

# The Role of Assurance in Judgment and Memory<sup>1</sup>

Edward Hinchman

Florida State University

[8 July 2020]

“Don't you even remember that?” asks the manipulatively doubting husband played by Charles Boyer in *Gaslight* (1944). “Yes, of course, I do,” answers his wife, played by Ingrid Bergman. “But suddenly,” she immediately clarifies, “I am beginning not to trust my memory at all.” Does she remember what happened to that broach? Though her memory is intact psychologically, it has begun epistemically to fail her. An epistemological theory of memory ought to explain this distinction: how one can lose touch with one's memory in a way that does not involve amnesia or dementia. Since the 2016 elections ‘gaslighting’ has served as a political diagnosis; earlier it more generally marked a form of emotional abuse. But *Gaslight* appears to reveal an important link between memory and testimony: one treats one's memory as a source of epistemic warrant only insofar as one can get relevant others to acknowledge it as capable of providing warrant through testimony. I'll argue that this formula captures only part, and not the deepest part, of the link between memory and testimony. But the film vividly depicts how the epistemic faculty of memory puts us all at the mercy of at least some potential interlocutors. If they do not believe us, we may lose a feel for how to ‘believe ourselves.’

The film shows us how the link may emerge in practice. How does it emerge in theory? We can begin from the popular idea that memory resembles testimony insofar as each does not

---

<sup>1</sup> Forthcoming in Sanford Goldberg and Stephen Wright (eds), *Testimony and Memory: New Essays in Epistemology* (Oxford University Press, 2020 or 2021).

generate but ‘preserves’ epistemic warrant.<sup>2</sup> My question is how such ‘preservation’ does its epistemic work – and whether the idea (or metaphor) of preservation captures how it works. Theories of testimony that emphasize something like preservation generically divide into those that do and those that do not emphasize an interpersonal relation between testifier and addressee. Should we stop at minimal non-reductionism, at the claim that the testimonial warrant ‘preserved’ from testifier to address does not reduce to other sources of warrant, or should we go further and explain it as partly grounded in the interpersonal relation wherein the speaker addresses her audience? The latter approach, distinctive of an ‘assurance theory,’ draws a key distinction between how the addressee can get warrant from being thus addressed and how a non-addressed overhearer can get warrant from merely witnessing the speech act, a distinction that more minimal non-reductionists reject. The approach is distinctive in its emphasis on the epistemic significance of how the testifier invites the addressee’s trust, an invitation that she does not extend to overhearers. I have defended an assurance theory of testimonial warrant in other work.<sup>3</sup> Here I’ll develop an assurance theory in the epistemology of memory.

How could the ‘preservation’ of warrant through memory involve an assurance? What would even count as an *intrapersonal* assurance? On an assurance-theoretic approach, we can understand the relation that ‘preserves’ warrant by contrasting it with a relation that has gone pathological in specific respects. An assurance-theoretic approach to testimony explains the

---

<sup>2</sup> For some canonical formulations, see Tyler Burge, “Content Preservation,” *Philosophical Review* 102:4 (1993); Michael Dummett, “Testimony and Memory,” in A. Chakrabarti and B. K. Matilal, *Knowing from Words* (Springer, 1994); and David Owens, *Reason without Freedom* (Routledge, 2000), Chapters 9-11. I’ll engage Burge’s view directly in the final three sections below.

<sup>3</sup> “Telling as Inviting to Trust,” *Philosophy and Phenomenological Research* 70:3 (2005); “Can Trust Itself Ground a Reason to Believe the Trusted?,” *Abstracta*, Special Issue VI (2012): A Symposium on Paul Faulkner’s *Knowledge on Trust*; “Assertion, Sincerity, and Knowledge,” *Noûs* 47:4 (2013); “Assurance and Warrant,” *Philosophers’ Imprint* 14:17 (2014); “On the Risks of Resting Assured: An Assurance Theory of Trust,” in *The Philosophy of Trust*, edited by P. Faulkner and T. Simpson (Oxford University Press, 2017); and “Assertion and Testimony,” in *The Oxford Handbook of Assertion*, edited by S. Goldberg (Oxford University Press, 2020).

pathology by appeal to the illocutionary norms governing the speech act of telling, with an emphasis on violations on the speaker's side. Pursuing the parallel, what norms govern the contribution of the earlier self in an intrapersonal memory relation? There are two sides in the species of memory that we're considering: the earlier self judges that *p*, then memory keeps the later self in epistemic touch with that judgment by 'preserving' its warrant. Nothing like an illocutionary relation figures here. (When you 'tell yourself that *p*,' you're trying to 'convince yourself' that *p* and as such typically do not yet even judge that *p*.) Despite the absence of a speech act, I'll argue, the earlier self offers an assurance by extending an invitation to trust. There is an invitation to trust in your earlier self's judgment that *p* precisely here: that judgment projects a future in which you will continue to trust yourself on the question whether *p* when you lose access to the evidence that informed the judgment. That trust in yourself posits a worthiness of trust that ought to lead others to accept the invitation to trust informing your testimony that *p*.

We grasp the normative complexity of this intrapersonal relation by seeing how pathology emerges when it breaks down – as it so vividly does in the victim's plight in *Gaslight*. Kate Abramson has rightly emphasized how gaslighters aim to disable the victim's judgment in order to avoid challenges to their dialogical authority.<sup>4</sup> Our question is *how* gaslighting can do this. As I'll explain, gaslighting works by exploiting a vulnerability built into a key dimension of rational responsiveness: your responsiveness to evidence about your own status as relevantly trustworthy when you judge. You need others, and a responsiveness to others, to serve as a corrective when your judgment is not relevantly trustworthy. You are not the prisoner of your judgment – passively willing to go wherever it takes you. You could not be fully rational if you

---

<sup>4</sup> Kate Abramson, "Turning Up the Lights on Gaslighting," *Philosophical Perspectives* 28 (2014).

were. But your responsiveness to others' resistance to your judgment does more than make you fully rational. As we see depicted in the film, it also makes you vulnerable to manipulation.

I'll present the case from *Gaslight* in Section I. In Sections II and III, I'll develop a dialectic between two interpretations of it, an Evidential Interpretation that may seem plausible at first glance and an opposing Assurance Interpretation that I favor. The point isn't that the latter is the best interpretation of the film but that it is coherent and compelling, and that as such it helps clarify how an emphasis on assurance might vindicate a normative parallel between testimony and memory. The Assurance Interpretation will lead to questions about the Socratic nature of interpersonal assurance (Section IV), how there could be such a thing as intrapersonal assurance (Section VI), and in what sense, if any, an assurance can count as (in Tyler Burge's term of art) 'preserving' content or epistemic status across perspectives, whether interpersonal or intrapersonal (Section V). I'll conclude, in Section VII, by emphasizing the Socratic nature of intrapersonal assurance, which imposes, as I'll argue, a normative constraint on judgment. I'll emphasize throughout a dimension of normativity that imposes not merely an epistemic constraint but a constraint on how judgment manifests a grasp of concepts.

### *I. How might testimony cure a pathology of memory? The case of Paula Anton*

Let's begin by seeing why we must understand the pathology of gaslighting in this deeper dimension, not merely as a pathology of judgment but as a pathology of the understanding – that is, as lying in one's grasp of the concepts that one would apply in judgment. How does the link between memory and testimony emerge in *Gaslight*? Briefly put, Paula Anton lets herself be manipulated by her husband Gregory into the fear that she is going insane. The denouement

arrives as she escapes from that fear through conversation with a detective, Brian Cameron (Joseph Cotten). The scene depicts vividly how her words not only mark that escape but enact it:

**Paula:** I'm sorry, I haven't talked to anyone for a long time. I can't talk to you either. I'm afraid... I'm not... I'm afraid I...

**Brian:** You're afraid you're going out of your mind? Well, I'm here to prove to you that you're not. To help me do that, you've got to answer my questions.

Paula proceeds to answer his questions. And answering them cures her affliction, a cure expressed by the self-recognition: "At last I can tell this to someone."

