Free & Equal | A Journal of Ethics and Public Affairs

Schroeder, Mark 2025 Tipping Points: Abuse and Transformative Discovery. *Free & Equal: A Journal of Ethics and Public Affairs*, 1(1): 1–35. DOI: https://doi.org/10.16995/fe.17907

Open Library of Humanities

Tipping Points: Abuse and Transformative Discovery

Mark Schroeder, Philosophy, University of Southern California, US, maschroe@usc.edu

This paper explores how philosophical accounts of the nature of persons and attributive responsibility can help us to make sense of the kinds of characteristic errors that people make in interpreting what is attributable to one another. I show how this gives us an important tool for understanding some important kinds of interpersonal conflict, with particular attention to understanding why Maya Angelou's advice that "when someone shows you who they are, believe them the first time" can seem easier to have followed from the privileged perspective of retrospection than it really was in prospect.

Tipping Points: Abuse and Transformative Discovery

Mark Schroeder

I. INTRODUCTION

This paper is an exercise in exploring how philosophy can help us to understand and navigate the dynamics of interpersonal conflict. It is motivated by a conviction that interpersonal conflict is by definition conflict between *persons*, that conflicts accelerate and deepen when they become more *personal*, and that it is to philosophy that we must turn, to understand what it means to be or be treated as a person. But this conviction can only be tested by its fruits. So let's go look for some of those.

I.A Transformative Discovery

Sylvia has been married to her husband, John, for over forty years.[1] They have children together, and grandchildren—a busy and complicated and intertwined life. But their marriage is not perfect. Far from it. John is verbally, emotionally, and physically abusive toward Sylvia. He keeps mistresses and flaunts that in Sylvia's face, threatening to humiliate her publicly. He taunts her about her weight issues and appearance. He uses control over money as leverage. He mocks; he yells; he hits. And over time all of this has gotten worse.

Sylvia always understood that John was not perfect, and their relationship far from it. But in the last ten years she has come to understand this in a new way. Before she thought that she had an imperfect marriage with an imperfect man, but that the core of it was good and there was room to improve. But one day she woke up and saw things differently, and it devastated her. Sylvia now understands that John is irredeemable—that he never truly loved her but only saw her as a prop, and that his past actions reflect not imperfection, but, in her words, "evil." Some nights she lies awake revisiting events that happened early in their relationship whose significance she now disorientingly feels like she completely misunderstood until now. It is unsettling—even debilitating—to feel like she had so little grasp at the time of what was happening to her in her own

[1] Names in this case have been changed to protect identities.

life. But she is still married to and lives with John, and has no plans to change that—indeed, her life plan is now simply to outlast him.

Sylvia's story is just one of millions. It is easy to think that Sylvia is making a mistake to stay with John. However trapped she feels in this relationship that she now understands is irredeemable, the costs of exit are not as high as they now appear to her from the inside, while John's contemptuous words are gnawing inside her head. Whatever financial leverage John holds over her is not worth her dignity, her children and grandchildren will understand and support—indeed, applaud—her for leaving him, and the fallacy of sunk cost reasoning could never be clearer. Women—and it is important to remember not just women, but it is also important to acknowledge often women—stay in abusive relationships for a wide variety of complex reasons, and for different reasons at different times.

But in this paper I will be interested not in Sylvia's current choices, but in how she got here. I will be interested in how Sylvia ever stayed in her relationship with John for long enough for there to be sunk costs, complicated family considerations, or financial leverage. I will be interested in what light we can shed on the power of abusers to entrap their victims. And most of all, I will be interested in how it can be possible to wake up one day and see not just another person, but the events in your own life that involved that other person, in a new and deeply unsettling light—a light that reveals you to have deeply misperceived and misunderstood yourself and what was happening to you. Sometimes, as in Sylvia's case, such revelations are so long delayed that this delay is crucially implicated in the evolving entanglements that still make it hard for Sylvia to respond by exiting the relationship. But I suspect that you will identify this same unsettling kind of discovery in your own life, either on shorter time horizons or in less intimate relationships.

When Lenore Walker published her seminal study of abusive relationships, *The Battered Woman*, in 1980, it was in many ways a counterargument to the ease of blaming the victims of abusive relationships, especially after two decades of cultural changes had made divorce and financial independence easier than ever before for women, which made it easy for people who had never experienced such a relationship themselves to wonder, "Why do you stay?" Walker argued, based on her extensive experience working with victims of intimate partner violence, that in many ways abusive relationships don't look quite as dramatically different from non-abusive relationships from the inside as it might seem from the outside. In particular, the abuse is characteristically not constant, but comes in *cycles*—a good period followed by building tension, an abusive incident, and then reconciliation, starting the cycle all over again. And somewhat more

controversially, Walker also claimed that these cycles tend to become faster and more extreme over time, making their nature and direction less obvious at the beginning than you might think.[2]

Walker's twin focuses on the cyclical and accelerating nature of abuse helped many people, in the following decades, to sympathize a little bit better with the choices made by women in abusive relationships. But in the twenty-first century we are the inheritors of Maya Angelou's wisdom that "when someone tells you who they are, believe them the first time."[3] Angelou's advice is the advice that Sylvia wishes that she would have had and heeded, when she is lying awake late at night reliving things that John did early in their relationship. It is informed by Walker's insight that abuse comes in cycles and gets worse. And it tells us what to do in light of that. But it can also lead us back to victim blaming. For now when someone ends up entrapped in an abusive relationship we know that they have failed to follow Angelou's advice. Sylvia knows this about herself when she lies awake at night reflecting on the telltale events from early in her relationship with John.

Walker's work helped us get over one kind of victim blaming deriving from understanding too little about the dynamic of abusive relationships. My hope, in this paper, is to help us get over a different kind, that derives from understanding too much. My aim is to explain, in general terms, why Angelou's advice is not, after all, quite so easy to follow in prospect, and at the same time, by making rational sense of transformative discovery, why in retrospect it can seem like it *should have been* easy to follow. Along the way we will develop some tools that can be used to analyze many other forms of interpersonal conflict.

---

[2] Lenore Walker, *The Battered Woman* (New York: Harper & Row, 1979). Walker's methodology has been criticized at various times as being anecdotal, failing to include analysis of many intersectional traits that can make it especially difficult for some women to escape abusive relationships, overgeneralizing the intermittent and accelerating nature of abuse after the initial incident, collapsing distinctions between different kinds of sociological and psychological motives of the abuser, and omitting many important kinds of detail. These criticisms appear not all to be consistent with one another, however, and sometimes seem to reflect differing goals about what we might hope to understand about abusive relationships. But the patterns described by Walker remain commonly cited by victims as accurately, even if not completely, characterizing their experience as victims and commonly form the core of public-facing materials educating the public about the experience of abusive relationships, and it is that characteristic pattern, without prejudice to how fully it covers the range of abusive relationships or what it leaves out that is also worth saying, in which I will be interested.

[3] Angelou, in conversation with Oprah Winfrey in 1997. See, for example, Austin Kleon, "When people show you who they are, believe them," *Austin Kleon* (blog), October 3, 2018, https://austinkleon.com/2018/10/03/when-people-show-you-who-they-are-believe-them/.

I.B Interpretation and Conflict

When Sylvia awakens to the sudden realization that she has misunderstood the events in her own life and her own relationship with John, there is a shift in her interpretation. To understand why we are susceptible to such shifts, it is important to understand what it is that shifts about her interpretation, and hence to understand more about what goes into interpreting one another.

It is a familiar and well-trod observation that interpreting someone else requires constructing a kind of model of their mind—representing them as having beliefs and priorities of some kind. And philosophers working on moral responsibility have spent a great deal of time and effort shedding light on the connections between this kind of interpretation on the one hand, and emotions like anger, resentment, and hurt feelings, on the other. According to *quality of will* theories, such emotions are responses to the quality of will revealed by someone else's actions—or in other words, to what someone's actions reveal about their priorities or values.[4]

If someone's actions can sometimes reveal her priorities, however, at other times the evidence that they provide can be misleading. If you are walking down the sidewalk when you suddenly feel two hands planted firmly on your back and then find yourself sprawling on the sidewalk with your hot coffee spilled all over your clothes, you are likely to respond with anger, and to direct that anger at whomever you see standing above you with their arms outstretched. This makes sense, because this is excellent evidence that the person who pushed you has placed too low a priority on whether you skin your knee and stain your clothes with coffee. But this evidence can be misleading.[5] You might go on to learn that they were themselves pushed into you—or that there was an active shooter and they were trying to protect you by pushing you down. Prioritizing your life over coffee stains would not be an objectionable priority for them to have, and so it would not make sense to be upset about it.

So how we feel about the people around us and with whom we have relationships is a complex matter that is informed by how we interpret their mental lives. This makes sense, because we care deeply not only about what other people do, but about what they think and feel about us. These interpretations can be wrong, and because they can be wrong, they can shift—typically as a consequence of identifying that we *were*

---

[4] Compare Peter Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962): 187–211. For a survey of some different quality of will theories with comparisons, see David Shoemaker, "Qualities of Will," *Social Philosophy and Policy* 30, vols. 1–2 (2013): 95–120, https://doi.org/10.1017/s0265052513000058.

