Synthese (2023) 202:151
https://doi.org/10.1007/s11229-023-04311-2

ORIGINAL RESEARCH

Hot-cold empathy gaps and the grounds of authenticity

Grace Helton[1] · **Christopher Register**[1]

Received: 31 December 2022 / Accepted: 18 August 2023 / Published online: 1 November 2023

Abstract

Hot-cold empathy gaps are a pervasive phenomena wherein one's predictions about others skew 'in the direction' of one's own current visceral states. For instance, when one predicts how hungry someone else is, one's prediction will tend to reflect one's own current hunger state. These gaps also obtain intrapersonally, when one attempts to predict what one oneself would do at a different time. In this paper, we do three things: First, we draw on empirical evidence to argue that so-called *hot-cold empathy gaps* arise when one projects one's own current state into a simulation about another. Second, we argue that this process does not typically confer knowledge, even when the predictions it produces happen to be accurate. Third, we suggest that these results can be used to develop a challenge for L.A. Paul's view that authentic action sometimes requires a certain kind of experience-based knowledge of one's own values and of how these values relate to relevant outcomes. We then sketch an alternative view of the epistemic grounds of authenticity, one on which authenticity requires a kind of understanding. The relevant form of understanding can be achieved by knowledge from first-personal experience but can also be achieved elsewise, such as through testimony from a close friend about what one values.

Keywords Empathy · Hot-cold empathy gaps · Authenticity · Transformative experience · Subjective knowledge · Self-knowledge · Rationality · Understanding · Cognitive science · Epistemology · Seductions of clarity

Both authors contributed equally.

✉ Grace Helton
ghelton@princeton.edu

Christopher Register
chris.register@princeton.edu

[1] Philosophy Department, Princeton University, 08544 Princeton, NJ, USA

It is a platitude that you shouldn't go grocery shopping while hungry. The thought is that if you do, you risk leaving the store with a cart brimming with more food than you could possibly eat, much of it of little nutritional value. As it turns out, empirical evidence vindicates the advice encapsulated by this platitude. It suggests moreover that the wisdom of that advice is likely explained, at least in part, by a broader phenomenon involving how subjects in *hot states*, i.e., affective states, make predictions about their behavior in relevant *cold states*, i.e., non-affective states. In particular, subjects in relevant hot states tend to predict that, even were they *not* in that state, they would behave in a way that is congruent with being in that state. For instance: people who are currently hungry overestimate how much they would want to have spaghetti for breakfast.[1] Notice that this prediction is skewed 'in the direction of' the subject's current state of hunger.

This effect also occurs in the opposite direction.[2] Subjects in cold states are poor at predicting how they would act if in some complementary hot state. For instance: when considering the distant possibility of performing an embarrassing dance, people tend to underestimate the effects of social anxiety and overestimate their willingness to perform.[3] Notice again that the misprediction is skewed 'in the direction of' their current state.

The examples so far involve *intrapersonal* predictions about oneself at other times, but one's current hot or cold state also tends to affect predictions about how *others* would act, i.e., *interpersonally*. For instance: people who are thirsty tend to overestimate how thirsty other people are,[4] and doctors who are not in pain underestimate how much pain their patients are in.[5]

Psychologists refer to this broad class of phenomena, wherein one's current hot (or cold) states systematically skew one's behavioral predictions about both oneself and others 'in the direction' of those states, *hot-cold empathy gaps*. This label refers both to the intrapersonal and interpersonal variants of the phenomenon. It also picks out both instances in which the subject making the prediction is in a hot state and instances in which she is in a cold state.[6] The effect has been studied most in the context of *visceral states*, which are roughly states of the body which serve immediate needs of survival, such as hunger, thirst, fatigue, warmth, sexual arousal, and pain. However, other results show similar effects for other felt states, such as curiosity, social anxiety, attachment to possessions, and social pain.[7] The label 'empathy gaps' reflects the common presumption that the gaps result from a failure to first-personally

[1] Gilbert et al. (2002).

[2] Fisher and Rangel (2014).

[3] Van Boven et al., (2012).

[4] Van Boven and Loewenstein (2003).

[5] Loewenstein (2005a, b).

[6] In some contexts, 'hot-cold empathy gaps' refers exclusively to versions of the phenomenon wherein subjects in a hot state mispredict the behavior of subjects not in that state. We aim to make clear by context which of these usages we are employing. See also Van Boven et al. (2013).