How exactly does Paula's regained capacity to engage in acts of testimony constitute her cure? Best to flip that question around: how does her fear for her sanity lie in her feeling that she cannot talk to people? As a first approximation, one might hypothesize that gaslighting works by exploiting a constitutive link between interpersonal reason-giving through testimony and intrapersonal rational self-governance through judgment. On this view, the distinctive epistemic injustice that we call 'gaslighting' embodies a more complex form of the silencing than we find in illocutionary disablement.<sup>5</sup> Unlike the illocutionarily disabled, the victim of gaslighting can perform the speech act in question, since she can get her illocutionary intention recognized by those whom her speech act addresses. What she cannot do, on this view, is thereby regard herself as a source of epistemic reasons for those addressees. In systematically failing to regard

---

<sup>5</sup> For this term, see Rae Langton, "Speech Acts and Unspeakable Acts," *Philosophy and Public Affairs* 22:4 (1993), 315. For development of the idea, see Jennifer Hornsby and Rae Langton, "Free Speech and Illocution," *Legal Theory*. 4:1 (1998).

herself as a source of reasons for addressed others, she gradually loses grip on how her judgment could provide rational guidance for herself. And so the victim becomes more deeply silenced in a dimension of epistemic self-mistrust. She can speak but won't. She can give reasons by speaking her mind – epistemic reasons to believe her testimony – but because she cannot herself feel the force of those reasons it seems to her as if she hasn't got a mind to speak. On this view, an understanding of these connections between reason-giving through testimony and rational guidance through judgment enables us see how it could cure her merely to find someone who, like Brian Cameron, will actually listen and take her at her word. In giving what this interlocutor will accept as reasons grounded in 'her word,' Paula relearns how to treat her judgment as a rational guide – as a source of rational requirements for herself and of epistemic reasons or warrant for others. (I'll use 'reasons' as the summary term for all such forms of rational guidance. Though this usage obviates a debate that I lack space to consider, nothing turns on it.<sup>6</sup>)

It is, I believe, a good first pass on an account of how gaslighting works to say that it disables its victim's judgment by thus trading on a normative parallel between memory and testimony. But we now confront a legitimate worry that the account cannot do justice to two core elements in the explanandum presented by gaslighting. To see the worry, we need to step back and ask what makes gaslighting a distinctively important concept in understanding our social world. (I'll take for granted that it is both important and distinctive.) There are two core elements in gaslighting: the *pathology* that lies in the victim's feeling that she is, as Brian Cameron puts it in the film, 'going out of her mind' and the *cure* that lies (somehow; we're

---

<sup>6</sup> In fuller treatments ("Rational Requirements and 'Rational' Akrasia," *Philosophical Studies* 166:3 (2013), and "On the Risks of Resting Assured: An Assurance Theory of Trust," I clarify that – in light of John Broome's work (e.g. "Are Intentions Reasons? And How Should We Cope with Incommensurable Values?" in C. Morris and A. Ripstein (eds), *Practical Rationality and Preference* (Cambridge: Cambridge University Press, 2001), 98-120) – these 'reasons' are better called 'rational requirements' or 'dimensions of rational coherence.'

trying to understand how it lies) in the restoration of her capacity to give testimony – as Paula puts it, to ‘tell this to someone.’ If we can understand how gaslighting includes both elements, we can explain the most fundamental datum that it is dialogical rather than monological. Though Gregory Anton’s aim is ultimately monological – to have his wife institutionalized in order to gain access to jewels hidden in her attic – his gaslighting tactic specifically targets Paula’s ability to enter into dialogue with people.<sup>7</sup> But how can we explain the two core elements? How does Paula’s testimony to Brian provide a cure for her pathology of memory? It is possible, of course, that *Gaslight* does not reveal any such deep connection between testimony and memory. But my hypothesis is that we ought to try to vindicate the appearance that it does.

The proposal that gaslighting is dialogical in a way that reveals a connection between testimony and memory derives from an apt observation: both Gregory’s and Brian’s interventions target Paula’s capacity to give – and, more fundamentally, to view herself as capable of giving – reasons. But if we treat this observation as the key to explaining the two core elements, we confront the worry that Paula’s response to Gregory’s intervention is not, from her own perspective, irrational – not anything that, as such, stands in need of a cure. Consider what disables Paula’s judgment – that is, her self-assessed entitlement to a memory judgment – about the location of a painting or the contents of her purse. One central cause of the disability is her appreciation of the evidence that Gregory has given her of her unreliability as a judge. Both well-placed onlookers and we the audience can see that this evidence is manipulatively

---

<sup>7</sup> A word on my methodology using this film. Since there are no actual gaslights in our world, the term ‘gaslighting,’ used in the way we’re investigating, obviously functions as a metaphor. More fully, the usage comprises two figures of speech: it is a metaphor built on a metonymy. Thinking about the 1944 Hollywood film (I’ll ignore the 1940 British film and the earlier play, since few have seen these), wherein there are actual gaslights, we metonymically treat that actual gaslighting as standing for the psychological dynamic to which it bears a key relation. As I’ll emphasize, I present my interpretation of the film merely as a possible interpretation. What’s important is merely that it be coherent and compelling as a possible case about which we can reflect philosophically.

misleading. But part of what defines her predicament is that Paula cannot. As far as she can tell, the evidence really does discredit the reliability of her judgment. So she mistrusts her judgment – as it would have been reasonable for her to do had the evidence not been misleading.

How, we might generally wonder, does Paula's self-mistrust amount to a pathology? She is in a nasty predicament – with the evidence available to her and with the interpersonal relation between her and her husband. But there is no pathology, it seems, in her. She is responding to her predicament as well as she can, and the self-mistrust that disables her judgment seems warranted by the evidence at hand. The film's melodrama requires that this plot contour be simplified to the point of caricature. But she is not 'going crazy' in letting this evidence weigh with her, even if that evidential responsiveness leads to self-mistrust. What we have, it seems, is a subject who comes to mistrust herself by responding to evidence of her unreliability and then regains self-trust by responding to new evidence that she is reliable in addressing testimony to Brian Cameron, since he treats her and her judgment as reliable. Her testimony 'cures' her only insofar as it generates evidence that either outweighs or defeats the evidence that warrants her self-mistrust. (The former defeats the latter evidence if she regards the former as evidence that the latter is misleading; otherwise it merely outweighs it.) If our aim is to explain the pathology induced specifically in the victim by gaslighting, this approach appears to miss the mark.

In the next two sections we'll consider two very different ways of responding to this worry. In Section II, we'll consider what I'll call the Evidential Interpretation of the proposal that we've considered in this section. In Section III, we'll consider an opposing Assurance Interpretation. As we'll see, the key issue between them lies in how they respond not to this worry about Paula's affliction but to a different worry about how her testimony to Brian



Cameron could constitute a cure. I'll argue that the Assurance Interpretation offers a compelling explanation of this contribution.

## *II. The Evidential Interpretation*

The Evidential Interpretation of our proposal emphasizes the evidence available to Paula about her condition: first the evidence that Gregory gives her that she is in a pathological condition, then the evidence that Brian gives her that she is not. To see what's distinctive of Paula's perspective in the first stage, it helps to compare her predicament to the predicament of an elderly subject whose spouse confronts her with evidence of her failing memory. Consider two scenarios with such a subject – two variants on what I'll generically call the **Dementia Case**.

- Nice** Donna does not realize that she suffers from short-term memory loss, but her ailment is obvious to her spouse, who, concerned for her welfare, does his best to alert her of her symptoms.
- Nasty** Like Donna, Delilah does not realize that she suffers from short-term memory loss, but her spouse sees an opportunity to manipulate her, for his own selfish reasons, by making her aware of her symptoms. It is convenient for the spouse's manipulative gambit that he can speak truthfully in testifying to her symptoms, but he would have offered the same testimony (with further manipulation of evidence, as needed) even if it had been false.

Donna's predicament in *Nice* resembles Paula's in *Gaslight* (with, say, a misplaced broach or a misremembered painting) but the evidence lovingly presented by her spouse counts as good evidence that Donna ought not to trust her memory, and generally her dispositions to judge on that basis, in the way that she used to do. This subject is not at all 'going crazy' – though it may feel disorienting for Donna to mistrust her own memory and judgment in the way required by the evidence her spouse presents. In *Nice* we're imagining a healthy marriage, nothing like the pathological relation into which Paula has unknowingly entered with Gregory. But in *Nasty*, by contrast, we imagine that our dementia patient, Delilah, confronts a second problem, over and above the dementia itself: her spouse aims to make her believe that she has dementia not in order to help her cope with the symptoms but in order to manipulate her into falling in line with some otherwise unrelated plan of his own. Delilah's predicament thus exactly resembles Paula's – with the one exception that Delilah really has the ailment that Gregory aims to make Paula falsely believe that she has.

How might we get a Brian Cameron figure into this story? The figure needn't, of course, be just one person, but whoever it is must – to follow this model – offer the possibility of testimonial relations that that would counter the manipulative testimony of Delilah's spouse. We might begin with this case:

**Noncon** Exactly as in *Nasty*, except now Delilah is able to engage in testimonial relations with other interlocutors, and those relations reveal to her that these others regard her judgment and memory as reliable in a way that her spouse does not.

This new case sets up a conflict between the manipulative intervention of Delilah's spouse and these non-manipulative testimonial interventions from others. (Think of the 'noncon' as deriving from how these others aim to provide the antidote to the spouse's 'con.')