[5] I explore the significance of illusions of ill will in Mark Schroeder, *When Things Get Personal* (unpublished manuscript), Chapter 3.

wrong. When Sylvia's interpretation of John shifts, it involves many of these pieces—she comes to interpret him in a new way, and to identify her earlier interpretation as mistaken. As a result, she comes to feel new things about earlier events. And many of the new feelings that she has are among the reactive emotions that Strawson identifies as responses to the thoughts, attitudes, and feelings of other people. So it is right, I think, to look to what we know about interpersonal interpretation to find the basis for the transformative interpretive shift that leads Sylvia to see her relationship with John in a new light.

But I don't think that this is enough for us to fully understand what happens to Sylvia when she goes through this shift. When Sylvia wakes up one morning and comes to understand John and their relationship in a new way, she does not just have one new piece of information about John's feelings, motives, beliefs, or priorities. Rather, she comes to see almost everything that has happened in their relationship in a new light. For her it is like reaching the ending of *Fight Club* or *The Usual Suspects*. A single shift in interpretation engenders a holistic new perspective on what has happened throughout her relationship with John, and now she sees everything about it differently—not just his relative priorities on one or another occasion.

An even better example to compare to the flip in Sylvia's interpretation of John comes from a famous photo known as "the dress" that went viral in early 2015.[6] This photo became an international sensation overnight when people noticed that while it looked to many people like a photo of a gold and white dress, to many others the dress was obviously black and blue. When you see "the dress" as gold and white, it is difficult to imagine how it could look black and blue to anyone else. White doesn't look like blue, after all, and gold doesn't look like black. That is what made the dress photo an internet sensation. But even more strikingly, once you manage to flip the way that you see the dress from gold and white to black and blue, it is now bewildering how it could have looked gold and white to you before. After all, there it is, obviously blue and black—and blue looks nothing like white, and black looks nothing like gold.

When you realize that the person who pushed you to the ground thought that there was an active shooter, you are revising your understanding of their beliefs

[6] For the "the dress" photo and an explanation, see Wikipedia's entry on "the dress": Wikipedia contributors, "The dress," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=The_dress&oldid=1260536790 (accessed December 20, 2024). It turns out that I am not the first to draw a connection between the dress photo and domestic violence; in South Africa, the Salvation Army actually ran a domestic violence campaign using this photo with the slogan "why is it so hard to see black and blue?" *CBC News,* "Why is it so hard to see black and blue?" March 06, 2015, https://www.cbc.ca/news/trending/salvation-army-uses-the-dress-in-ad-targeting-violence-against-women-1.2985043.

and motives. Once you do, you see them in a new light—and will probably respond with gratitude or at least appreciation, rather than resentment. But you won't stay up late at night berating yourself for not seeing it right away, "the first time." It is perfectly clear to you how you could have made that mistake, and why your evidence supported anger. But Sylvia does stay up late at night, wondering how she could have gotten it wrong. She replays early moments in her relationship with John in her head, and it is now so dreadfully, agonizingly *obvious* what was going on. Her friends even told her at the time, and she ignored them. Her transformative discovery is much more like flipping from seeing the dress as gold and white to seeing it as black and blue.

I.C Signal and Noise

It could be, I suppose, that all that happens when Sylvia makes her discovery about John is that she gains some new piece of evidence about her husband's psychology that had eluded her before, as you gain a new piece of evidence when you see how nervous the person who pushed you is about an active shooter. I can't prove that is not the case. But if you think that it is so, then I want to draw your attention to just how common this mistake is, to the characteristic way in which these discoveries holistically transform our understanding of the past and not just the present, and to how obvious things can look, as they do to Sylvia after the discovery—the feature that makes Angelou's advice seem easier to follow, in retrospect, than it really is, in prospect.

These are the things that I find puzzling to understand, about transformative discoveries like Sylvia's, and about similar ones that I have made in my own life. These are the things of which I am going to try to offer an explanation, before the end of this paper. My explanation will take some work, so I won't have the space to also imagine alternative possible explanations and try to knock them down. The test for my explanation will be whether its benefits—and the other benefits that we can glean from its machinery—are worth its costs.

The dress illusion works because our eyes do not pick up directly on the surface reflectance properties of fabrics. Rather, our color vision works by triangulating on the average wavelength of light that impacts on each of three different kinds of cone receptor on our retinas. But in different lighting conditions, different wavelengths of light will reflect off of the same surface and make it to our retinas. So in order to perceive surfaces as having a constant color, our brains have to interpret the wavelengths of light that we actually receive against a background hypothesis about the lighting conditions. And it turns out that there are some lighting conditions in which a black and blue dress will reflect the same wavelengths of light to our eyes as a gold and white dress would in

other lighting conditions.[7] So interpreting from the wavelengths of the light whether the dress is gold and white or black and blue requires adopting a hypothesis about which lighting conditions you are in. It requires a global hypothesis about what is introducing noise into the channel of information that you are getting about the surface colors of the dress. When this global hypothesis shifts, you come to see the dress completely differently.

By the end of this paper, I am going to suggest that Sylvia's interpretation of John undergoes a global shift because it, too, trades on a global background hypothesis about what is introducing noise into the channel of information—about John. The reason why Sylvia's interpretation of John shifts in a radical way that transforms not just how she understands the present, but also the past, is that this hypothesis about noise is global. And the reason that Angelou's advice now looks like it should have been easy to follow, even though it wasn't, is that as with the dress, it is hard to shift between these background hypotheses about noise. In order to sustain this hypothesis, I am going to have to convince you that there is an important distinction between signal and noise that we use in interpreting one another.[8] But I suspect that you already recognize this, at least implicitly.

Suppose, for example, that we are discussing philosophy and you ask me a question about what I mean by the distinction between signal and noise and I respond in a harsh tone of voice, snapping back at you. To know how to respond, you have to identify whether my snap reveals that I am a rude asshole, or whether it perhaps instead (or also) reveals that you've been making a persistent mistake. If it's the former, then it makes sense to be upset at me, and if it's the latter, you might be embarrassed. But a third possibility is that I missed lunch and am simply hangry. The third interpretive possibility requires a different sort of response. Whereas the first two interpretations find meaning in my snap, but locate in it different kinds of meaning, on the third interpretation my snap is where meaning breaks down. The correct response is not to feel one way or the other, but perhaps to pass me a Snickers bar and ignore it.[9]

---

[7] The evening before this paper was accepted, I encountered a powerful illustration of this when my daughter got upset at me for buying her black tissue paper to distribute holiday gifts. The tissue paper that I had purchased was in fact gold, but she was seeing it in dim lighting conditions.

[8] The following ideas are further developed in *When Things Get Personal*.

[9] I don't mean to claim that it is always right to dismiss actions that are influenced by hanger—far from it. I just mean, by choice of an example familiar from a famous international ad campaign, to call your attention to the fact that you *do* distinguish between signal and noise in at least some cases, and to describe a familiar kind of case in which you might draw just this distinction. I also don't mean to imply that you cannot be upset at me for being hangry—for example, if you had reminded me that I would become so if I didn't eat.

If distinguishing between signal and noise in this way is an important part of interpersonal interpretation, then it is a new way in which interpersonal interpretation can go wrong. Just as we may fail to correctly identify someone's beliefs and priorities, so also we may fail to correctly identify what is signal and what is noise. And if we do so, then our interpretations will also appropriately shift, once we come to realize that we were mistaken.[10]

II. SOME PHILOSOPHY OF ACTION

So we do, I think, distinguish between signal and noise, in interpreting one another. And we do it all of the time. Philosophers of action have a name for this distinction, when it is applied to actions, in particular. It is the distinction between actions for which you are *attributively responsible* and those for which you are not—or for short, between which actions are, and are not, *attributable* to you.[11] When you interpret my snap as signal—either as revealing me to be an asshole, or as revealing me as responding to a mistake of yours—you are attributing it to me. When you interpret me as "merely hangry," in contrast, and pass me a Snickers bar to calm me down so that we can proceed, you are instead attributing my snap to my circumstances.

II.A Attributive Responsibility

Although the term "attributive responsibility" was introduced only later by Gary Watson, the classic pair of philosophical examples introducing the concept of attributive responsibility comes from Harry Frankfurt's classic paper, "Freedom of the Will and the Concept of a Person."[12] The willing and the unwilling addict are both, according to Frankfurt, bound by the strength of their addiction to end up taking the drug before the end of the day, no matter what they do. But whereas the willing addict awakes looking forward to getting her hit and planning her day around it, the unwilling addict wakes up believing that today is the day that she will finally get through without caving. She

[10] Throughout this paper I will refer to "distinguishing" between signal and noise, as if it is a binary distinction. I rarely present this work to an audience of philosophers, however, without someone wondering whether they can replace this binary distinction with a graded one. I invite you to do so, if you like, but the issues that I am going to discuss are complex enough that I believe nothing is added, and much is taken away from understanding, to constantly frame things in the more complex way that this would require. I also suspect that it could be graded in different ways, and that we won't know enough to think through these choice points unless we first have a better understanding of what work the distinction can do. So I will continue to speak throughout as if the distinction is a binary one.

[11] Gary Watson, "Two Faces of Responsibility," *Philosophical Topics* 24, no. 2 (1996): 227–48, https://doi.org/10.5840/philtopics199624222.

[12] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68, no. 1 (1971): 5–20, https://doi.org/10.2307/2024717.

deletes her dealer's info from her phone, throws her needles away and takes her trash out to the dumpster, and then spends the whole morning in an online support group for recovering addicts. But then when she takes a break for coffee, she runs into one of her old addict pals. Soon she is rummaging through the dumpster for her own trash bag, going through it to find an uncontaminated needle, putting her dealer's number back into her phone, and driving across town to get the goods.

