an intention-to. Yet unlike the baseball example, "blurting out" would not result from an automatic reflex. At the least, such an utterance is not completely out of one's control—many times, I can suppress or at least cut short an utterance before I cause embarrassment. (In contrast, it seems nearly psychologically impossible to suppress a defensive reflex triggered by a speeding headward projectile.)

[15] This is not to say that I reject the semantic middleman. I am rather agnostic on whether semantic contents are real; see Parent (ms.). Nevertheless, the best model of linguistic behavior is provided by neural nets (a.k.a. "connectionist" models), and I elaborate on the connectionist underpinning in Parent (2017, §8.7). Here too, the view is not incompatible with folk psychological explanations. But it does not include a commitment to such explanations, and that strikes me as more prudent relative to the current state of cognitive science. More than that, I regard the connectionist underpinning as dialectically important in avoiding an objection from Kornblith (2013, n. 24). Kornblith worries that transparency accounts presuppose self-knowledge insofar as they assume that speakers reliably express what they really believe. However, if expressive behavior can be explained without invoking folk-psychological "knowledge" of what one judges (as with neural nets), then transparency accounts need not be defeated by Kornblith's objection. See again Parent (2017, §8.7).

3.2 The Main Argument

Using Anscombe's distinction, we may now define a few quasi-technical terms as follows. Suppose that "*p*" is a "standard" declarative token, i.e., it means that *p* on the occasion of use, and its meaning is neither ambiguous, nor metaphorical, nor overly vague, nor otherwise obscure to a problematic degree. Then:

> (EX) A speaker expresses the judgment that *p* iff she competently[16] utters "*p*" and one intention *in* the uttering is to communicate the judgment that *p* (which may or may not be her own judgment).

It is allowed that a speaker might express a judgment that *p*, and yet not judge that *p* herself. The speaker can very well express the judgment that *p* in an act of lying, for example.

One potentially unexpected feature of (EX) is that it speaks of communicating a *judgment* that p, rather than communicating that *p simplicitur*. The latter way of speaking is more common in the literature, and the former way of speaking might suggest that one intent in uttering "*p*" is to communicate that *the speaker* judges that *p*. But while that is sometimes true (e.g., when answering a pollster), it often is not true. Normally, when someone asks me the time, the intent in my uttering 'It's 12 o'clock' is to communicate that it is 12 o'clock, and not to communicate something about my current mental state.

However, this is not how my talk of "communicating a judgment that *p*" should be interpreted. I am influenced here by the observation that sometimes I utter "*p*" not to indicate a belief, but merely to draw attention to a hypothesis I am entertaining. Indeed, the linguistic act may convey that the audience *should not* take *p* to be true, at least not yet, but instead just consider it for discussion's sake. As a different sort of case, "*p*" might be uttered sarcastically,

---

[16] A speaker utters "*p*" competently only if, in routine communicative conditions, the intent *in* the uttering succeeds. With a declarative sentence, this includes communicating the judgment that *p* (which need not be the speaker's own judgment). Routine conditions are where the speaker is not impeded in producing the utterance, the audience is not impeded in perceiving the utterance, the utterance uses only terms that the audience understands, etc.

as a way of indicating that one should judge ~*p* rather than *p*. (Note: "*p*" used sarcastically can still count as a "standard" declarative token, per the earlier definition.) Such examples suggest that when "*p*" is uttered, the linguistic act suggests an *attitude* that should be taken toward *p*—whether it should be regarded as true, as false, or as hypothetical, for instance.

This just illustrates the well-known point that what is communicated by uttering a declarative is not just a bare semantic content, but rather a "speech-act content". The intent in uttering "*p*" may well be to cause the audience to believe that *p*, but it instead may be to cause the audience to merely entertain *p* as a hypothesis—or even to reject *p.* Thus, when I speak of an uttering of "*p*" as communicating a judgment that *p*, I am describing it as a speech-act where the intent-in is to cause the audience to occurrently believe *p* (rather than to merely entertain or reject *p*). Note, moreover, that this describes the intent in the linguistic act even if I am lying. In a lie, I still utter "*p*" with an intent of causing the belief that *p*, even though I myself reject *p*.[17]

Next, it will be convenient to utilize the following notion of "assertion":

(A) A speaker asserts "*p*" iff she expresses a judgment that *p*.