In Noncon, the new testimony makes evidence available to Delilah that her spouse may be manipulating her, in something like the way that Gregory manipulates Paula in the film. And this new evidence counters the evidence that Delilah's spouse gives her that she is suffering from dementia. Which evidence is real, and which merely apparent (perhaps because intentionally misleading)? We have stipulated both that Delilah really is suffering from dementia and that her spouse really is manipulating her, testifying to her dementia for reasons that have little to do with evidence that the testimony is true. What if we lift both restrictions? That leaves us with the next case, and a new protagonist.

**Nebulous** As in Noncon, except that we lift the stipulation that the subject – now call her Deborah – suffers from dementia and that her spouse is manipulating her. We now say that she *may* suffer from dementia, and that he *may* be manipulating her. But it is possible that she does not and he is not – and that the other interlocutors are themselves manipulating her, or perhaps merely speaking in error. Everything, in other words, is up for grabs. All we specify in this case is the testimony.

Nebulous thus returns us to how things look from the perspective of someone in Paula's predicament. All Deborah and Paula have is this testimony – of their spouses, and of others who belatedly opposed that spousal testimony – to weigh against the felt authority of their own

memory judgments. How should they weigh this testimony? It has the virtue of simplicity to say that they should weigh it simply as evidence – alongside whatever other evidence they have – in their efforts to understand where things stand for them.

How might this comparison help us make sense of the idea that Paula fears for her sanity? On first glance, this emphasis on evidence appears to make the problem sketched in the previous section even worse. How you ought to respond to apparent evidence in a case with this structure can be complicated. But in navigating the complexities, as we saw, you need not be ‘going out of your mind.’ You might, of course, navigate them badly. But, as we noted in the previous section, Paula does not navigate her complicated predicament badly. Though the film’s melodrama distracts us from a realistic understanding of her struggles, she appears to manage her predicament in a broadly reasonable manner. That said, however, the parallel with the Dementia Case suggests a simple diagnosis of Paula’s fear that she is losing her sanity: she fears for her sanity because she does not believe that any plausible medical diagnosis could justify the charge that she is suffering from dementia. The parallel with the Dementia Case accordingly suggests a general account of gaslighting: Gregory gaslights Paula insofar as he (i) aims to give her what she will receive as evidence that she is suffering a serious psychological impairment but (ii) does not believe that she actually suffers from that impairment. Gaslighting would thus amount to a simple species of lying. The gaslighter lies to – in general, willfully misleads – his victim about the state of the victim’s mental health, aiming to give her misleading evidence that her mental health is impaired. If the victim has independent evidence to believe that her mental health may be compromised in the respect at issue, the manipulation gives her evidence that she may be able to cope with constructively. But if there is no such independent evidence, and therefore no constructive story with which to frame the gaslighter’s intervention, then the victim cannot

respond constructively. In that case, all she can do is lapse into despair about her condition, a despair that is likely to include despair about her ability to serve as a source of reasons, whether for testimonial addressees or for herself.

Although I'm going to reject this interpretation of our proposed account of gaslighting, I believe we can learn something important from seeing why it's wrong. The first step in that dialectic is to note that while this account appears to provide an effective explanation of one of the two core elements discussed in Section I, it does so at the cost of rendering inexplicable the other core element. Perhaps we can now make sense of Paula's feeling that she is losing her sanity, but how can we make sense of Brian Cameron's success in helping her regain it? As far as the Evidential Interpretation takes us, testimony cannot on its own constitute a cure, or even part of a cure, for the pathology of memory that we find in gaslighting. The fact that Brian treats her judgment as reliable may give Paula evidence that it is reliable, but she could equally get evidence of its reliability by learning that a snooping Brian treats her diary as reliable or that a neuroscientific Brian treats her beliefs as reliable by reading them directly from activity in her brain – not to mention more direct evidence of reliability that does not emerge from others' responses to it. A response to Paula's testimony can give her evidence of her reliability, but if that's what she needs she would do better to keep written records of her verified judgments, and bear witness to the reliability herself. How then might testimony on its own contribute to a cure for this pathology of memory?

As I'll now explain, the challenge lies in explaining how Paula could get the evidence that she seeks – evidence about her capacity, as testifier, to give an addressee reasons – specifically from her addressee's receptivity to those reasons. I'll argue that the evidential model that we have developed in this section cannot explain this possibility but that an alternative

model can. Toward that end, I'll propose an alternative interpretation of Brian Cameron's intervention in the concluding act of the film. In this section, we interpreted Brian's uptake as confirming for Paula the reliability of her memory judgments. In the next section I'll argue that we should interpret his uptake as also confirming for Paula the communicative efficacy of some important concepts that she fears she may not properly grasp. This new interpretation does not conflict with the earlier but supplements and deepens it. What's crucial for our purposes is that it deepens our understanding in a way that highlights how Brian's reception of her testimony helps her cure the pathology of memory induced by her husband's gaslighting.

### *III. Introducing the Assurance Interpretation*

There are two ways we might describe the conflicts in *Noncon* and *Nebulous*. We might present the conflict as a question of whom to believe: one's spouse or these other testifiers? Or we might say that this broader interpersonal context gives our subject further evidence, evidence that may outweigh (or otherwise defeat) the evidence generated by her spouse's interventions. Focusing on the film, should we ask whom Paula ought to believe, Brian Cameron or Gregory? Or should we ask instead about the evidence – whether the evidence made available through Brian Cameron's testimony outweighs the evidence made available through Gregory's? The account presented in the previous section endorsed the second way of posing the question. We're now ready to consider how we might develop a principled alternative.

To develop the alternative, it will help to step back and review how we've reached this point. On the proposal we considered in Section I, the pathology at the core of gaslighting lies in how it disables the victim's judgment specifically by undermining her conception of herself as a

source of reasons for potential interlocutors. If it cannot serve as a source of reasons for recipients of her testimony, Paula fearfully reasons, then her judgment cannot serve as a source of reasons for herself. When I say that she ‘fearfully reasons,’ I mean that this train of thought articulates her fears, not that it informs a judgment that would be disabled by a self-application of this reasoning. Her predicament is shaped, all sides agree, by her fearful or anxious reluctance to judge and her consequent reliance on the influence of the gaslighter. The question is how to explain that reluctance, and specifically what, if anything, the reluctance has to do with her capacity to give testimony.

In Section II we developed the Evidential Interpretation, and that interpretation has led us to our present impasse. On the Evidential Interpretation, Paula’s case motivates a positive account of gaslighting insofar as it parallels a version of the Dementia Case: the pathology at the core of gaslighting lies simply in the victim’s belief – or if ‘belief’ is too strong a term, in her active disposition to believe – that she is suffering from something very like dementia. The interpretation begins from the observation that Gregory treats her in a way that approximates how one would naturally treat an elderly person who appeared to display the symptoms of dementia. Looking at the treatment itself – that is, at what one says and at how one presents oneself as regarding the recipient – all sides agree that there need be no significant difference between gaslighting and the Dementia Case. And, on the Evidential Interpretation, there need likewise be no significant difference from the perspective of the recipient of the treatment. The only significant difference, for this interpretation, lies here: the victim of gaslighting is being intentionally manipulated into this predicament by the perpetrator, whereas someone who is merely treated as if she had dementia is not being thus manipulated. Though Paula may be coping with the treatment as well as she can, the same may be true in a Dementia Case, and this

has nothing to do with the nature of the pathology in gaslighting. Putting this upshot of the Evidential Interpretation bluntly, the pathology in gaslighting lies only in its status as manipulation, not in the predicament into which the victim is manipulated.

The Evidential Interpretation draws a simple comparison between gaslighting and any version of the Dementia Cases described in Section II. This comparison reveals how Paula might, despite responding reasonably, nonetheless feel that she is ‘losing her mind.’ If the dementia is real, its victim really is losing her mind, with evidence of that reality available to her in all sorts of ways. But even if it is not, she is being treated as if she is losing her mind, and she thereby receives evidence that she is from this very treatment. Either way, then, she might intelligibly come to feel that she is losing her mind, and we might regard the latter case – the second variant – as an all-too-common instance of the kind of manipulation that lies at the core of gaslighting. Whether we call it ‘gaslighting’ may rest on our assessment of the manipulator’s motives. Gregory aims at the manipulation directly – not for its own sake, but as a key instrument in his project of gaining access to Paula’s family’s jewels. A manipulative dementia case may differ not only in degree of manipulation but also in kind: perhaps the manipulative spouse acts merely from anxiety over the future of their life together, earnestly misreading benign evidence as malignant. On the Evidential Interpretation, these cases lie on a spectrum.