Frankfurt says that the willing addict takes the drug *freely* but the unwilling addict does not. Watson describes this case differently. He says that the willing addict but not the unwilling addict is attributively responsible for taking the drug. I suggest that we will find the shift in Sylvia's interpretation of John in what she attributes to him and what she does not.

Despite its obvious applicability in examples like these, some philosophers are skeptical about the concept of attributability. This skepticism comes in two forms. One concerns some of the high-falutin' metaphors that surround it and I'll return to it in section III. But another form of skepticism about attributability is more prosaic. Maybe the distinction between attributable and non-attributable actions is a genuine distinction, but not a new one. Perhaps it is really just the distinction between full-blooded actions and behavior that doesn't rise to the level of full-blooded action.[13]

Indeed, it is natural to interpret some cases in this way, where non-attributable actions are impulsive, like my hangry snap, rather than reasoned. And if this were right, then it might be that the distinction between attributable and non-attributable actions adds nothing to the quality of will theory, because only full-blooded actions truly reflect our qualities of will.[14] But I don't think that this can be right. As I have described Frankfurt's unwilling addict, for example, her behavior *does* rise to the level of full-blooded action—indeed, it involves quite complex agency, requiring planning and coordination over time to search the trash, reconnect with her dealer, and drive across town.

So I conclude that the philosopher's concept of attributive responsibility gives us a way of identifying what we are doing when we distinguish between signal and noise.

---

[13] Compare, for examples, the account of attributive responsibility developed by Nomy Arpaly and Timothy Schoeder, *In Praise of Desire* (Oxford; New York: Oxford University Press, 2014), https://doi.org/10.1093/acprof:oso/9780199348169.001.0001. This thought is available, in a way, even in Donald Davidson's classic treatment of the distinction between actions and events, in Donald Davidson, "Actions, Reasons, and Causes," *Journal of Philosophy* 60, no. 23 (1963): 685–700, https://doi.org/10.2307/2023177.

[14] I have already said, in section I.B, why I am not going to try to explain Sylvia's transformative discovery solely in terms of the tools of the quality of will framework, but the remarks in this paragraph and the next are intended to acknowledge that you might think that the concept of attributive responsibility reduces to the tools of the quality of will framework after all and to address that directly.

Actions are noise when they are not attributable to their agent.[15] And attributability does not simply reduce to either full-blooded action or revelation of quality of will. It is something else.

II.B. Participant Responses

So if attributability is not the same thing as action, then what, exactly, is it? There is a well-established literature that tries to answer this question by saying what makes an action special in the way required to make it attributable to the person who performs it. And I am eventually going to have something to say about this literature, at least by implication. But first I want to answer a prior question, about what *kind of thing* we are saying about an action, when we say that it is attributable to someone. And I suggest that we can do so, by looking at the role that attributability judgments play when made by ordinary people as part of ordinary interpersonal interactions. Attributability, I suggest, is the "in" to what we can call, following Strawson, *participant relations*.

In "Freedom and Resentment," Strawson claims that we have two different ways of regarding or relating to people—two different perspectives that we can take when interpreting someone.[16] The first, which Strawson calls the "objective," is the same kind of way that we relate to any other kind of thing. This perspective is what we might call "clinical" or "detached," and it is characteristic of thinking about human behavior from a scientific perspective. But Strawson says that we also have a different way of relating to and regarding persons—a way that is distinctive to seeing them *as* persons. This is what he calls the "participant" perspective or stance.[17]

An important part of the evidence for the existence of a special "participant" way of regarding or relating to people marshalled by Strawson is that there is a whole class of ways that we have of relating to people that we can't relate to things that are not people. Paradigmatically for Strawson, while you can be angry *about* anything, you

---

[15] Later I will reverse this order of explanation, and say that attributability is a special case of the signal/noise distinction, applied to the case of actions. But attributability is the best-studied case, and so it will help us to focus on it and learn what we can from the philosophical literature on attributability.

[16] Strawson, "Freedom and Resentment."

[17] Strawson doesn't himself use the word "stance"—it was first introduced by Rae Langton, "Duty and Desolation," *Philosophy* 67, no. 262 (1992): 481–505, https://doi.org/10.1017/s0031819100040675, who distinguishes between the "objective stance" and the "interactive stance" to emphasize and echo, following Christine Korsgaard, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," *Philosophical Perspectives* 6 (1992): 305–32, https://doi.org/10.2307/2214250, Kant's distinction between the "standpoints" of theoretical and practical reason, and Richard Holton, "Deciding to Trust, Coming to Believe," *Australasian Journal of Philosophy* 72, no. 1 (1993): 63–76, https://doi.org/10.1080/00048409412345881 first introduced the now-ubiquitous phrase "participant stance" to refer to Strawson's idea.

can't be angry *at* a rock. You can only be angry at a person, or perhaps more carefully, only at something that you are regarding or seeing as a person at the time of your anger. Similarly, you can only forgive a person—or something that you are regarding at that time as a person. This is evidence for a special way of regarding or interpreting something as a person because we don't typically regard rocks in this way. But Strawson also points out that we can also occasionally step back from regarding persons in this way and think about them in the same sort of disinterested way that we normally relate to rocks. Again, this is evidence that we have two interpretive perspectives on persons—one that is distinctive to seeing or treating them as a person, and one that is not.

The participant stance makes possible, and the objective stance "excludes," in Strawson's words, the distinctively interpersonal responses that I have been calling, following Strawson, participant relations.[18] Of course, Strawson's only examples of participant relations are what he calls "attitudes"—mental states or mental acts—and in particular they are all what he calls "reactive" attitudes—attitudes *towards* or *responding to* the attitudes or perceived attitudes of other people. But I want to suggest that once we appreciate the distinction between participant and non-participant responses, we can see that this list is much too narrow. Just as you cannot be angry at a rock, you cannot complain *to* a rock, but only to someone who is, or at least who you see as, a person.[19] But complaining isn't an attitude—it's a speech act.

Similarly, as any woman who has ever offered an argument in a meeting only to see their point taken up later when reiterated by a male colleague knows, there is an important distinction between someone changing his mind in a way that is caused by what you said, and them being *persuaded* by you. The latter puts you in company with Descartes and Galileo, and the former puts you in company with Archimedes' bathtub and the apple that fell on Newton's head. Being persuaded by someone requires seeing them as a person and the argument as part of their personal contribution. And that is why it is demeaning when someone responds to what you said only in the way that Archimedes responds to his bathtub or Newton to his apple. So being persuaded by someone is a participant response. But being persuaded by someone is not a reactive attitude, either.

---

[18] On Strawson's claim that the objective stance tends to "exclude" participant relations, I find Langton's "Duty and Desolation" to be both especially helpful and especially forceful. I explore this in Mark Schroeder, "Persons as Things," *Oxford Studies in Normative Ethics* 9 (2019): 95–115, https://doi.org/10.1093/oso/9780198846253.003.0005.

[19] Thanks to Irene Bosco for discussion of complaining.

So there are a wide variety of participant relations—distinctive relations that we can only bear to something we are regarding as a person. These relations answer to "to whom" or "by whom" questions, rather than "to what" or "by what." As Strawson notes it is necessary to bear these relations to someone that you are regarding or seeing them as a person. But as Strawson also notes, it is not sufficient. Sometimes even while overall regarding or treating someone as a person, we nevertheless exclude some of their traits or some aspects of what they do from participant responses. Strawson seems to think that this amounts to a limitation or restriction of the participant stance—a way in which it is encroached on by the objective stance—and his characteristic examples of how it arises are all cases of moral *excuses*. So, for example, you may blame someone for arriving late for lunch at the same time that you excuse them for belching as they arrive, excluding their belch from eligibility for participant responses, but not exempting them from participant responses altogether.

Strawson makes it sound like when you do so, this amounts to a kind of limitation on the participant stance. But on the contrary, as I have argued elsewhere, this is really *part and parcel* of taking the participant stance toward someone.[20] Because we are all imperfectly embodied as persons, interpreting someone as a person always requires recognizing that their personhood has limits, and being able to recognize the boundaries of where their actions reflect them as persons (as in, their arriving late) and where they instead reflect merely their imperfect embodiment (as in their belching as they arrive). The distinction between attributable and non-attributable actions is just this distinction, for the case of action or behavior. It is the line between what is eligible for participant responses and what is not. It is a line that we draw from *within* the participant stance—a line that determines the boundaries of our participant responses. It is, for short, the "in" to participant responses.

This is why, when your colleague repeats your argument back to you in the meeting as if he has just thought of it himself, it is demeaning. It is not that he is not treating you as a person at all—he may even thank you for being the occasion of his thinking of this argument. But the fact that you are only the occasion of his changing his mind, and he is not actually persuaded by you, shows that even though he has identified your argument in what you said, and even though he has been convinced by this argument, he does not actually attribute making this argument to you. It is, in a way, there in what you have said, but he does not see you as responsible for offering it. And this is demeaning, because it makes you out to be a little bit less of a person, and a little bit more of a thing, than you really are.

[20] Schroeder, "Persons as Things."