This is a nonstandard use of 'assert', since typically it is propositions rather than sentences that are said to be "asserted". But if preferred, one can see my talk of asserting a sentence "*p*" as shorthand for 'asserting the proposition which is semantically expressed by "*p*" on the occasion of use'.

It is now opportune to introduce the psychological law that is relevant to defending the strengthened Sellars-Williams thesis. I mark it as "(H)" so it is clear that the law has the status of a hypothesis:

---

[17] We should probably acknowledge here the possibility of sarcastic lying, where one utters "*p*" sarcastically, i.e., with the intent of causing a belief that ~*p*, where one nonetheless believes *p*. One might also ask about cases of pretense, e.g., where "*p*" is uttered by an actor in a play. In Parent (2017), I classified this as expressing a judgment that *p*, but this was a mistake. (Fortunately, nothing important depended on the point.) After all, in the usual cases, the intent in the actor's uttering of "*p*" is not to cause the belief that *p*. It is more to cause *p* to be entertained much like a hypothesis. Yet to be sure, there are differences in pretending versus hypothesizing.

(H) It is a psychological law that, if a speaker reflex-like asserts "*p*", then one intention *in* the uttering is to communicate *her own* judgment that *p*.

Drawing upon earlier remarks, the import of (H) is that the adaptive *telos* in reflex-like assertive behavior is that of causing the audience to believe something the speaker herself believes, namely, *p*. That is what the behavior is *for* in evolutionary terms; that is what explains the proliferation of the behavior, even if the behavior often fails to achieve this *telos*.

There is at least one objection to (H), but first, consider that if (H) is granted, a strengthened version of the Sellars-Williams thesis immediately follows.

(2SW^{+}) It is a psychological law that, if a speaker reflex-like asserts "*p*," then she expresses her own judgment that *p*. [From (H), (A), and (EX)]

A key idea in all this is that the intention in an assertive act inevitably leads to a successful act of a closely related kind, to wit, that of expressing a judgment. An assertive act occurs with the intent of communicating a judgment, and such an act *suffices* to express that judgment. Granted, communicating the judgment to the audience may fail—the audience may misperceive the utterance, or not understand the terms it uses, etc. But since an assertion is a competent utterance by definition, and occurs with the intent of communicating that judgment, an assertion will always succeed in at least being *expressive* of the judgment.

So when we hypothesize the law that *reflex-like* assertions occur with an intention of communicating one's own judgment, reflex-like assertions automatically succeed in expressing one's own judgment. In which case, (2SW^{+}) can guide me to recognizing when, as a matter of psychological law, I judge that *p*.[18]

[18] More details are desirable on how one recognizes that an utterance is a "reflex-like assertion" in the relevant sense, but I cannot present these here. Instead, see Parent (2017, §8.6).

4. Two Objections

Again, the Humean problem was to judge what your first-order credence is without "eroding" that credence. The partial solution being offered is that, if you reflex-like assert "$P(p)=n$", then ($2SW^{+}$) can indicate that as a matter of psychological law, you are judging $P(p)=n$ at the time of the assertion. And so, under the supposition that (PPj) applies, this is a reason to assign credence 1 to $C_t(p)=n$ at that time.

4.1 The Incidence of Reflex-Like Asserting

One objection is that this avoids Hume's problem only in a very unusual kind of case, to wit, one where you reflex-like assert a sentence about an objective probability. After all, assertions about probability are rarely produced in a reflex-like way. In fact, I demur, although explaining why requires further clarity on what counts as a "reflex-like" assertion.

The notion was first introduced using the clearest sort of example, one where I "blurt out" something that causes embarrassment. But strictly speaking, a reflex-like assertion is any assertion that is produced without an intention *to* produce that utterance. And far from being the exception, this is normally how assertive behavior goes. It is certainly possible for an utterance to result from an intention-to, e.g., I can plan out what I say in my conference talk—or even in conversation, I can take a mindful approach where I plan what to say before I say it. But normally, we just "think out loud" or just start talking without a plan for what to say, letting speech flow without any self-conscious consideration.