The Evidential Interpretation regards the interlocution in all such cases as merely in the business of generating evidence, and this feature of the proposal generates two interesting implications. One implication is that the distinctive element in the Evidential Interpretation appears to capture one side the dialectic between two opposing positions on the epistemology of testimony. On what I’ll call an ‘assurance theory’ of testimony, the epistemic warrant generated by testimony derives in part from an aspect of the interpersonal relation between testifier and



addressee that cannot be understood simply in terms of evidence made available to the latter. The Evidential Interpretation stands opposed to that explanation, supporting instead what I'll call an 'evidential theory,' on which testimonial warrant derives entirely from evidence that the testifier is speaking the truth. In saying that the warrant 'derives from' this evidence I do not mean that the addressee gets the warrant only insofar as he weighs the evidence or otherwise engages in evidential reasoning. On some 'evidential theories' (as I'll use the term), the addressee can obtain the warrant without any evidential reasoning or even any pertinent evidential responsiveness – so long the evidence exists, typically in the form of the testifier's truth-conducive reliability. On an 'assurance theory,' by contrast, there is a crucial aspect of testimonial warrant that cannot be understood evidentially, even in this extended sense of averting to evidence from which the addressee cannot reason or to which he cannot respond. One upshot of this difference is that the proponent of an assurance theory holds that a testifier may make available to an addressee an epistemic warrant that she does not at the same time make available to overhearers, even when addressee and overhearers share all relevant evidence concerning the testifier's truth-conducive reliability. Proponents of evidential theories deny this: if the evidence is the same, they hold, the testifier must make the same warrant available to testifier and overhearer alike.

We'll now pursue a dialectic about the nature of gaslighting – and therefore, I'm assuming, about the relation between testimony and memory<sup>8</sup> – that mirrors this debate insofar as the Evidential Interpretation replicates the evidential emphasis of the evidential theory of testimony. Insofar as it replicates that emphasis, it places the proponent of an alternative

---

<sup>8</sup> Remember: both interpretations target our original proposal, that the pathology of gaslighting lies in how it disables the victim's confidence in herself as a source of reasons both for others, through testimony, and for herself, in memory. That proposal constitutes the core rationale for using gaslighting as a guide to understanding the relation between testimony and memory.

interpretation in the position of a proponent of an assurance theory in this one key respect: the proponent of the alternative interpretation will argue that we cannot explain the pathology of gaslighting as a purely evidential pathology.

So how might an alternative emphasis on interpersonal relations challenge the Evidential Interpretation? Our question, again, is how to interpret our original proposal, which targets the victim's capacity to give reasons to an addressee. The Evidential Interpretation emphasizes the evidential – that is, the truth-conducive – nature of those reasons, but note that the victim's capacity to give reasons also operates in the interpersonal dimension of communicative engagement. The Evidential Interpretation places all its emphasis on how a testifier's truth-conducive reliability generates evidence that the proposition asserted is true. It places no particular emphasis on the meeting-of-minds that constitutes communication between testifier and addressee. An alternative interpretation might therefore emphasize this communicative engagement. But how could such communication engagement figure in gaslighting?

On the Assurance Interpretation that I'll develop, gaslighting marks the predicament not so much of Deborah in *Nebulous* but of the protagonist in this brief tale:

**Confusion** Paula has come to worry that she misunderstands the consequences of marriage, and specifically how entering into a marital contract changes the distribution of authority between spouses. Newly married to Gregory, Paula finds that Gregory is disposed to 'speak for' her memory – to claim authority over what she does and does not remember – in a way that confuses and troubles her. Is this like how a husband might commit his wife to a social engagement,

claiming authority over her future? With only unusual exceptions, Paula is willing to accept that the marital contract gives her husband the latter authority. But does this authority to say what she will do extend to what feels to her like a very different authority to say what she has done – and to correct her memory judgments, if it comes to that? Speaking to Brian Cameron, she finds it reassuring that he accepts her testimony with an attitude that confirms her assumption that it is no part of the marital contract that either party has authority over the other’s memory.

I do not claim that the predicament of ‘Paula’ in this brief tale coincides perfectly with the predicament of Paula Anton in *Gaslight*. While I regard this as a possible and indeed a compelling interpretation of the film, our task is not to make sense of this particular film but to see how an account of gaslighting can shed light on the link between testimony and memory. As I’ll argue in the next section, Paula’s predicament in *Confusion*, unlike Deborah’s in *Nebulous*, fits quite well with an emphasis on the link between testimony and memory that we’ve been trying to develop since Section I. Our understanding of that predicament can thereby serve as the basis of a theory of that link and – most importantly for our purposes – as the basis of an explanation of the role of assurance in the epistemology of memory.

#### *IV. Assurance as an occasion for Socratic reflection*

Our aim, again, is to understand Paula’s predicament in a way that takes at face value what I’m calling its two core elements: that her good-faith attempt to navigate it makes her feel like she’s

‘going out of her mind’ and that it cures this pathology for Brian to receive her testimony about what has happened. I’ll now argue that gaslighting works by inducing uncertainty at a deeper communicative level, not merely about one’s capacity to give reasons but about the stability of one’s conceptual commitments. I’ll introduce that deeper level by mapping it onto the issue theorized over roughly the past thirty-five years as the normativity of meaning.

In terms borrowed from Kripke’s reading of Wittgenstein,<sup>9</sup> gaslighting trades on how one might worry that there’s a ‘bend’ in the rule constituting the meaning of a term where one had not expected a bend. Paula thus worries that getting married introduces a bend in the concept of one’s authority over one’s memory: before entering into the marital contract, one has full authority (at least, in normal cases) over the contents of one’s memory, but marriage gives to a husband new and possibly conflictual authority over his wife’s memory. Kripke’s famously, and more fancifully, used as his example the bend in the concept that he invents for the term ‘quus’: the extension of ‘quus’ is identical to the extension of ‘plus’ for numbers below 57 (by hypothesis, you’ve never before added a number higher than 56) but the ‘bend’ appears when you add 57 and 68, yielding not the sum but the ‘quum’ of 5. (Kripke defines quaddition as coextensive with addition for addends below 57 but as diverging thereafter: instead of the sum for any addend of 57 or higher, quaddition generates the quum of 5.) The worry that one means *quus* rather than *plus* in one’s current application of ‘+’ (and therefore ought to answer ‘5’) could arise only in a philosophical thought experiment, but what I’m calling a ‘bend’ in one’s conceptual commitments isn’t anything esoteric. A ‘bend’ arises whenever there is a shift in the application conditions of the concept tied to the introduction of a new background context.

---

<sup>9</sup> Saul A. Kripke, *Wittgenstein on Rules and Private Language* (Harvard University Press, 1982).

Many non-esoteric concepts admit the possibility of such a ‘bend.’ Normative concepts provide especially common examples. Say you’re a man who regards all unattached women as ‘dateable,’ till you receive a promotion and encounter an unattached woman whose status as your employee makes her surprisingly *undateable* – surprisingly only because you’re clueless about how this background power differential affects the application conditions of ‘dateable.’ An aesthetic example may, by contrast, mix normative with more purely perceptual elements. Say you go well into adulthood without hearing atonal music, and you are surprised that a flat-ninth interval does not sound dissonant without a background tonal center: remove the tonality, and with it go both dissonance and consonance – a surprise to you. You may likewise be surprised by how color concepts apply in a novel context of neon lighting – not a normative matter at all. In all three cases, you may feel confused about this ‘bend,’ and pursue Socratic reflection and debate with interlocutors as you articulate and adjudicate your confusions. Or you may need no such adjudication as interlocution immediately reveals the dispositions informing your usage as out of step not merely with your fellows but with the normative contours of the concept.

What’s disorienting about being gaslighted is that one becomes confused whether the new context should have this ‘bending’ effect and generally how to go forward in one’s applications of the concept on the hypothesis that it does. Thus confused, one may benefit from a testimonial exchange that confirms how one’s conceptual commitments actually work. Since gaslighting works by exploiting the confusion for nefarious ends, the gaslighter works to prevent such a testimonial exchange from occurring. In the film, Gregory both induces the confusion and prevents Paula from seeking normal contexts of testimonial exchange. When Paula does have a normal conversation about her predicament with Brian Cameron, his receptiveness to her testimony reassures her that the rule is not bent and that she therefore does not misunderstand

what it is to have entered into the relation that she now bears to Gregory. With her conceptual commitments thus clarified, she can address her predicament constructively, by confronting Gregory with the contents of her memory in full confidence of her authority over them.

What Paula gets from Brian is a reassurance that derives from his responsiveness to the assurance that informs her testimony. I am happy to say that she gets reassuring *evidence* from his responsiveness; what distinguishes the Assurance Interpretation from the Evidential Interpretation, is how the former characterizes what he gets from her testimony. To mark this distinction between what Brian gets and what Paula gets, I'll say that Brian gets an 'assurance' from Paula, whereas Paula gets a 'reassurance' from Brian. But again, I claim only that the assurance given by Paula cannot be understood in terms of evidence, not that the reassurance given by Brian cannot be. To defend the Assurance Interpretation, we need to show that the Evidential Interpretation cannot make good sense of the narrative in Confusion. The Evidential Interpretation cannot explain how Paula could get this reassurance, I'll argue, because she could not get it from Brian's response if the only thing her testimony gave him were evidence.