So the attributable/non-attributable distinction, I have been arguing, is not just a philosopher's distinction. It is deeply embedded in our ordinary way of understanding one another, and required for us to relate to one another as persons at all. Employing it successfully keeps us on the same page with one another, allowing us to identify the same meaning in what each of us does, and to respond to each other as persons, rather than getting distracted by typos, belches, and the side-effects of hunger or hormones. But the fact that it is embedded in our ordinary way of relating to one another also means that if we mis-apply this distinction, then things can go wrong.[21]

Sylvia, I think, does misapply this distinction to John. When John tells her who he is the first time, she doesn't listen. Later, after her interpretation shifts, she has a very different participant understanding of what has happened in her relationship with John. She attributes things to him that she did not, before. But I think that we can still learn more about *why* Sylvia was apt to make the particular kind of mistake that she did, and about what happened when she fixed it. My suggestion is that if we better understand what *makes* an action attributable to someone, we will better understand what kinds of mistakes we should expect people to make about it.

III. SOME METAPHYSICS

In "Freedom of the Will and the Concept of a Person," Frankfurt argues that in order to identify the conditions of what I have been calling attributability, we must look to the philosophical concept of a person, and identify what it is to be a person, in the first place.[22] Most of the literature downstream from Frankfurt has either forgotten about this claim, or has set it aside as not important. But for Frankfurt it was important enough to feature in the title of his paper. And it sounds implausibly strong. Indeed, given what I have just been arguing, it sounds wildly strong. If, as I have been arguing, the distinction between attributable and non-attributable actions is something that we need to apply *whenever* we are having ordinary participant responses to someone, then Frankfurt's claim implies that the answer to philosophical accounts of the nature of persons are implicated in our ordinary ways of relating to one another.

[21] The project of working out in much greater detail the scope, diversity, and ramifications of what goes wrong when we misapply this distinction is the project of *When Things Get Personal*.

[22] Of course, Frankfurt doesn't use the word "attributable" or "attributive responsibility," which come from Watson, "Two Faces of Responsibility." But Frankfurt's theory is one of Watson's paradigms of a philosophical account of attributability, and so I will proceed to interpret Frankfurt as making claims about attributability.

III.A True Selves

Nevertheless, despite the patent logical strength of this claim, I think that there are excellent reasons to think that it is true. And I want to suggest that we can see this by returning to our twin observations that attributability is the "in" to participant responses and that participant responses are distinguished by the fact that they are distinctive ways of relating to persons, in contrast to other kinds of things—to *whos*, rather than to *whats*.

Start with the observation that attributability is the "in" to participant responses. As we have seen, this means that the boundary of what is eligible for participant responses is set by what we interpret as attributable to someone. We may be angry at them for stepping on our toe at first, but then, once we realize that they were shoved by someone else and so it was actually inadvertent, we cease to attribute stepping on our toe to them, and instead interpret it as a reflection of their circumstances. Once we see it in this way, anger no longer makes sense. Or in reverse, we start by seeing an argument in what someone says but think that it was us who identified this argument and were only inspired by their words to do so. But then we reflect on the fact that this is uncharitable and come to appreciate that it was, in fact, their argument. In coming to attribute it to them it now makes sense for us to be persuaded by them.

The fact that participant responses are responses to persons is enough, by itself, to make sense of why we only have them to someone who we are regarding as a person, and hence of why we cannot have them to rocks. But it is not enough, by itself, to make sense of why *attributability* is also required for participant responses. For the fact that you don't attribute someone's belch or hangry snap to them does not mean that you are not regarding them as a person—on the contrary, as I have argued, recognizing the things that are not attributable to someone is *part and parcel* of regarding them as a person. So that leaves, in a way, a puzzle about *why* attributability would be necessary for participant responses.

The answer to this puzzle, I believe, is that there must be some sense in which attributable actions belong more to someone *as a person* than non-attributable actions, so that responding to someone's attributable actions counts as a response to them *as a person*, while responding to someone's non-attributable actions is only responding to them *as a thing*. Only if attributable actions count as belonging more intimately to the person can we use the same fact about participant relations—that they are ways of relating to *persons*—to explain why it is both necessary to see someone as a person in order to have a participant response to some action of theirs and also necessary to interpret this action as attributable to them.

It is no wonder, then, that so often when philosophers discuss the concept of attributability, they reach for the metaphor that attributable actions are more "truly yours" or express in some way your "true self." These metaphors about the true self are attempts to articulate the idea that there must be a distinction between actions that belong directly enough to you as a person for responses to them to count as responses to you as a person, and actions that while belonging to you, belong to you in a way that is not sufficiently intimate to support this connection. It will be helpful, then, to try to see through this metaphor and make something of this distinction in relative degrees of intimacy with respect to which an action can belong to you "as a person."

III.B Inherence and Incidentality

A good place to turn, I suggest, in order to make sense of how an action could belong to you in a more or less intimate way, is toward the more general question of whether any properties can belong to any things in more or less intimate or direct ways. And here I think that the distinction between attributable and non- attributable actions is in excellent company.

Consider, for example, the statue and the clay. Goliath is a statue made out of a lump of clay, Lumpl.[23] Goliath is innovative, baroque, grey, and heavy. Lumpl is likewise innovative, baroque, grey, and heavy. But I contend that there is an important difference in how it is that Goliath is innovative and baroque and how Lumpl is—and likewise an important difference between how Lumpl is grey and heavy, and how Goliath is. Goliath, I hope you will agree with me, is innovative and baroque in its own right. But Lumpl is only innovative and baroque because it composes a statue that has those properties. Lumpl, in contrast, is grey and heavy in its own right, but Goliath is only grey and heavy because it is composed out of Lumpl, which is grey and heavy.[24] So in general, I want to suggest, things have some properties in their own right—*inherently*, as I will say—and some properties only in virtue of the fact that they are the properties of things with which they are co-located, or which they constitute or by which they are constituted—as I will put it, *incidentally*.

Much ink could be spilled over the inherent/incidental distinction. It could be, as I have characterized it, a distinction between two ways of having properties. On this way of interpreting the distinction, it is in some sense metaphysically deep. On another

[23] Allan Gibbard, "Contingent Identity," *Journal of Philosophical Logic* 4, no. 2 (1975): 187–222, https://doi.org/10.1007/bf00693273.

[24] Some of my informants suggest that, depending on other aesthetic properties of Goliath, Goliath may also count as heavy in its own right. I am happy to allow that this could be the case. What is important to me is only that there is a distinction, not that only one object counts as having each property inherently.

possible view, the distinction is shallow. Maybe if we are speaking most carefully and precisely, what we should really say, along with Kit Fine, is that only Goliath has the properties of being innovative or baroque, and that only Lumpl has the properties of being grey or heavy—and that calling Goliath "grey" (incidentally) and Lumpl "baroque" (incidentally) are just forms of loose talk, or perhaps uses of "grey" and "baroque" in an extended sense that really ascribes the indirect property of being co-located with something grey, or co-located with something baroque.[25] It could also be that we can make sense of the inherent/incidental distinction without endorsing the view that Goliath and Lumpl are distinct, and instead say that one and the same thing which is both a statue and a lump of clay is innovative-qua-statue and heavy-qua-lump.[26]

I am going to trust that the reader who wishes to adopt one of these alternative theories about what is going on in the intuitive difference between the way in which Goliath is innovative and baroque and the way in which Lumpl is, will be able to reconstruct everything that I say from here forward in their preferred idiom, or according to their preferred general metaphysical view and division of labor between metaphysics and the philosophy of language. What is important to me is not exactly how we theorize about this more general distinction, but rather *that there is* this more general distinction. And my claim is that we should see the distinction between attributable and non-attributable actions as just a special case of it.

On the view that I am suggesting, therefore, there is no great mystery in how the hangry snap could belong less to me as a person than the content of what I am saying. It is exactly the same as the respect in which the statue's weight belongs less to it as a statue than its aesthetic properties. Just as treating the statue as an aesthetic object requires responding to the properties that it has inherently—its innovativeness and its period characteristics, for example—and making arrangements for moving it from one place to another requires taking account of the traits of how it is constituted—its weight and overall dimensions, for example—similarly treating me as a person requires responding to the properties that I have inherently—to the actions that are attributable to me, for example.

The hypothesis that attributability of action is just inherence therefore explains not only what the metaphor of the "true self" is all about, but why it is that attributability is the "in" to participant responses. And it does so while only appealing to a very general

[25] Kit Fine, "A Counter-Example to Locke's Thesis," *The Monist* 83, no. 3 (2000): 357–61, https://doi.org/10.5840/monist200083315.

[26] Thanks to Paul Pietroski for this last, possibly less ontologically profligate, suggestion.

distinction that we need to be able to make sense of, in some way, for other kinds of things.

But most importantly, it reveals that attributability of actions is really only a special case of the signal/noise distinction that we use in interpreting and relating to one another. For the property of having performed some action is just one kind of property of persons, and hence it is just one place where we can look for an inherent/incidental distinction. Persons also have thoughts, emotions, attitudes, traits of character, and are embodied in different ways. For each of these properties we can ask whether it is always inherent to persons, always incidental, or whether in fact like the property of having performed some or another action, it might often be inherent but sometimes be incidental.