In fact, even if deliberation precedes an assertion, that does not mean that the speaker had an intention to produce that particular assertion. For our purposes, the most relevant example is when someone is calculating a probability. Suppose a teacher asks a student to calculate the probability of getting a face card on the next draw. We can imagine the student working the numbers on paper, carrying the '1', counting the decimal places, and so forth,

until her work reveals that the answer is .07. The student then immediately asserts 'It's 7%'. In the normal course of events, there would be no prior intention to assert that particular sentence. Granted, there was a prior intention to assert whatever answer she ended up with on paper. But a prior intention to assert "whatever answer I arrive at" is not the same as a prior intention to assert 'It's 7%' specifically. (This illustrates the opacity of intentions, which Anscombe discusses as well.)

4.2 Lying in a Reflex-Like Way

So the example is one where ($2SW^+$) would be applicable; further, this type of example makes clearer the wide scope of application of ($2SW^+$). But now, the worry might be that ($2SW^+$) applies *too* widely. For it is easy to imagine a speaker who ends up lying even though she had no prior intention to lie. Indeed, one might have a firm intention *against* lying, but a lie might still be triggered under stress (cf. Parent 2017, §8.8). And a reflex-like lie is not an expression of the speaker's own judgment. So these seem to falsify ($2SW^+$) outright.

My excuse here is to plead expository convenience. That is, my intention all along was really to defend a qualified version of (SW^+) along the following lines:

($3SW^+$) It is a psychological law that, if a speaker reflex-like asserts "*p*", then unless she knows otherwise, she expresses her own judgment that *p*.

However, it made the initial exposition easier to leave off the additional clause. But at this stage, it is appropriate to detail what I really mean to be defending.

Note that the additional clause does not trivialize the thesis, for the consequent has a form that is equivalent to "either she knows ~*p* or *p*", which is non-trivial. Moreover, the additional clause does not create any difficulty on applying the thesis to recognize your expressed judgments. That is so, on the assumption that lying means expressing a judgment

you know is not your own. The assumption here is quite plausible, moreover. If you express a judgment that is not your own, but you don't realize it, then you are guilty not of lying but of a kind of self-ignorance.[19]

But this leads us to a follow-up objection. Consider for instance the case of lying to *oneself*, of thinking you judge *p* when really you are agnostic or even judge ~*p* instead. Thus, suppose you reflex-like assert 'My mother really loved me!' when in fact you unconsciously harbor the belief that she never really loved you. Then, since your true feelings are hidden in your unconscious, you will have no knowledge that the assertion fails to express your own judgment. And so, ($3SW^{+}$) will falsely suggest that you indeed judge that your mother really loved you, even though that is merely self-deception.

Freudian cases are the most salient examples of concern, but other relevant examples feature the widespread phenomenon of *confabulation*. There is plenty of psychological evidence that people confabulate their beliefs routinely (saying things that they don't really believe), in a way that strongly suggests that people have little idea what they really believe. (For an overabundance of references, see n. 2 of chapter 2 in Parent 2017.) This, moreover, puts in jeopardy the venerated practice of self-reflection. After all, if you don't know what you really believe, then you could not meaningfully reflect on what you believe. Parent (2017) is a book-length attempt to rescue self-reflection from this threat.

At any rate, confabulation means, potentially, that a speaker routinely and unwittingly expresses judgments that are not her own judgments. Indeed, I suspect that the reflex-like nature of much linguistic interaction contributes to the frequency of confabulation. Perhaps it is incorrect that in all such cases the speaker is *lying* to herself: if a speaker reflex-like asserts in full confidence 'The winning lotto numbers will be 3, 14, 29, 35, and 48', she *could* be

[19] Interestingly, there may be atypical lies where the speaker simultaneously regards the judgment as her own and also not (thus exhibiting inconsistency). See Krstić (forthcoming). But the point would remain that the speaker *at least* regards the judgment as not her own.

saying something true. Moreover, she could be saying something that she *thinks* she judges. But it will often be a case of overconfidence where, deep down, she knows very well that she has no idea what the lotto numbers will be. In which case, what she reflex-like asserts will not be expressive of a judgment she actually has.