It will help if we frame the issue in terms that parallel Kripke's treatment of the bent concept *quus*. Here, again, is Kripke's definition of 'quus':

Quus             $x$  *quus*  $y$  = (a) if  $x$  and  $y$  are both less than 57, then  $x$  plus  $y$ ; (b) if at least one of  $x$  and  $y \geq 57$ , then 5.

Now consider this definition of what I'll call 'quauthority':

Quauthority     $X$  has *quauthority* over  $Y$  = (a) if  $X$  and  $Y$  are not married, then  $X$  has relational authority over  $Y$ ; (b) if  $X$  and  $Y$  are married, then  $X$

has relational authority over *Y* *plus* authority over *Y*'s memory judgments.

By 'relational authority' I mean the normal sort of authority that people have over each other's actions and commitments when they are in an established relationship and are, as we say, sharing their lives. Nothing turns on how we define this, except that we can see that no one would think that people who are sharing their lives in this ordinary sense, even as a married couple, have authority over their memory judgments. We can see this, but Paula fears that it may well be true – and, even more alarmingly, that she herself has all along been exercising a concept of marital authority with this implication. This fear creates the need for reassurance from an interlocutor like Brian. Will he treat her as exercising the concept of authority or the concept of quauthority? If the conceptual commitments that inform her testimony align her with the former, then Gregory's efforts to 'speak for' her memory judgments can be justly criticized as an illicit claim of authority. But if those commitments align her with the concept of quauthority, then those efforts are licit on their own terms and can be justly criticized only on broader grounds – for example, on the grounds that the whole concept of quauthority is unjust. She looks to Brian for help in settling this question about her conceptual commitments, and she is interested specifically in how he responds to her testimony as she narrates her struggles with Gregory's claim of authority. In a different setting, she could ask him directly how to define the concept that they share, but such an explicit Socratic exchange is, in her present circumstances, practically impossible. Like most of us most of the time, she pursues such Socratic reflection implicitly, by seeking a reflected image of her commitments in the response of an addressee to her speech acts.

Consider how the exchange would have to work on the evidential model: Paula's testimony would give Brian evidence that *p*, and his responsiveness to this evidence for *p* would

have to give her evidence in turn that she is exercising the concepts that figure in  $p$ , including specifically the straight version of the concept at issue, rather than the bent version. The problem is that the latter evidence – the evidence deriving from Brian’s response – is sensitive to the specific difference between the bent concept and the straight concept only insofar as it is supplemented by evidence that he affirms or disaffirms  $p$  – that is, by evidence that he agrees with her or evidence that he does not agree with her on the question whether  $p$ . If she has evidence that he agrees with her that  $p$  – that is, that he affirms both that  $p$  and that, in so affirming, he agrees with her testimony – then Paula gets evidence that she is exercising the concepts that figure in  $p$ , including (for example) the straight concept *authority* rather than the bent concept *quauthority*. She would get the latter evidence for the twofold reasons that (i) Brian’s affirmation that  $p$  would generate evidence about the truth conditions of  $p$  and thus of the application conditions of the concepts that figure in  $p$  and (ii) his agreement with her assertion would generate evidence that she is asserting that  $p$ . For Paula to get the evidence she seeks, on the Evidential Interpretation, that evidence would have to lie in what would amount to a discussion between them, with a burden on her to understand whether Brian agrees or disagrees with her assertion. But that is not where she is seeking the reassurance she seeks.

Paula is not looking merely for evidence that she means the straight concept. Such evidence could take many forms, but in every such form but one, interpreting the evidence will require substantial background assumptions – assumptions that, like the assumption that Brian agrees with her, Paula will need additional evidence to vindicate. The one form of reassuring evidence that does not require substantial background assumptions lies in the uptake that Brian provides when he merely consummates her illocutionary act (in Austin’s metaphor). To say that this evidence does not require substantial background assumptions is not to say that she needs no



evidence that Brian is in the room, that he hears her, and so on for everything that makes Brian's uptake an empirical reality. What's crucial is that the reassuring evidence be evidence simply that the empirical reality of his uptake has occurred. How could it give her the reassurance she seeks simply to appreciate evidence that her illocutionary act has been consummated? It would give her the reassurance if the consummation consisted in his taking her word for it that p by trusting her. By 'taking her word' I don't mean believing her. Believing her – that is, being convinced by her testimony – is a perlocutionary, not the illocutionary, aim of her speech act: it is not what she presents herself as aiming to do *in* telling him that p. What she presents herself as aiming to bring about in telling Brian that p is that he treat her testimony that p as giving him a reason to believe that p. I argue elsewhere that she aims to give him a reason that would count as sufficient or conclusive in his context, but that further claim is not crucial to my present point.<sup>10</sup> What's crucial is that, even if she aims to give him a conclusive reason, she need not also aim that he acknowledge that reason *as* conclusive and thereby form the belief. Again, the former aim is her illocutionary aim, the latter her perlocutionary aim. He satisfies her illocutionary aim by acknowledging the reason – that is, by trusting her to the extent that he treats her as giving him the reason – even if he does not also acknowledge the reason *as* conclusive for him. But in getting evidence that he trusts her for this reason, she gets evidence that he understands the content of her testimony as she understands it – that is, as containing the straight rather than the bent concept of authority at issue – as containing *authority* rather than *quauthority*. She gets this evidence not from background assumptions and evidence that supplement the uptake through which he consummates her illocution but from that uptake itself.

---

<sup>10</sup> My account in "Assurance and Warrant" builds on my account in "Telling as Inviting to Trust" by emphasizing that the speaker aims to give her addressee a reason that would count as conclusive in his context.

I'm building here on the assurance theory of testimony that I have defended at length elsewhere.<sup>11</sup> My strategy is to apply that theory in developing an epistemology of memory that emphasizes a parallel between memory and testimony. The parallel that I propose rests on the Assurance Interpretation of gaslighting, and the core of that defense lies here: Paula seeks reassuring evidence that derives from Brian's mere consummation of her speech act. There is no room, on the Evidential Interpretation, for the idea she gets such reassuring evidence from his mere consummation of her speech act, because there is no room there for the idea that he accepts her testimony as giving him a reason to believe that *p* grounded in part in her assurance – except, of course, insofar as that assurance gives him pertinent evidence. Again, I'm happy to say that Brian's responsiveness gives Paula reassurance by giving *her* evidence about her conceptual commitments. The problem with appealing to evidence does not lie there but in the claim that *she gives him* nothing but evidence for *p*. If she gives him nothing but evidence for *p*, then she does not give him the assurance that could, as we're understanding it, be reflected in his consummating response to it, in turn giving her the evidence that she seeks about her conceptual commitments. In seeking to understand her conceptual commitments as reflected in an addressee's receptivity to her 'her word' – in her addressee's willingness to 'take' her word – Paula implicitly pursues what I would theorize as a Socratic dimension of concept possession. It forms part of the Socratic dimension of concept possession that you monitor your usage through responsiveness to how others are receptive to your concept-expressive acts. Do they treat you as giving them a reason grounded partly in your assurance? If not, how might this failure of illocutionary consummation lie in a specific failure to communicate by sharing a repertoire of concepts? As we'll see in the next three sections, we cannot understand this 'meeting of minds'

---

<sup>11</sup> See the papers cited in note 2.

simply as the ‘preservation’ of content across perspectives, as if all one had to do to consummate a speech act is to grasp a shared content. In both the interpersonal case and the intrapersonal case that we’ll now discuss, it is a normative accomplishment to share a content in this way.

#### *V. Content preservation and the dynamic stability of belief*

The Assurance Interpretation thus explains the interpersonal dynamic through which Paula’s sensitivity to how her testimony is received can help cure her conceptual confusion. But what of the intrapersonal dynamic informing her memory judgments? Our shift to this conceptual dimension, with its emphasis not primarily on the act of judgment but on grasp of the conceptual rule that informs judgment, may appear to have left behind Paula’s concern with the reliability of her memory as she confronts the questions whether and how to judge and therefore what, if the question arises, to tell others about what she remembers. Have we thereby changed the subject away from the primary link between memory and testimony? We have not, for the simple reason that talk of rule-following is always implicitly talk of memory.