And here I think there is excellent evidence both intuitive and theoretical that we should apply the inherent/incidental distinction more widely. For example many people have thoughts that they experience as intrusive. These are no one else's thoughts—they are happening inside the head of the person who is experiencing them. So they have the property of having these thoughts. But if thoughts can be incidental as well as inherent, then the experience of these thoughts being intrusive can be veridical. So I suggest that the phenomenology of intrusiveness *just is* the experience of one of one's thoughts as being one's only incidentally. People also experience features of their own embodiment in different ways—sometimes as central to who they are, and sometimes as getting in the way. Again, I suggest, the *content* of this experience, whether or not this content is veridical in any given case, is that some of these features are one's inherently, and some only incidentally. And it is often alleged that you can have values—even deeply held values—that do not belong to you authentically because of the way in which you have been indoctrinated. Here too, I suggest, what we are identifying is the distinction between values that are yours inherently and those that are yours only incidentally.

Indeed much of the literature on attributability downstream from Frankfurt and Watson is implicitly committed, I believe, to the idea that the distinction between attributable and non-attributable action should turn out to be just a special case of some more general such distinction. For by far the vast majority of accounts of the actual conditions of attributability locate the source of attributable actions in a special kind of source in the agent's psychology—in their *second-order volitions*, or in their *values*, or in *traits with which they identify*, or in *attitudes that they accept*, or in their *character*.[27] All

[27] Compare Harry Frankfurt, *The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press, 1988), https://doi.org/10.1017/cbo9780511818172; Michael Bratman, *Structures of Agency: Essays* (Oxford; New York: Oxford University Press, 2007), https://doi.org/10.1093/acprof:oso/9780195187717.001.0001; Christine Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford;

of these accounts are attempts to single out some part of your psychology that is you inherently, and not just incidentally, with the idea that only actions that come from internal causes that are inherently you can be actions that are inherently yours. So this is further evidence, I think, that we are on the right track to identify attributability of action as just the inherent/incidental distinction applied to the case of persons and actions.

Allow me to spell out this hypothesis more explicitly. The distinction between inherence and incidentality can be applied to every kind of thing, and to every property of that thing. It can be applied to statues, and it can be applied to clay. It can be applied to shoelaces, to stars, and to populations of Wildebeests. So it can also be applied to persons. When we apply this distinction to persons, we get what I earlier called the distinction between signal and noise. Part of interpreting someone as a person, I argued, requires identifying what is signal and what is noise. I am now claiming that this is just a special case of applying the more general distinction between inherence and incidentality—but applying it to persons, in particular. It is no wonder, on this view, that relating to a person requires interpreting what is signal and what is noise—it is for the very same reason that relating to a work of art like the statue requires identifying what is inherent and what is incidental.

So the signal/noise distinction is just the inherent/incidental distinction applied to *persons*. Similarly, the attributable/non-attributable distinction is just the signal/noise distinction applied to *actions*. When we apply this distinction, we are identifying which actions belong to someone in the way that innovativeness belongs to Goliath and greyness belongs to Lumpl.

III.C The Concept of a Person

So let's return, then, to Frankfurt's strong—wildly strong, as I admitted— claim that our account of the conditions of attributability must extract those conditions from our concept of a person—from our answer to the philosophical question of what it is to be a person, in the first place. Since we have seen that the attributable/non-attributable distinction is just a special case of the inherent/incidental distinction, we can now use this understanding in order to see why Frankfurt is right.

New York: Oxford University Press, 2009), https://doi.org/10.1093/acprof:oso/9780199552795.001.0001; David Shoemaker, *Responsibility from the Margins* (Oxford: Oxford University Press, 2017), https://doi.org/10.1093/acprof:oso/9780198715672.001.0001; and August Gorman, "The Minimal Approval View of Attributability," *Oxford Studies in Agency and Responsibility* 6 (2019): 140–164, https://doi.org/10.1093/oso/9780198845539.003.0006, though Shoemaker and Gorman go to great lengths to be much more liberal about which source in the agent's psychology counts than the others.

The *reason why* different properties are Goliath's inherently and Lumpl's inherently surely, after all, has something to do with the kinds of thing that Goliath and Lumpl are—with the difference between what it is to be a statue, and what it is to be a lump of clay. It is *because* Goliath is a statue that it has properties like innovativeness and baroqueness inherently, and (typically) properties like greyness and heaviness incidentally. And it is *because* Lumpl is a lump of clay that it has properties like greyness and heaviness inherently and never properties like innovativeness or baroqueness except in a way that is merely incidental. Surely it is because we know something about what sort of thing statues are that we can discern which kinds of properties they have inherently, and which incidentally.[28] So likewise, by similar reasoning we can infer that it must be through understanding what kind of thing it is to be a *person*, that we can understand which properties are apt to be yours inherently, given that you are a person.

And we can see the fruits of this line of thought by taking seriously some of the different ways in which philosophers have theorized about what it is to be a person, each of which I think lends itself naturally to different expectations about what we should expect the conditions of attributability to look like, and therefore what sorts of mistakes we should expect reasonable interpreters to make, in interpreting what is attributable to one another.

Take, for example, the common thesis that persons are in some important sense self-made. This idea pervades much of philosophical thinking about persons and the self across different topics and applications over the last fifty years. It is visible in Frankfurt's idea that the distinction between attributable and non-attributable action lies in which mental states you endorse or identify with.[29] It is visible in Korsgaard's discussion of practical identities and self-constitution.[30] And it is visible in many forms of psychological continuity accounts of personal identity over time.

From the thesis that persons are self-made it follows that persons are not animals. For animals are not self-made—they exist in many cases, and certainly in the case of humans, long before they are capable of doing very much at all, let alone self-making. So proponents of the thesis that persons are self-made—in whatever form that thesis

---

[28] The right way to put this point depends, of course, on how we understand the distinction between inherence and incidentality. If you think that Goliath is the same thing as Lumpl and that it is just innovative-qua-statue and heavy-qua-clay, the point is that we have to know the difference between the sortals *being a statue* and *being a lump of clay*, in order to understand which properties Goliath-cum-Lumpl has qua-statue and which it has qua-clay.

[29] Not just in Frankfurt, "Freedom of the Will," but in various forms throughout the essays in Frankfurt, *The Importance of What We Care About*.

[30] Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), https://doi.org/10.1017/CBO9780511554476.

takes—should allow that you the person are co-located with and constituted by the animal in whose life you have constituted yourself. But you are not identical to that animal. You are related to it as the statue is related to the clay. And that opens up the possibility that not all of its actions are yours inherently. Your human animal has all kinds of thoughts and desires, and is subject to all kinds of influences—whether from being hungry, or tired, or affected by hormones or SSRIs. As a result of this it does all kinds of things. So you count as doing all of those things, because it is your body doing them. But not all of those thoughts, desires, or actions are yours inherently. Somehow your act of self-making constitutes some but not all of these thoughts, desires, and actions as inherently yours. And thus we get a distinction between what is attributable to you and what is not. Frankfurt, Bratman, and Korsgaard all give us slightly different answers about exactly how this self-making happens, but their answers are to this extent structurally the same.

The thesis that persons are self-made therefore offers an explanation of why it makes sense to distinguish between attributable and non-attributable actions. This is because this tracks a genuine distinction. And it offers an attractive picture on which each of us has great authority over ourselves and our own lives—a kind of authority that fits with liberal ideas about autonomy and respect. In all of its forms, of course, it faces problems with circularity, due to the fact that for any act of self-making we can intelligibly ask whether that act itself belongs to the person inherently or only incidentally. Making this kind of view work therefore requires there to be something that we do that can count as the relevant kind of making and is also secure, in virtue of its nature, against the charge of ever being possibly merely incidental.[31]

I doubt that there is any such thing that we can do.[32] But in this paper I'm interested in a more general problem with self-making views of persons. And that is that they don't just grant each of us *some* important authority over who and what we are. Really, they grant *too much* such authority.[33] If persons are self-made, then the right way to discern which actions are attributable to someone must be to discern their own acts of self-making. On this view, each of us is in a way, the ultimate authority on what is attributable to us. But this makes it unintelligible, I think, how often each of us interprets one another in ways that flout each other's self-interpretation. It may well be that we are wrong to flout each other's self-interpretations so often, and that we

[31] Compare Mark Schroeder, "Narrative and Personal Identity," *Aristotelian Society Supplementary Volume* 96, no. 1 (2022): 209–26, https://doi.org/10.1093/arisup/akac009.

[32] See Schroeder, *When Things Get Personal*, Chapters 7 and 8.

[33] Here I am painting with a very broad brush—there are of course many moves that a self-making view can make to try to explain why our own acts of self-making are sometimes obscure to us.

would do better to listen more carefully to one another and try to respect one another's self-interpretations.

I agree with all of that. But explaining why this is true by appealing to a general principle that each of us is the ultimate authority on our own self-interpretation is too powerful. It justifies us in criticizing people who are insufficiently attentive to how others interpret themselves only by making it unintelligible why anyone would make such a mistake in the first place. And more importantly for our purposes in this paper, like our original quality of will hypothesis, it makes the mistake that Sylvia makes earlier in her relationship with John out to be a kind of scientific mistake about which events transpired in John's head, and when. This leaves unclear why there should be a characteristic experience, like Sylvia's, of suddenly holistically reinterpreting someone and your relationship with them, and why Angelou's advice seems easier to follow in retrospect than it really was, in prospect.

Another approach to the nature of persons that has been recently gaining in popularity is the idea that persons are not strictly *self*-made, but rather are a kind of social construct. On this view, you don't just make yourself, but rather get help from those around you.[34] Like the idea of persons as self-made, the social construct view allows us to distinguish between person and animal. The person and the animal are co-located, and so the person shares the properties of the animal—including its actions—*incidentally*, but also has some properties *inherently*. And on this view, the question of which properties the person has inherently, including which of her actions are attributable to her, is ultimately to be answered by how she was socially constructed. And the view of persons as social constructs has an answer to my objection to persons as self- made, because it grants much less authority to the individual as author of herself. On this view, it is intelligible to not always respect your own self-interpretation, because your self-interpretation is not always definitive.