In chapter 2 of Parent (2017), psychological experiments on confabulation are scrutinized in detail, and I cannot recap that discussion here. But what is most relevant for present purposes is that, contrary to what one might expect, it is natural to interpret cases of confabulation as cases where the subject *in some sense* believes what she says.[20] The Freudian patient who reflex-like asserts 'My mother really loved me!' is naturally ascribed the corresponding judgment, if only to explain the reflex-like assertion. The subject is "confabulating", however, because it is also natural to ascribe the opposite belief in her non-conscious mind. Yet this would be a case where the subject believes and rejects the proposition that her mother really loved her. It would not be a case where one of the beliefs was simply fabricated *ex nihilo*. That is so, even though the subject may later admit (after extensive therapy) that she knew all along that her mother never really loved her. It is not so much that she has identified that one judgment was actually an illusion; it is rather that she has relinquished one of her *bona fide* judgments that was inconsistent with a more well-founded belief.

The moral of the story is that ($3SW^{+}$) remains correct even when a reflex-like assertion amounts to an unwitting confabulation. For "confabulation" is not sheer metacognitive make-believe; it is rather a case of a judgment that conflicts with other beliefs. But at the time of the reflex-like assertion, the judgment was indeed the speaker's own.

[20] Such a defense of self-knowledge against Freudian cases was first offered in Bilgrami (1998). Thanks to Ben Winokur for the reference.

5. Closing Remarks

In this paper, I've tried to convince the reader that Humean erosion is a real problem. And I've tried to suggest a partial solution in the form of a transparency view. The view is roughly that as a matter of psychological law, your reflex-like assertion of "$P(p)=n$" implies that you judge $P(p)=n$, at least at the time of the assertion. Under some natural assumptions, (PPj) then necessitates that your credence in p is n. So, upon recognizing all this, you will have a reason to assign credence 1 to your first-order credence. And that is a second-order credence which leaves your first-order credence intact.

I've tried to be forthright, however, that your reason to assign credence 1 will not be an unassailable reason. But if there remains doubt about the second-order credence, then Humean erosion will remain a threat, and all will be for naught. I cannot fully respond to this objection here. As mentioned earlier, it is still better to justify a second-order credence of 1 rather than assume it without argument (which has been the status quo). But more than that, the objection reflects a need for us to get clearer about talk of "certainty".

I suspect that the word 'certain' is frequently used to tag something that is taken for granted. But what is taken for granted varies from context to context. So it happens that one statement gets tagged as "certain" in one context, but not in another. Moreover, when a statement is deemed uncertain, the success of its defense depends on whether its premises are taken for granted. Thus, depending on what is taken for granted, a statement may be successfully defended in one context, and unsuccessfully defended in another context.

We often take for granted first-order credences, according them the status of certainty. But we can rightly question our credences in other contexts, and this is where Hume's regress arises. In response, I have defended self-knowledge of credences by appeal to the psychology of expressive verbal behavior. If the psychology is accepted as a matter of course, then I

submit that the defense succeeds. Otherwise, the defense will fall flat, as the objection suggests.

My point, however, is that this is just business-as-usual. Whether an argument succeeds depends on whether its premises are taken for granted—and since that varies with context, so too will the success of an argument. There is no such thing as premises that cannot be questioned. This flies in the face of Descartes, who believed that some things were immune from doubt. But rational inquiry allows any statement to be called into question—even the statement that I exist or that I think. (Think of Hume on the self and of eliminative materialism.) The insight here is not mine of course: "empirical knowledge, like its sophisticated extension, science, is rational, not because it has a *foundation* but because it is a self-correcting enterprise which can put *any* claim in jeopardy, though not *all* at once" (Sellars 1956/1963, p. 170). Philosophers often act as if rational inquiry requires premises that are immune from doubt. But what it actually demands is that *nothing* is immune from doubt. The consequence is that an argument can succeed only relative to a context of inquiry.[21]

References

Andreotta, A. (2024). Transparency, Moore's paradox, and the concept of belief. In A. Andreotta & B. Winokur (Eds.), *New Perspectives on Transparency and Self-Knowledge*. Routledge.

Anscombe, G.E.M. (1963). *Intention*, 2nd edition. Harvard University Press.