[7] Respectively shown in: Loewenstein et al. (1998), Van Boven et al., (2012), Van Boven et al. (2000), Nordgren et al. (2011).

model the target subject's psychological state.[8] On this presumption, the reason that (say) a hungry person fails to correctly predict how much food a full person would eat is that she fails to imagine 'from the inside' what it is like to feel full, rather than hungry. The predictive error is rooted, on this view, in a failure to accurately or fully render the subjective experience of some 'other', whether that other is oneself in a different scenario or another person altogether.

In this paper, we have three broad aims. The first is to develop a partial, empirically plausible account of hot-cold gaps. Specifically, we will argue that this process involves a subjective or first-personal encoding of one's current visceral states (§1.1) and that hot-cold gaps typically cannot be 'closed' by way of correcting one's simulation (§1.2). Our second aim is to argue that the predictions generated by this process very often do not amount to knowledge, even when they happen to be accurate (§2).

Our third aim is to build on the previous results to draw out some implications for a compelling and influential view of authentic decisions developed and championed by L.A. Paul, on which authentic decisions must, in certain cases, be guided by knowledge of one's values based on first-person experience, or *subjective knowledge* (§3). We will argue that when it comes to deciding what to do on the basis of one's values, hot-cold empathy gaps often block subjective knowledge. As a result, the requirement that authentic decisions about what to do be guided by subjective knowledge of one's values is triggered in far fewer cases than one might have thought. In light of this result, we briefly motivate an alternative view of the epistemic grounds of authentic action, one on which authentic action should be based on a kind of *understanding*, where understanding can be achieved through subjective knowledge but can also be achieved elsewise, such as through testimony from those who know you and what your values are.

Before proceeding to our main claims, it will be helpful to say something about the broader significance of hot-cold gaps. First, it might seem, from the cases we've glossed so far, that hot-cold empathy gaps are *generally a bad thing*, in the sense that they tend to stymie practical ends and, in some cases, moral ends. For instance, because of hot-cold gaps, the recovering alcoholic who is in the throes of a craving will underestimate how much she will regret her drinking later, when she is no longer craving alcohol, and this predictive difference might plausibly contribute to her choosing to drink later, with consequences she might sharply regret.[9] On the moral front, hot-cold gaps may play a role in the unjust stigmatization of impulsive behaviors, such as overeating and addiction,[10] in the undertreatment of pain by doctors who

[8] See, e.g., Van Boven et al. (2013) for a view on which empathy gaps are due to two judgments.

[9] This suggestion is indirectly supported by Poggiolini (2019). Our suggestion is that hot-cold gaps might play some role in addiction; we are not claiming that these gaps are the sole or primary drivers of addiction. For an interesting take on the contextual factors in addiction, see Pickard and Pearce (2013). For evidence that hot-cold gaps plausibly play a role in sub-optimal negotiation processes, see, e.g., Loewenstein and Adler (1995) and Van Boven et al. (2000).

[10] Nordgren et al. (2007).

are not currently in pain,[11] and perhaps even in the public's acceptance of policies that permit torture.[12]

We resist the sweeping conclusion that empathy gaps are generally a bad thing, even as we acknowledge that empathy gaps sometimes stymie practical and moral ends. In some cases, empathy gaps help *facilitate* practical ends and even, potentially, moral ones. For example, someone's excitement in the moment might help them start a worthwhile future project they would avoid were they to fully appreciate how bored it would make them by the end. Likewise, the addict who is not currently craving might be more willing to check themselves into rehab than they would be if they were to fully appreciate how powerful their craving will become after that decision is made. The decision to enter rehab is plausibly a good decision, both practically and morally. Thus, while empathy gaps sometimes stymie practical or moral ends, in other cases, they seem to facilitate them.[13]

A second point about the significance of hot-cold empathy gaps concerns the connection between these gaps and empathy deficits more generally. On the psychological view of hot-cold gaps we will defend, hot-cold gaps emerge when one 'projects' one's own hot or cold state into a first-personal rendering of the experience of someone who is in a different state. While this kind of projection *might* undergird empathy deficits in general, it is no part of our claim that they necessarily do. Rather, we are more inclined to see the process of projection at play in hot-cold gaps as merely one in a grab bag of methods by which we 'read minds,' that is, attribute mental states to others. Other psychological strategies of mindreading plausibly include: the use of background theories, perceptual processes, emotional resonances, and heuristics.[14] In defending the view that hot-cold gaps arise as a result of first-personally projecting one's own state into a model, we are neutral on whether this kind of projection also figures in (for instance) more complex forms of empathizing, such as understanding a loved one's puzzling behavior in the wake of their having suffered a traumatic event or appreciating the ideological motivations of someone opposite you on the political spectrum.