The film may obscure this connection by its focus on Paula’s experiential memories. Gaslighting as a real-world interaction need not target the victim’s experiential memories, as the interaction between Gregory and Paula does in *Gaslight* the film. But it does always target the victim’s propositional memories. In *Gaslight*, it targets Paula’s beliefs about what being in love and being married will be like in many respects, beliefs formed earlier in her life that she has no reason to question till Gregory’s manipulation begins to falsify them. But she wonders: *Is* he falsifying them? Or is he instead revealing that her concepts do not apply in the way she thought they did? There is no dramatically effective way to depict this contrast in a film, so the film

targets not Paula's grasp on a remembered concept but her experiential memories themselves. Instead of 'I formed a belief that *a* is F, but F-ness doesn't appear to be what I thought it was' we get 'I vividly remember that I did such-and-such, but it appears that I didn't do such-and-such.' On the Assurance Interpretation, these experiential memories trouble Paula not only because they implicate her in a mistake about her experiences but also because they appear to show that she does not understand how her own concepts describe them.

When you apply a concept by forming a belief or making an assertion, you thereby undertake or present yourself as undertaking inferential commitments whose content depends on how that concept relates to concepts that you have previously applied and can remember having applied in beliefs that you retain from those prior applications. Let me elaborate one of the examples mentioned in Section IV. As a jazz musician, you know how to improvise over a flat-ninth interval when it adds dissonance to a dominant seventh chord – know-how informed by beliefs that you have retained from college classes in music theory – but now that the harmonic context is completely atonal, how do those memory beliefs apply? We needn't picture you responding in performance but reflecting on your improvisatory strategy in rehearsal. If this marks a mere 'bend' in the application-conditions of the concept of dissonance, then you'll have to figure out the right inferential relations between those memories and your present context of application. If, on the other hand, this new context does not mark such a bend, then your dispositions to call it 'dissonant' will bear a different sort of relation to those memory beliefs. Perhaps the term would now express an entirely different concept, or perhaps its present use would count as a metaphorical application, with correspondingly looser inferential relations between the memory beliefs. We can imagine similar circumstances for the other bent concepts that we considered. In each case, the question directly or implicitly targets your memory: how

would a present application bear on the inferential content of past applications, as retained through your memory in the form of stable beliefs?

Applied intrapersonally to the stability of belief, the assurance model generates a new way to think about cases that Tyler Burge theorizes in terms of ‘content preservation.’<sup>12</sup> These are cases coined as counterexamples to semantic-externalist or anti-individualist theories of meaning as ‘slow switch’ cases, though I’m going to call them ‘shift’ cases, since they address how meaning responds to a shift in broader context. In a standard example, your context at an earlier time determines that you refer to H<sub>2</sub>O by ‘water’, but a shift in your environmental context ensures that you later refer to a different chemical XYZ by ‘water’. When you make a judgment that you would express by saying ‘There was some water,’ aiming to ground that judgment in a remembered earlier judgment that you would then have expressed by saying ‘There is some water,’ you’ve gone astray: you aim to ground a judgment about XYZ in a remembered judgment about H<sub>2</sub>O.<sup>13</sup> This appears to undermine the assumption that your ‘water’-beliefs are stable across such a shift, and if they are not thus stable you appear to lose your entitlement to reason from them. Burge aims to vindicate the stability of belief by conceding that if you are not reasoning from a belief across a shift but instead merely attempting to re-report it, then you may go astray in this way. But if you are reasoning from your earlier belief, he argues, then you are entitled to the presumption that it has remained stable. Burge argues that in such a case we aren’t usually describing or reporting or commenting on what we thought – at least, not most fundamentally. What are we doing then? Burge argues that at the later time we are typically *enacting* or *following through* on what we thought at the earlier time.

---

<sup>12</sup> See the papers cited in notes 13 and 14 below.

<sup>13</sup> To save space, I’ll work from this rough formulation of the problem. For a more refined formulation, see Paul Boghossian, “Content and Self-Knowledge,” *Philosophical Topics* 17 (1):5-26 (1989).

When we think, Burge argues, our perspective is typically the perspective of critical reasoning. You cannot reason unless you undertake specific conceptual commitments – commitments that essentially have implications for future and past times. To think about your thoughts is to think about these commitments as such – that is, to commit or uncommit yourself to them. Sure, you *can* step back and merely report them. If that’s all you’re doing, then there is room for a skeptical worry. But typically that won’t be all you’re doing. Disengaged self-commentary is not what we’re typically up to when we report our thoughts. We’re *engaged* in thinking. We aren’t merely reporting on our thinking in the way that we might report on someone else’s thinking. Burge’s larger argument is that preservative memory is a necessary condition for our being able to be engaged self-critical thinkers.

This species of intrapersonal criticism parallels a critical stance that you adopt when you reason with others. I agree with Burge on the general parallel, though I gloss it with the assurance model’s distinctive emphasis on the Socratic dimension of concept possession. Though there may be no actual criticism in play, I agree with Burge that when you reason from an earlier belief you commit yourself to the possibility of an intrapersonal species of criticism that parallels, in some key respects, the potentially critical perspective on another that shapes how you can accept their testimony as giving you a reason irreducible to evidence of its truth. (I lack space to discuss Burge’s complex argument for content preservation in testimony.<sup>14</sup>) Though my emphasis on assurance is crucially different from Burge’s emphases (as we’ll see presently), the possibility of criticism in each case presupposes that the concept with its conceptual commitments is stable across these intrapersonal or interpersonal perspectives.

---

<sup>14</sup> See “Content Preservation” (1993), “Interlocation, Perception, and Memory” (1997), “Computer Proof, Apriori Knowledge and Other Minds” (1998), all reprinted in Tyler Burge, *Cognition through Understanding: Philosophical Essays, Volume 3* (Oxford University Press, 2013). I do discuss Burge’s view of testimony in “Telling as Inviting to Trust.”

The parallel is implicit in a shift case, but it becomes especially salient in the sort of case that Paula fears she's in, which we're calling a 'bend case.' A bend case more directly addresses how meaning is determined by one's context of ongoing use. In Kripke's cartoonish version, you've spent your life using '+' to mean a function coextensive with addition for numbers smaller than 57, never having added a number that large, and you now conclude your arithmetical deliberation '58 + 67 = what?' by answering '125.' Is your answer correct? It is correct if you mean *plus* by '+'. But it is not correct if you have all along meant by '+' not *plus* but *quus*, in which case you should have answered '5.' In Paula's much less cartoonish version (apart from the silly 'q'-term we've coined), she judges that Gregory lacks 'authority' to question her memory but fears that by 'authority' she means not *authority* but *quauthority*, which Gregory does not lack because she is now his wife. She does not worry that her meaning has shifted from what she used to mean by 'authority' but that she has all along meant *quauthority* – that is, used a concept with those conceptual commitments, commitments which make this key difference to how she conceptualizes her own 'authority' relations only now that she and Gregory have wed. Though there is no shift, fear of a bend in conceptual commitment functions as if she feared a shift. In each case – whether bend or shift – the challenge lies in knowing where you stand vis-a-vis your own past conceptual usage. Are you entitled to presume that you understand how the grasped conceptual content of your past usage continues to commit you in present usage? For our purposes, the key difference between a shift case and a bend case does not much matter. If your presumption is false in a shift case, the falsity lies in how your present commitment differs from your past commitment. But if your presumption is false in a bend case, you were all along committed in this way that falsifies your presumption. What matters for our purposes is the possibility that your presumption is false – whatever else that entails.

How do you maintain the stability of a belief that vindicates your presumption to understand how your present conceptual commitments are grounded in your past conceptual commitments? I answer this question in the next two sections, where I also confront the inevitable question: how could an assurance function intrapersonally? Let me first emphasize that the assurance model's answers will appeal to a complex and dynamic intrapersonal relation. Even though it's just you there and no speech act is in play – which points to a key disanalogy that we'll explore in Section VI – this intrapersonal relation partakes of a complexity that parallels the complexity that we explored in the interpersonal case. On the broad approach that I share with Burge, we explain the stability of belief by explaining how a belief can remain stable across a change in context that might have made it go unstable through a shift or a bend. For Burge, our entitlement to reason in a way that presumes stability derives from very general considerations about how the self-critical stance that informs a capacity to reason rests on assumptions about the nature of selfhood and the capacity for first-personal reference. (As with Burge's treatment of testimony, I lack space even to do begin to discuss this view here.<sup>15</sup>) My very different assurance model uses the parallel between memory and testimony to develop a model of the stability of belief that grounds this stability in a normative stance that functions as an intrapersonal assurance. It follows that a proponent of the assurance model cannot rest with the observation that conceptual content is simply 'preserved' through memory. Conceptual content is no more simply 'preserved' through memory, on the assurance model, that it is through a testimonial relation. In each case, a Socratic dimension of the exchange makes the stability of belief more dynamic than that. I agree with Burge that the epistemic and conceptual

---

<sup>15</sup> See "Content Preservation" (1993), "Memory and Self-Knowledge" (1996), "Our Entitlement to Self-Knowledge" (1996), and "Reason and the First-Person" (1998), in Burge's *Cognition through Understanding*.



dimensions of ‘content preservation’ – that is, of doxastic stability – go hand in hand. But I disagree in holding that doxastic stability is a normative accomplishment in both dimensions.