There are many things to like about this view—too many to detail here—and I like many of them. But like the view of persons as self-made, I think that it lends itself to a picture of the conditions of attributability that makes unintelligible many of the ways that we actually apply the attributable/not attributable distinction to persons. And that is because whereas the self-making view gives too much authority to the person herself, the social construct view lends too much authority to the group or society.

[34] Compare especially Susan Brison, *Aftermath: Violence and the Remaking of a Self* (Princeton: Princeton University Press, 2003), https://doi.org/10.1515/9781400841493; Hilde Lindemann, *Damaged Identities, Narrative Repair* (Ithaca, NY: Cornell University Press, 2001); and Hilde Lindemann, *Holding and Letting Go: The Social Practice of Personal Identities* (New York: Oxford University Press, 2016), https://doi.org/10.1093/acprof:oso/9780199754922.001.0001.

That makes this view ultimately too *conservative* about how to interpret someone. But sometimes we buck the trend and interpret ourselves or someone else in innovative ways. And sometimes, as I will argue later, our interpretations shift in drastic ways, as in Sylvia's transformative discovery about her own life and relationship with John. Rather than helping us to make sense of these drastic and disorienting shifts, the view of persons as socially constructed should lead us to find them unintelligible.

A more attractive view of persons, I suggest, should not only make plausible predictions about the conditions of attributability, but it should render intelligible the kinds of *mistakes* about attributability that people routinely make. Sylvia, as we saw, makes just such a characteristic mistake about John, that is only rectified late in their marriage. This isn't to say that it should *validate* all of the ordinary claims that people make—far from it. On the contrary, no matter what the right account of the conditions of attributability is, since it won't always be transparent to us we should *expect* ourselves to make characteristic kinds of errors in identifying it—just as we make characteristic kinds of errors in identifying when to invest in the stock market, characteristic kinds of errors in identifying whom to marry, and characteristic kinds of errors in perceptual experience. So it is worth turning, I suggest, to see whether there is any alternative way of thinking about what it is to be a person that makes better sense of the characteristic kinds of errors that we make in interpreting one another.

IV. The Interpretive Account of Persons

My own suggestion, to which I have been led back again and again the more that I try to think about the nature of persons and different applications of the concept of the person in different areas of philosophy, is that persons are what I call *interpretive objects*. Whereas self-making views say that you are who and what you are because you interpret yourself in a particular way, and social construction views say that you are who and what you are because of a collective act of interpretation, the interpretive view says that you are who and what you are because there is a *best interpretation* of you. This makes sense of some of the allure of self-making and social construction views, because given that who and what you are is set by your best interpretation, identifying you requires trying to interpret you. But it also explains why there is no ultimate authority on what this interpretation is, and does a much better job of making intelligible why we make the kinds of mistakes that we are prone to.

Let's take a look at what this idea entails in more detail, before returning to apply it to the question of what the conditions of attributability turn out to be, or the characteristic mistakes in attributability interpretation that we should thereby expect to find people making.

IV.A Interpretive Objects

It will help, in order to sharpen how I am thinking about persons, to zoom out briefly in order to examine the concept of an interpretive object more generally.[35] There are many interesting kinds of objects, each of which is individuated in a different kind of way. Natural objects like rocks, molecules, planets, and galaxies are bound together by the greater relative cohesion of their parts with one another than with other things. Artifacts are individuated by acts of creation. Functional objects are bound together by the parts constitutive of the performance of their particular function. Biological objects cohere around systems of self-maintenance. And social constructs are individuated by social practices. The category of interpretive objects, though it has been neglected by philosophers, is on a par with each of these other important and familiar categories of objects. But interpretive objects are bound together by their *interpretability*.

Because interpretive objects are often closely associated with other kinds of object, it helps, in order to distinguish them, to isolate cases in which there is no overlap. One way to collect such examples is by googling "rocks that look like animals." This search will bring up outcrops that look like elephants, hillsides that look like horses' heads, and more. A particularly simple case that I find it helpful to think about is a rocky outcrop above the 134 freeway in Los Angeles, between Glendale and Pasadena. The neighborhood that spreads out below this outcrop is called "Eagle Rock" for the obvious reason, once you get a good look at the rock, that it looks like there is an eagle with spread wings about to fly straight out of the rock. The eagle in this rock, for which the neighborhood is named, is an interpretive object. It's not a natural object, because it's smaller than and does not contain all of the rock. No one carved it, so it's not an artifact. It is clearly not a biological or functional object. And although we have a social practice of calling it an "eagle," it was there, so far as I know, before there were people to name it. So it's not a social construct, either. It exists because it looks enough like—and therefore is sufficiently interpretable as—an eagle. It's an interpretive eagle.

Each kind of object, as I have noted, is individuated in different ways. Natural objects like rocks are spread out as far in space as they need to be, in order to track the boundary of the relatively greater cohesion of their parts among one another than to other things. In contrast, functional objects like pairs of shoes are spread out in space in a way that includes all of the parts required for them to perform their function. This is why pairs of shoes can be spread out across two disconnected locations, but rocks

[35] The following remarks summarize the more comprehensive presentation of interpretive objects in Mark Schroeder, *Interpretive Objects: Meaning in Language, Life, and Law* (Princeton: Princeton University Press, forthcoming).

cannot. Interpretive objects are spread out in space just as far as they need to be, in order to be most interpretable as what they are interpretable as. So, for example, the interpretive eagle in the Eagle Rock is not as large as the whole rock, but only contains the part of the rock that is most eagle-shaped.

Similar points go for how different kinds of objects are extended in time, for their modal profiles, and (insofar as we distinguish this from their extensions in space and time), their parts. Natural objects last for as long as they cohere together. Functional objects last as long as they perform their function. And artifacts last as long as the changes that were implemented as part of their acts of creation. Similarly, an interpretive object like the eagle in the Eagle Rock lasts for just as long as makes it most interpretable as an eagle. Its parts are whatever makes it most interpretable as an eagle. And the best interpretation of it as an eagle also tells us under what counterfactual conditions it would have still been there.

The boundary between inherent and incidental properties, I suggest, is no different than boundaries in space, time, parthood, or possibility.[36] The question of which properties an object has inherently and which it has merely incidentally depends on what kind of object it is. Functional objects have inherently those properties that enable them to perform their function, and merely incidentally those properties that are merely incidental to their function. Artifacts have inherently those properties that are instilled or selected by their acts of creation and merely incidentally those that are otherwise. Biological objects have inherently those properties that play central roles in their systems of self-maintenance and merely incidentally those that are otherwise.

So by this reasoning interpretive objects, too, have their boundaries between inherent and incidental properties fixed by whatever it is that makes them most interpretable as the kind of thing as which they are interpretable. So, for example, the fact that the rock of the Eagle Rock is blotchy does not make it a blotchy eagle—it is just an eagle in rock that happens to be blotchy. And that, I claim, is because real eagles aren't blotchy, and so being blotchy does not make it more interpretable as an eagle. In contrast, if the rock looked like a cat, rather than an eagle, and it was striated, rather than blotchy, then rather than an interpretive eagle there could have been an interpretive tiger. In that case, the striations would be part of what made it interpretable as a tiger. So it would be a striped tiger in striped rock, and not just a cat in striped rock. In general, interpretive objects get to have whichever properties inherently make them more interpretable as what they are interpretable as, and their incidental properties are whichever properties get in the way of or distract from their being more fully interpretable as what they are.

---

[36] See Schroeder, *Interpretive Objects*, Chapter 1.

IV.B Interpretive Persons

In general, as we have seen, the boundary between inherent and incidental properties for interpretive objects comes from whatever makes them most interpretable as the kind of thing as which they are interpretable. An interpretive object gets to have whatever properties inherently make it more of what it is interpretable as, and its merely incidental properties are those that arise from the way in which it is imperfectly embodied. But earlier I argued that attributability is like this, too. Most of your actions are attributable to you. But some of your behavior, and some of your actions, reflect your imperfect embodiment—your nerves, your hunger, your hormones, your hangups. Rather than helping us to see you for who you are, these things get in the way. So they are in that way like the blotchiness of the Eagle Rock, which we need to ignore, in order to see the eagle, rather than like the striations in the Tiger Rock, which we have to pay attention to, in order to see the tiger.

But earlier I also argued that the attributable/non-attributable distinction is just the special case, for action, of the distinction between properties that are inherent to you as a person and those that are not. So this gives us what I believe is an excellent kind of evidence that you and I are interpretive objects—interpretive persons. Our inherent properties are those that bring us out as persons, and our merely incidental properties are those that get in the way of seeing us as persons, reflecting instead the way in which we, like the eagle in the Eagle Rock, are imperfectly embodied.

Because the eagle is an interpretive object, no one is an authority on which features of the eagle belong to it inherently and which merely incidentally. It is an interpretive matter, and different interpreters may reasonably disagree, within limits. Similarly, because you and I are interpretive objects, no one is an authority on which of your features are yours inherently and which only incidentally. It is an interpretive matter, and different interpreters may reasonably disagree, again within limits. Of course, each person is the one who gets to make the choices about how to live their own life, and the events of each person's life are the text of the interpretation of them as a person. So each person has a lot of power—indirect power, like that of an author writing the text of their own novel—to influence which interpretation of their life is correct.