Atkinson, D. & Peijnenburg, J. (2020). 'Till at last there remain nothing': Hume's *Treatise* 1.4.1 in contemporary perspective. *Synthese*, 197(8), 3305–3323.

Bar-On, D. (2004). *Speaking My Mind: Expression and Self-Knowledge*. Oxford University Press.

Bartlett, G. (2018). Occurrent states. *Canadian Journal of Philosophy*, 48(1), 1–17.

Bilgrami, A. (1998). Self-Knowledge and resentment. In Wright, C., Smith, B.C., & Macdonald, C. (Eds.) *Knowing Our Own Minds* (pp. 207–241). Oxford University Press.

Byrne, A. (2005). Introspection. *Philosophical Topics*, 33, 79–104.

[21] My thanks to Adam Andreotta, Chandler Hatch, Jim Hutchinson, Mike Titelbaum, Siegfried Van Duffel, and Ben Winokur for valuable feedback on earlier versions of this paper.

Demircioğlu, E. (2021). Reason, rationalizations, and rationality. *Philosophia*, 51(1), 113–137.
Eder, A-M. A. (2023). Evidential probabilities and credences. *British Journal for the Philosophy of Science*, 74(1), 1–23.
Evans, G. (1982). *The Varieties of Reference*. Clarendon Press.
Kornblith, H. (2013). Naturalism versus the first-person perspective. *Proceedings and Addresses of the American Philosophical Association*, 87, 122–142.
Hall, N. (1994). Correcting the guide to objective chance. *Mind*, 103, 505–518.
Hume, D. (1888/1960). *A Treatise of Human Nature*. L.A. Selby-Bigge (Ed.). Oxford University Press.
Hutto, D. & Myin, E. (2012). *Radicalizing Enactivism: Basic Minds without Content*. MIT Press.
Ismael, J.T. (2008). Raid! Dissolving the big, bad bug. *Noûs*, 42(2), 292–307.
Ismael, J.T. (2015). In defense of IP: A response to Pettigrew. *Noûs*, 49(1), 197–200.
Krstić, V. (forthcoming). Lying by asserting what you believe is true: A case of transparent delusion. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-023-00700-1.
Lewis, D.K. (1980). A subjectivist's guide to objective chance. In R. Jeffrey (Ed.), *Studies in Inductive Logic and Probability, Volume II* (pp. 263–294). University of California Press.
Millikan, R.G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.
Millikan, R.G. (2005). *Language: A Biological Model*. Oxford University Press.
Parent, T. (2017). *Self-Reflection for the Opaque Mind: An Essay in Neo-Sellarsian Philosophy*. Routledge.
Parent, T. (2019). Colivan Commitment vis-à-vis Moore's Paradox. *Philosophia*, 47(2), 323–333.
Parent, T. (ms.). Ontology after Folk Psychology; or, Why Eliminativists should be Mental Fictionalists. Available at http://tparent.net/ElimMF.pdf.
Savage, L.J. (1972). *The Foundations of Statistics*, 2nd edition. Dover.
Sellars, W. (1956/1963). Empiricism and the philosophy of mind. In H. Feigel & M. Scriven (Eds.), *Minnesota Studies in the Philosophy of Science vol I* (pp. 253–329). University of Minnesota Press. Reprinted with additional footnotes in his *Science, Perception, and Reality* (pp. 127–196). Routledge & Keegan Paul.
Sellars, W. (1975). The Structure of Knowledge. In H.N. Castaneda (Ed.), *Action, Knowledge, and Reality: Critical Studies in Honor of Wilfrid Sellars* (pp. 295–347). Bobbs-Merrill.
Tang, W.H. (2016). Transparency and partial beliefs. *Philosophy and Phenomenological Research*, 95, 153–166.
Thau, M. (1994). Undermining and admissibility. *Mind*, 103, 491–504.
Titelbaum, M. (2004). How to think like a Bayesian. *Psyche*. Available at https://psyche.co/guides/how-to-think-like-a-bayesian-and-make-better-decisions
Williams, B. (2002). *Truth and Truthfulness*. Princeton University Press.
Williamson, T. (2000). *Knowledge and Its Limits*. Oxford University Press.