### *VI. What is intrapersonal assurance?*

How could there be a purely intrapersonal assurance – an assurance without speech or any other public act? To see how memory rests on an intrapersonal assurance, we must invert the temporality of memory, switching our perspective from that of the later remembering self looking back to that of the earlier judging self looking forward. The key to understanding intrapersonal assurance lies in understanding the intrapersonal normativity of judgment. I’ll now argue that the intrapersonal normativity of judgment rests crucially on the Socratic element that we isolated in our treatment of the interpersonal normativity of testimony. The key to understanding how judgment functions as an assurance is to see who it is that’s being assured.

Once again, it helps to consider the phenomenon of gaslighting – now from the perspective of the victim’s capacity for judgment. How does a victim of gaslighting lose confidence in her judgment? As we’ve seen, gaslighting exploits the vulnerability built into a key dimension of rational responsiveness: your responsiveness to your own status as trustworthy when you judge. Since there are not literally two people in this relation – trusted judge and trusting recipient of the judgment – we need to exercise caution in how we frame it. Moreover, any proposed parallel between judgment and testimony encounters a more immediate problem: as we use the expression, ‘telling yourself that p’ does not entail judging that p. If you’re acrophobic, and you’ve unfortunately walked out on the Skywalk over the Grand Canyon, you might ‘tell yourself’ that the structure is safe – whispering to yourself “It’s safe, it’s safe” –

precisely because your acrophobia prevents you from straightforwardly or simply judging that it is safe.<sup>16</sup> Whatever intrapersonal assurance amounts to, it therefore cannot be simply the internalization of a speech act. What else then might it be? As we'll now see, the key to understanding the normative stance of judging that p lies in how it projects an audience whom you *could* sincerely tell that p. The parallel lies not with any speech act of 'telling' or assuring yourself but with the speech act of telling and thereby assuring possible others.

We can use the locution 'telling myself that p' to develop a case in which, unlike Paula's, a disrupted judgment does not indicate pathology. A Google search for the locution yields this newspaper report on a teenager whose breast cancer was diagnosed as benign, leading her to judge it benign – until a 'feeling of unease' began to wear away at that judgment.

*Elisa's Cancer.* Over the next few months, Elisa's life went on as normal. She carried on at uni, spending long afternoons rummaging through shops, building up her record collection. But all the while, there was a feeling of unease that she couldn't shift. '*Every day, I kept telling myself it was nothing,*' she says. '*But every day as I dressed I'd feel the lump there. I had to keep reminding myself that the doctors had done tests and they'd said it was fine – so how could it not be?*'<sup>17</sup>

Does Elisa judge that the lump in her breast is nothing – that is, not cancerous – when she "tells herself" that it is nothing? (As the report makes clear, her unease is correct, as it turns out: the

---

<sup>16</sup> Tamar Gendler uses a case with roughly this content to motivate a different but related idea: that your acrophobic stance manifests a sub-doxastic state that she calls an 'alief.' (See her "Alief and Belief," *Journal of Philosophy* 105:10 (2008), 634-663; and "Alief in Action (and Reaction), *Mind and Language* 23:5 (2008), 552-585.) My idea targets instead your anti-acrophobic state, which no longer functions as guiding you in the way that we mark with the word 'judgment.' (Readers who have the intuition about Gendler's case that your acrophobic state runs contrary to your judgment – that is, to a judgment that you've made and retain – should think of this as a different case from Gendler's. Here your acrophobic really does prevent you from judging that the Skywalk is safe.)

<sup>17</sup> *The Independent*, 27 April 2010; italics added.

diagnosis turns out to be mistaken.) The locution suggests that while she somehow thinks that she ‘should’ judge that the lump is nothing, that judgment is not one that she actually makes. It suggests that she is trying to ‘talk herself into’ believing that the lump is nothing – but unsuccessfully, given the doxastic dispositions informed by her “feeling of unease.”

How has Elisa wound up in this state? The newspaper reports that, relieved by testimony from medical authorities, she begins by confidently judging that the lump is nothing, only to see her daily encounters with what turns out to be the lump erode that confidence. As her confidence diminishes, she passes through a state in which she continues to judge that the lump is nothing but acts in some ways that seem to indicate that she does not believe that the lump is nothing. Finally, she winds up merely ‘telling herself’ that the lump is nothing, which figures as a halfway house en route to abandoning this judgment. The transition thus takes this form: her final state of ambivalence resolves the akratic irrationality that marks the middle stage in which she loses confidence in this judgment that she has not yet abandoned. Schematically put, Elisa shifts between two poles. Earlier she counts as irrational insofar as she judges that *p* but is not doxastically disposed to act as if *p*.<sup>18</sup> (She is, of course, disposed to act as if *p* in some respects. The problem is that she is not so disposed in other crucial respects – despite judging that *p*.) Then later Elisa ceases to be irrational because she no longer judges that *p* but merely ‘tells herself’ that *p* from a stance of non-committal ambivalence on the question whether *p*.<sup>19</sup>

---

<sup>18</sup> By a doxastic disposition, I mean a disposition that serves to structure and guide practical and theoretical reasoning in the way that a belief does, but without necessarily manifesting the subject’s judgment.

<sup>19</sup> Is this ‘upstream’ reasoning in the sense discussed by Niko Kolodny in “Why Be Rational?” (*Mind* 114 (2005), 509-563)? Kolodny argues that ‘upstream’ reasoning is impossible, and I have attempted to counter this argument in “Rational Requirements and ‘Rational’ Akrasia, and “On the Risks of Resting Assured.” But I need not claim that abandoning your judgment in the spirit of self-mistrust amounts to ‘upstream’ reasoning for any purpose served in the present paper. For present purposes I take no stand on whether the rational requirement linking judging that *p* with believing that *p* has narrow scope or wide, or on whether it is a state requirement or a process requirement, in the senses discussed by Kolodny.

How does Elisa's predicament compare with your predicament as an acrophobe on the Grand Canyon Skywalk? Since your contrary doxastic dispositions brought you to a state in which it is more accurate to say that you were merely 'telling yourself' that the Skywalk is safe, the Skywalk case is precisely similar to Elisa's Cancer in this respect: you are now ambivalently undecided on the question, since you no longer count as judging that the Skywalk is safe. What explains the dynamic? Simply this: you thereby cease to persist in the irrational state of doxastic akrasia wherein you judge that p without being doxastically disposed to act as if p.<sup>20</sup> Despite this parallel between Elisa's Cancer and the Skywalk case, there is one very important difference between them: Elisa is not irrational to resist the force of her own judgment, and arguably would not have been irrational to resist it even if she had in the end been proven wrong, whereas you on the Skywalk clearly are irrational. How should we explain this difference between the cases? I think the difference lies here: Elisa abandons her judgment that the lump is benign when she can no longer imagine herself confidently and sincerely telling others that it is benign, whereas you abandon your judgment that the Skywalk is safe without any such projection into a context of possible testimony. It is to Elisa's credit that her inability to feel entitled to tell others – marked by her stuttering failure to assure herself ("I kept telling myself") – leads her to abandon her judgment that the lump is benign. But it is not to your credit that you cannot succeed in assuring yourself ("It's safe, it's safe," you mutter to yourself) despite feeling perfectly entitled to give this assurance to others.

How do both these cases compare to Paula's predicament in *Gaslight*? Note first a parallel in how manipulation of the normative structure might generate pathology. Where a

---

<sup>20</sup> Like Elisa, you are doxastically disposed to act as if p in some respects. The problem is merely that you are not so disposed in other respects – despite judging that p.

relation to an interlocutor, real or imagined, can serve as a proxy for your intrapersonal relations, your relation to that interlocutor, again real or imagined, makes you vulnerable to exploitation. Imagine that Elisa practices a religion that preaches mistrust of medical authorities. On that basis alone, Elisa might find herself unable to imagine interlocutors whose trust she could sincerely and reasonably invite in testifying that the lump is benign. It would make her predicament worse if she were married to such a coreligionist. This case could share the normative structure of Elisa's Cancer but with the normative valence of the Skywalk case: an irrational mistrust in her own judgment now mediated by interlocution rather than acrophobia. That is the combination of structure and valence that defines Paula's predicament. Paula is not the prisoner of her memory judgments. Under the different circumstances that we imagined in Section II, she might come to doubt her memory in a way that, like Elisa's actual case, confirms her status as fully rational. But that would presuppose a capacity to imagine non-exploitative interlocutors. In the film she does not gain access to a non-exploitative interlocutor until Brian Cameron elbows his way through Gregory's defenses. Brian's appearance revives Paula's imagination in a way that resuscitates her judgment precisely where Gregory's stranglehold had led it to fail. She can now imagine telling interlocutors what she would like to tell them – about the broach, about noises in the attic, about the gaslights – but had lost a feel for her capacity to tell them through Gregory's manipulative insistence that she tell all such things only to him.