Consequently it makes sense to have a lot of deference toward others' self-interpretation. But this is the authority of influence, not the authority of dictatorship. Just as sometimes you can clearly enough discern in a novel an interpretation better than the author's own, so sometimes you can clearly enough discern in a person an interpretation of their life better than their own. And the fact that this is *possible* makes intelligible why people may often make the mistake of failing to defer to another's self-interpretation even when in fact they should. So the interpretive account of persons

validates a large and important role for deference to others' self-interpretations without making unintelligible, as I have argued that self-making views do, the prevalence by which people fail to so defer.

The interpretive view of persons also makes sense of some of the strengths of social construction views while avoiding their conservativism. As I noted above, social construction views can make sense of how someone can become a person before they have the capacities to make themselves and can persist as a person after they have lost these capacities. The interpretive account has an easy time with these features, because what makes something interpretable as a person is a matter of the shape of their life over time, not a matter of its shape at each cross-section in time. Social construction views also, I think, make sense of why certain forms of life are not really possible unless social circumstances allow it—for example that while same- sex attraction and gender dysphoria may be universal features, identities like being *gay* or *trans* may require a possibility of social uptake. But if you like these features of social construction view, the interpretive view of persons can make sense of them, too, because sometimes social uptake is required in order for people to live in certain ways, and living in those ways is the interpretive text to understand them as persons.

So the fact that persons are socially embedded does set a constraint that is in some ways conservative on what kinds of life are possible. But it does it through constraining what kinds of life it is possible to live, not by directly constraining which interpretations of this life are admissible. And that means that there is much more room, on the interpretive view than on the social construction view, to interpret someone's life (including your own) in a way that departs sharply from the socially accepted interpretation. You can be right to interpret someone (including yourself) in a way that is deeply unconservative, and to carve a new and different path for yourself.

Finally, the interpretive account of persons explains why it is right to draw an attributable/non- attributable distinction at all. As an interpretive object, you are physically constituted by a human animal and its life. But you are not identical to it, any more than the eagle is identical to the rock. You may last longer or shorter in time, and you may be extended farther or less far in space or in parthood, including for example your prosthetic limb or cochlear implant though your human organism does not. And so likewise some of your actions and behaviors make you more interpretable as a person – and those are therefore attributable to you—while others get in the way of seeing you as a person, and are noise, rather than signal. Those are not attributable to you, and this is why they are not.

This is why Sylvia's interpretive shift is like the visual shift from seeing the dress as gold and white to seeing it as black and blue. Seeing John for who he is, like seeing

the Eagle in the rock, requires adopting a global hypothesis about which features in his life reveal him for who he is, and which get in the way of seeing him for who he is. It requires identifying how noise gets in the way of signal.

IV.C Agency and the Good

We require just one more step, in order to have the tools to see why Sylvia makes her characteristic mistake, early in her relationship with John, and why this is the kind of mistake that tends to become corrected over the course of an extended relationship. The answer to this, I want to suggest, comes from the difference between eagles and people.[37]

What makes the eagle interpretable is that it looks sufficiently like an eagle. It isn't really an eagle, of course, because real eagles have feathers and fly and lay eggs and so forth, and it does none of those things. But it is as close as it comes to that, within the confines of the rock. Similarly, I suggest, what makes you interpretable as a person is that you are sufficiently like a person—that you are sufficiently *person-y*. But unlike the eagle, there are no real, genuine, non-interpretive persons out there. Interpretive persons like us are as good as it gets. And here what I have to offer is more of a sketch, than an argument. The true argument that we are interpretive persons goes backwards—from how well this makes sense, as I will show in section V, of the distinctive patterns in the kinds of mistakes that we make in interpreting one another.

The problem of free will, in its most general form, is the problem of making sense of how our experience of ourselves and others as the ultimate reasoned sources of our own decisions fit into the world of causes as revealed to us by science. Of course, there is a respectable tradition of denying that it can be so fit—and concluding that we are not free. And in contemporary philosophy there is a substantially larger tradition of locating freedom and reasons in special, distinctive, locations within the space of causes. The suggestion that I'm about to make follows a third way.

Our experience of ourselves as persons, I suggest, *is* an experience of a kind of freedom and agency that does not perfectly exist anywhere in the world. The best efforts of compatibilist metaphysicians do not succeed in showing us where it is. But the fact that compatibilist metaphysicians can get as far as they do, and that they can say such plausible things, are symptoms of the fact that we sufficiently *resemble* this kind of absolute agency, that we can get along sufficiently well *interpreting* ourselves—and

[37] For further development, see Schroeder, *Interpretive Objects*, Chapters 4–5, and Schroeder, *When Things Get Personal*, Chapter 12.

others—as if we have it. Just as we can get along for limited purposes in seeing the rock as containing an eagle—but in this case for a bit more general purposes.

My suggestion is that our experience of ourselves is *as protagonists*. Protagonists, most importantly, *do* things. They are the ultimate sources of agency. True, they cannot do just anything that they desire—they are shaped in complex ways by their circumstances—by the predicaments in which they find themselves. But they make their own choices about how to respond to those circumstances. Whenever we make a choice, we have to draw this distinction between ourselves and our circumstances—the distinction, in decision theory, between the rows of our decision table, on the one hand, and its columns, values, and associated probabilities. As Kant observed in section 3 of the *Groundwork*, making a choice requires seeing yourself as the one with the freedom to go with either option.

Protagonists also have a second important feature, however. They do, at least to some important extent, *good* things. Kant's version of this thought crystallizes it into the idea that perfect noumenal agents never do wrong. The thesis of the guise of the good offers one important attempt at articulating what this looks like from the inside, when you see yourself as a protagonist. And the principle of charity is its manifestation in our interpretation of others. Cooperating with another person requires interpreting them charitably—at least within the bounds of the limits of your cooperative endeavor.[38] It requires seeing them not just as loci of agency, but as doing things that contribute positively toward your shared goals.

We are not, I think, perfect protagonists. We are too imperfectly embodied for that. We are neither perfect agents nor perfectly good. But we are enough *like* perfect protagonists, for us to see one another in this way. And that is what is involved in seeing ourselves and one another as persons. To interpret someone as a person is to strip away all of the features of their imperfect embodiment that get in the way of seeing them as a perfect protagonist. It requires overlooking typos, hangry snaps, and addictive urges.

The framework of interpretive objects allows us to add to this that to *be* a person is to be the object constituted by the best interpretation of you *as* a protagonist. The properties that you have inherently are those that help to constitute you as the most perfect protagonist that can be discerned in your life, and your incidental properties are those that get in the way of seeing you for this protagonist.

The concept of a protagonist, I suggested, includes the concept of the good. Seeing someone as a protagonist requires, other things being equal, seeing them in a positive

[38] See especially Korsgaard, "Creating the Kingdom of Ends" and Langton, "Duty and Desolation" for attractive developments of this idea.

light. So charity, on this view, is a fundamental feature of interpersonal relations. And even though charity involves a bias toward the good, the hypothesis that we are interpretive objects explains why this does not make it a bias away from the truth. But the concept of a protagonist also includes the concept of agency. Seeing someone as a protagonist requires seeing them as a doer.

Some of the characteristic mistakes that we make about attributability and the signal/noise distinction are, I think, consequences of the fact that this distinction, like everything about us, is ultimately interpretive. Others are consequences of the fact that attributability interpretation requires charity. But Sylvia's experience of suddenly comprehensively re-interpreting the whole of her relationship with John in a way that so destabilizes her own understanding of how she could ever have let the wool be pulled over her eyes is best explained by all of these pieces, put together.

V. BACK TO BUSINESS

We set out at the beginning with the suspicion that because interpersonal conflict is conflict between *persons*, and because one important way that conflicts accelerate is by becoming more *personal*, philosophical insight into what it means to be or be treated as a person might help us to understand one or more aspects of the dynamics of interpersonal conflict more generally.

Along the way, we have taken a detour through exploring one important dimension of interpersonal interpretation that lies at the center of how we relate to one another because it forms the "in" for participant responses—the matter of whether an action is or is not attributable to someone. I've argued that attributability is just the special case, for actions, of a much more general distinction between properties that belong inherently to a person and those that are hers only incidentally, that different kinds of objects have different kinds of properties inherently, and that it makes independent sense to think that persons are interpretive objects because none of us are—and indeed, many of the central puzzles about how our experiences of ourselves fit into the world revealed to us by science arise because none of us are—perfectly person-y.

And so we now know just enough, I believe, to be able to return to the general question of how this investigation might help us to think through some more general issues about interpersonal conflict—and to shed light on the phenomenon of transformative discovery in particular, and on its important role in the shape of some kinds of relationships, including, possibly, Sylvia's.

V.A The Consequences of Error

The most general consequence of the foregoing arguments is that our interpersonal relationships depend greatly for their shape on how we interpret one another, and

that there is great room for this to go wrong, because none of us have any direct or perfect access to the ultimate facts about what is attributable to whom. That means that we will often make mistakes in interpreting one another—and indeed that we may often be subject to *illusions* of attributability or its lack that are due to the evidence that is available to us. As a result, we will often respond to one another in ways that are mistaken—and we may fail systematically to recognize that this is what we are doing.