The parallel rests on a common element in all three cases: each illustrates how a capacity to imagine oneself testifying constrains the rational normativity of judgment. I elsewhere argue for this broad idea from several disparate angles.<sup>21</sup> Here I'm arguing from the role of projected

---

<sup>21</sup> For distinct arguments for this thesis, see my "Receptivity and the Will," *Noûs* 43:3 (2009); "Reflection, Disagreement, and Context," *American Philosophical Quarterly* 49:2 (2012); and (most fully) "Assertion, Sincerity, and Knowledge," *Noûs* 47:4 (2013).

stability in epistemic agency. What cases like these reveal, I believe, are the real-world stakes of the doxastic stability projected by judgment, and how those stakes place a burden on the one who judges. When Elisa merely ‘tells herself’ that the lump is benign, she has reached an impasse in exercising her capacity to project a stable future for the belief that it is benign, and the impasse involves the projection of an interpersonal relation because the parallel intrapersonal relation stretches into a future that has not yet come to pass. Doxastic stability requires of her that the invitation to trust at the core of her judgment project stably into the contexts both of future selves who will accept the invitation by reasoning from or acting on the judgment and of present interlocutors who will accept by believing her testimony. The dual-sided nature of doxastic stability permits her to use the synchronic testimonial case as a substitute for the diachronic case in which she judges only for herself. It not only permits that substitution but requires it, since the two sides stand in what we might regard as a relation of normative equivalence. One way to grasp the equivalence is to observe that if you invite another to believe on the basis of your testimony, you thereby also invite that other to project a future in which that belief will prove doxastically stable. If you cannot sincerely and coherently do the latter then you cannot sincerely and coherently do the former. (You could, of course, testify insincerely.) An interlocutor can thus serve as a proxy for your own future self in purely intrapersonal projection.

The basis of the parallel between these two projections – into the contexts of imagined interlocutors and into the contexts of your own future selves – lies in how both projects invite trust. As we’re imagining the cases, the subject is not trying to imagine that these interlocutors will treat her testimony merely as a source of evidence, believing what that evidence dictates, but that they will trust her – that, even if she also gives them good evidence, they will more fundamentally believe her, by taking her word. Therein, I’m claiming, lies its point: to imagine

how these interlocutors would react to her testimonial assurance. Strictly speaking, as I emphasized in Section IV, she is trying to imagine that she has satisfied her illocutionary aim in testifying, giving them good reason to believe what she tells them, not that she has succeeded in realizing the perlocutionary aim of convincing them. The point of this imaginative exercise derives from how these reactions would parallel – and could therefore substitute for – a reaction from her own future self: both projections amount to invitations to trust. It is here that we see the most fundamental parallel between testimony and memory: the remembering self exercises a capacity for reasonable trust, in response to the invitation to trust that informs judgment, just as an addressee who believes a speaker exercises a capacity for reasonable trust by accepting the invitation to trust that informs the speaker’s testimony. Let me again emphasize both that judgment is not a speech act and that trusting your judgment is not the same as trusting a speaker’s testimony. As we’ve seen, you can be stuck ‘telling yourself’ that p in part because you don’t simply trust your judgment that p. All I’m saying is that an invitation to trust lies at the core of both acts – at the core of both judgment and testimony.

### *VII. Assurance as monitoring a normative constraint on judgment*

Having thus explained the possibility of intrapersonal assurance, my argument confronts a new objection. Someone might accept both my view of testimony and my claim that imagined testimony plays the above-described role in constraining judgment and yet resist the idea that judgment invites trust, thereby also resisting the idea that memory embodies self-trust. Why then should we view judgment as an intrapersonal assurance? The most fundamental reason replicates our earlier dialectic, through which we developed the assurance model for testimony. We can view these intrapersonal relations with an assurance-theoretic emphasis on conceptual

normativity as follows. By the partial parallel with Elisa's Cancer, what disables Paula's judgment is her doubt whether the apparent memories informing it are sufficiently reliable to serve as the ground of an assurance offered to others. As we saw in Sections II and III, this yields dialectical space for an Evidential Interpretation, on which Elisa and Paula merely register their responsiveness to evidence of their unreliability. We can resist that interpretation with an extension of the assurance-theoretic reframing maneuver that we developed in Section IV.

On such an Assurance Interpretation, the full story of Paula's affliction must also include her fear that a bend in the concept that she expresses with a term like 'authority' will prevent others from receiving her assurance in trust. Although she may *also* be interested in evidence of her truth-conducive reliability, which may – as far as the present point goes – be generated as conceived by the Evidential Interpretation, she is more fundamentally interested in evidence that others use the term without the bend that she fears it may contain. Her interest in this evidence most fundamentally takes the form of her disposition to project an audience of imagined interlocutors for her judgment – say, that Gregory is abusing his authority – who will receive the testimonial expression of that judgment in trust, by taking her word for it. This disposition to project then runs aground on Paula's metalinguistic anxieties: what if these others do not believe her because they think that, as Gregory's wife, she has no right to issue such a complaint? Or something like that: her anxiety is amorphous and cannot be reduced to a single propositional content. Formulating it with metalinguistic content, she is anxious that addressees will regard the relevant concept of authority as undergoing a 'bend' in marriage and *therefore* will regard him, as her husband, as not lacking in authority to challenge her memory judgments in the way he has done. Because she cannot feel confident that this imaginative projection will be confirmed by actual testimony she might give – not because there will be no such actual



addressees, but because, if there are, they will not react as depicted by the projection – she loses confidence in her capacity to form judgments using a term like ‘authority’ in this intended usage. Her loss of confidence leaves her just as Gregory wants her: silenced because unable to judge.

Paula’s case thus goes deeper than Elisa’s cancer or your acrophobia, both of which do admit of an Evidential Interpretation, since in each of the latter the subject can be viewed (if one wants) as most fundamentally doubting her or your truth-conducive reliability. We cannot fully understand Paula’s case, by contrast, unless we treat her self-doubt as extending to her status as a concept-user. As I’m interpreting the film, Paula does doubt the truth-conducive reliability of her memory judgments, but she also, and more fundamentally, doubts the dispositions informing her use of some key concepts in the judgments that frame the controversy over those memory judgments. I think it’s better to treat all three cases as containing both dimensions, but it is only Paula’s case that forces us to acknowledge the conceptual dimension of self-reflection. Again, we are forced into that interpretation only if we are struck by the aspect of Paula’s predicament that I have emphasized: that her pathology of memory appears cured as soon as, and insofar as, she feels the force of Brian Cameron’s trusting receptivity to the assurances that she gives him in testimony. (In the film this happens twice. He leaves her in the final scene, Gregory returns, and she relapses into self-doubt. But Brian’s return as reassuring addressee re-enacts her cure.) Paula has worries about her memory as a guide to truth, but her core worry targets her grasp of the concepts that inform the judgments that she expresses in this testimony. When he accepts her assurances, she receives reassuring evidence that her grasp is solid.

This Assurance Interpretation of the intrapersonal case returns us to the topic of Section V, since we see here how the two dimensions of doxastic stability marked by Burge’s term ‘content preservation’ – the purely conceptual dimension and the properly epistemic – fit

together. A focus on the pathological disruption in Paula's predicament and the non-pathological disruption in Elisa's helps us see how it is far too simple to describe stability as a matter of something's being 'preserved.' The stability of belief is a normative accomplishment. Though not itself a speech act, judging that p projects both a possible interpersonal recipient of the act of testifying that p and an intrapersonal recipient who will receive the judgment in memory. This intrapersonal recipient is not an addressee, since judgment is not as such a speech act. But the intrapersonal recipient's responsiveness to evidence of your untrustworthiness as you judge constrains your intrapersonal invitation to trust in the way that your beliefs about how an addressee will receive it may constrain your testimony. That is why judgment projects this audience. While the constraint may engage your truth-conducive reliability on the question whether p, as you assess your reliability from this addressee's perspective, it may equally – and more fundamentally – engage your capacity to communicate with the addressee through exercising a shared repertoire of concepts. We need an assurance model of judgment and memory because we must explain how you monitor your conceptual commitments through monitoring your addressees' receptivity to the assurance that you give them through inviting their trust. The mechanism of this monitoring reveals how we can pursue Socratic reflection in the medium of judgment without speech. While you can pursue Socratic reflection in private, the assurance model provides a novel basis for ruling out 'private language.' The dynamic of assurance and reassurance is inherently communicative. The intrapersonal role of assurance in judgment and memory does its normative work by mimicking its interpersonal role in testimony.