One way of making a mistake about the inherent/incidental distinction is to overproject signal. If you do this, then you are interpreting some aspect of what someone is doing as meaningful to understanding and relating to them as a person when in fact it is just a reflection of their imperfect embodiment. You may, for example, become angry with me for snapping at you, or preoccupied with trying to figure out what mistake you were making when you asked your question, when in fact my snap is merely a reflection of the fact that I missed lunch and there is nothing else to it. Or you may spend hours trying to decipher the distinction that some philosopher is making between 'wants' and 'desires' when in fact they are using them as synonyms without really noticing that this is what they are doing. Overprojecting signal can distract us with things that are not there.

It is a quite different kind of mistake, however, if instead of overprojecting signal, you overproject noise.[39] We are all at risk for overprojecting noise, because given that none of us has perfect access to the inherent/incidental distinction, the only way to avoid at least *some* overprojecting of signal is to accept the risk of sometimes overprojecting noise. This is what the mansplainer in your meeting does, when he repeats your argument back to you as if he has just thought of it (prompted by the stimulation of speaking with you, of course). When we overproject noise, instead of getting distracted by things that aren't there, we miss out on what is right in front of us. The mansplainer does it to you, and as I will go on to argue shortly, for a long time Sylvia did it to John.[40]

In this paper I promised an account of transformative self-discovery and its role in some kinds of abusive relationships. So to that we must turn at last.

---

[39] Overprojecting noise onto a communicative act results in a particularly important and distinctive kind of *silencing*. Compare Mark Schroeder, "Attributive Silencing," *Oxford Studies in Normative Ethics* 12 (2022): 170–92, https://doi.org/10.1093/oso/9780192868886.003.0009, following Mary Kate McGowan, "On Multiple Types of Silencing," in *Beyond Speech: Pornography and Analytic Feminist Philosophy*, ed. Mari Mikkola (New York: Oxford University Press, 2017), 39–58, https://doi.org/10.1093/acprof:oso/9780190257910.003.0003.

[40] The role of charity in interpersonal interpretation can help to explain both of these errors. Sylvia's mistake arises from her being charitable to John. And since charity involves value judgment, widespread distortions in values, such as those over gender roles, are going to lead to people making systematic mistakes (in this case, in ways connected to gender).

V.B Transformative Discovery

Everything that I have said about error and its consequences survives, even if my thesis that what makes you interpretable as a person is some combination of agency and value—indeed, even if I am completely wrong in my speculation that you and I are interpretive objects at all. Indeed, those remarks survive even my hypothesis that attributability is just a special case of the inherent/incidental distinction. Although the conditions under which we should *expect* to find people in error about how to interpret one another depend on what we say about each of these things, the facts that error and discord are ultimately inevitable and that they are going to have predictable consequences for interpersonal relationships are independent of all of those further claims.

But each of these further theses that I have advanced does have a payoff, in thinking through why it should ever make sense to sometimes experience radical shifts in how we understand what has gone on in our relationship with someone else. And that is because given the twin values of agency and value in interpreting someone else as a person, and given the fact that these interpretive values are often in tension, it can relatively easily turn out that sometimes, large differences in the value that we attribute to someone can be compensated for by large differences in the degree of agency that we accord to their life. As a result, sometimes interpreting someone can require choosing between radically different conceptions of what is going on—on one of which they are a relatively decent person struggling with a highly encumbering predicament, but on the other of which they are a horrible person after all or at least have done some quite horrible things, through what in fact turns out to be quite powerful agency much less encumbered by their predicament.

To see what this would be like, imagine that after asking me your original question about the signal/noise distinction and getting my snapping response, you work out that I am merely hangry and so you give me a Snickers bar and we proceed to be able to have quite an interesting conversation about philosophy and interpersonal conflict. Then next week, when we are discussing philosophy again and I snap at you once more, you will have a tried-and-true interpretive strategy for making sense of me, and some evidence that it helps you to better identify what is going on, because your interpretive strategy worked for us to move past the snap and be able to have a productive conversation. So once more you pass me the Snickers bar and once more we move on. The more often that this happens, the more practiced you will become in seeing my hangry snapping as part of my imperfect embodiment—part of what you have to work around, in order to have a relationship with me.

But then one day, you start to realize just how much of our relationship revolves around me snapping at you and you giving me a Snickers bar. It dawns on you that you find yourself stopping by convenience stores late at night in order to keep stocked up on Snickers bars in case you run into me over coffee in the morning. And it occurs to you to start to think about just what portion of my caloric intake you are providing in the form of Snickers bars. Suddenly you see me not just as someone whose agency is limited by his embodiment, but as someone who has been using you as a Snickers bar pump. Although we have had many collegial conversations about many specific things—some of which were super helpful at the time—all of those now start to pale, in a way, next to the significance of how my hangry snapping has structured your needs to work around me by keeping Snickers bars stocked at all times.

This is transformative discovery, as experienced by Sylvia—and, I conjecture, as experienced by you in some form or another at some point in your life. It is made possible by the competing interpretive values of agency and value, and it often—especially in the most disorienting cases—goes in the negative direction rather than in the uplifting direction by which Elizabeth re-interprets Mr. Darcy precisely because so much of charitable interpretation requires overlooking small but negative things, because initially small but negative things can so easily accumulate into large things, and because our strategies for overlooking the small but negative things can so easily be stretched into sacrificing more and more agency in pursuit of a more positive interpretation. This is especially true when holding onto our interpretation of someone else is so closely tied to our own self-image—for example, as someone who would surely not let herself become trapped in the same kind of relationship as her mother did.

Maya Angelou tells us, "when someone tells you who they really are, believe them the first time." And all of us who like Sylvia have gone through a negative-shifting transformative discovery wish that we had better understood and followed this advice at the time. But a large part of what I have been endeavoring to show in this paper is that, like the advice to "buy low, sell high," its patent truth covers over the fact that it is easier to understand than to apply. Because not everything that someone does is inherently them, not everything that they do *is* telling you who they really are. And some forms of abuse, including many forms of emotional abuse, are much harder to detect than others. Whereas a hit is a hit whether it is attributable or not—and, I would argue, not worth it under either interpretation—whether something is emotional abuse or not is a much more holistic interpretive matter that depends on how it fits into a broader pattern, and hence often particularly difficult to see. Things look much less different from the inside than it might seem from the outside that they must.

The fact that Angelou's advice is harder to follow than it wears on its sleeve follows, I think, from the importance of the distinction between attributable and non-attributable actions in the first place, and from the fact that we can make mistakes about how to apply this distinction. But the fact that all interpersonal interpretation involves the exercise of interpretive charity in trading off agency against value also explains part of the distinctive power of abusers' narratives to entrap. At Walker's reconciliation stage, the abuser apologizes, and may plead the special circumstances of, for example, how he struggles with anger, or how he was hit by his own father, or his difficulty with certain emotional responses.

To endorse the thesis, as I have, that good interpersonal interpretation always and rightly involves charity is not to validate the abuser's narrative at the reconciliation stage, but rather to explain its power. This narrative has the power to ensnare not because it tricks people into trusting and interpreting one another in unnatural ways, but rather because it exploits and takes beyond its proper extension the *right* way of relating to and interpreting one another. The fact that good systems can be exploited by bad actors doesn't make them bad systems—it just makes them bad actors.

And finally, I would like to conjecture that the existence of transformative discovery may also help to explain why Angelou's advice can *appear in retrospect* so much easier to follow than it really was at the time. When Sylvia lies awake at night rehearsing the things that John did early in their relationship, it seems painfully, embarrassingly, obvious what mistake she was making at the time. But this, I think, is like the way in which, once we have shifted from seeing the dress as gold and white to seeing it as black and blue, it is so very difficult to appreciate what was compelling about seeing it as gold and white in the first place. In both cases, the gestalt shift is mediated by a holistic background. In the dress photo this is an implicit understanding of the lighting conditions of the photo, whereas in the transformative discovery case it is an understanding of the holistic way in which John's behavior combines to constitute his agency over time.

In both cases, our new perceptions of what is signal and what is noise get in the way of seeing it as we used to. And that can make it as hard to see in retrospect how we could have made such a glaring error at the time—an important part of why transformative discoveries like Sylvia's can be so destabilizing and threaten our confidence in our own abilities to understand what is going on around us and to act with power in the world. The bad news is that being a person is messy, and hence understanding someone for who they are, even more so. But the good news is that this is in part because you are the most person-y that you can be. You are as powerful an agent as the interpretation of your life will allow.

Acknowledgements

Thanks to hundreds of people along the way for listening to me talk about persons, interpersonal interpretation, and conflict. But especial thanks to "Sylvia" who in fact underwent such a transformative discovery and to those people in my life about whom I have made such discoveries, to Kathryn Pogin for a forceful question that long ago interested me in how philosophy might help us to understand the power and scope of injustice, to Erica Shumener for first pressing me about abusers' reconciliation narratives, to Shieva Kleinschmidt for insight that I have failed to incorporate into this paper about the difference between agent- and patient- centric conceptions of abuse, to Hille Paakunainen for helping me to appreciate the parallels between overprojecting signal and overprojecting noise, and to a retired legal advocate for victims of abuse in the audience for the first version of this paper at Leeds in May 2023 for sustaining my hope that something in these ideas might be true to life. This work has benefited from audiences of related talks in many places, but especially under this title at Leeds, Edinburgh, and Groeningen, and finally in close to its present form at Rutgers University.

Competing Interests

The author has no competing interests to declare.