[11] Loewenstein (2005b).

[12] Nordgren et al. (2011).

[13] For further discussion of the ways that empathy in general can be morally problematic, see for instance: Kate Manne's discussion of excessive empathy for men, what she terms "himpathy," Sukaina Hirji's discussion of the ways in which empathy for an abuser can make it difficult for the person abused to maintain a proper sense of herself, and Olivia Bailey's discussion of the ways in which empathy can warp testimonial trust (Manne, 2017; Hirji, 2022; Bailey, 2018). For an argument that there is a puzzle about empathizing with vicious perspectives, see Bailey (2021). For an argument that empathy in general can compromise the empathizer's authenticity, despite its moral and epistemic benefits, see Paul (2021).

[14] For helpful recent discussion of some of these myriad strategies in mindreading, see Spaulding (2018, 2020).

1 Hot-cold empathy gaps and the first-personal perspective

In this section, we will develop and briefly defend a partial psychological account of hot-cold empathy gaps. On the partial view we will develop, subjects employ their own current visceral states in order to simulate some situation 'from the inside,' i.e., first-personally, and this simulation yields a prediction about the behavior or preferences of either some other person or else of themselves at a different time.[15] Moreover, subjects by default treat their current states as inputs to this prediction. On our substantive view of hot-cold empathy gaps, these gaps arise when one 'projects' one's own current visceral state into some kind of subjective rendering of a scenario. Specifically, these gaps arise when this process of projection yields a distorted representation of some subject's preferences. This hypothesis we dub the *first-personal projection view*.

This account is neutral on further questions about how empathy gaps occur. In particular, it is neutral on the questions of whether the relevant first-personal process of prediction is inoculated from broader background theories, and whether it is encoded imagistically rather than propositionally, graphically, or otherwise.

We will often refer to the predictive process which underlies hot-cold gaps as one in which a subject *simulates* experience, by which we mean she somehow models it 'from the inside' or with respect to a subjective 'I.' It does not matter for our purposes whether this process is achieved imagistically, but the process will tend to involve some 'felt' quality insofar as it involves a model constructed around one's current visceral state. To illustrate the view with an example: When Calvin makes a prediction about how much food he will want when very hungry, he does so partly on the basis of his current feelings of hunger when imagining that prospect, and not (say) on the basis of third-personal information about what someone in his state would likely want.

Before proceeding, a point of terminology. On the view we will defend, in making predictions about visceral states, subjects tend to import their current visceral states into a simulation, regardless of whether doing so is appropriate. This tendency we have already been calling *projection*. Thus, projection is a sub-species of simulation, where simulation is the more neutral way to describe any kind of first-personal modeling, and projection in particular is the version of this first-personal modeling in which subjects import their own visceral state into that model.

1.1 Evidence for the 'first-personal projection' view

First, here is why we take hot-cold gaps to arise because of some form of first-personal process, as opposed to a process of third-personal prediction. Multiple studies have shown that, when asked about what someone else might be thinking or feel-

[15] In contrast to how we use the term 'simulation,' the broader literature does not always use this term to pick out first-personal processes. See, e.g., Gerstenberg and Tenenbaum (2017) and Moulton and Kosslyn (2009). Our view is similar to the simulation view in the broader debate about mindreading. However, unlike most species of the simulation view, we are neutral for present purposes on whether the first-personal predictive process proceeds by (perhaps implicit) theorizing and whether it is guided by mental imagery. For a helpful recent overview of this issue, see Barlassina and Gordon (2017).

ing, participants tend to first-personally imagine the circumstances that the other person is in.[16] There is also evidence that asking participants to imagine the feelings of someone else tends to increase the activation of the participants' own self-focused thoughts, relative to a control.[17] We take these studies to show that people tend, by default, to represent others' situations 'from the inside' when making predictions about those people. Additionally, other results suggest that people first-personally project at least as frequently when making predictions about their own experiences as they do when making predictions about others' experiences.[18]

Second, here is why we take empathy gaps to arise because of a process of projection, whereby subjects by default treat their own visceral states as inputs to a prediction. This 'projection' view straightforwardly explains the fact that hot-cold predictions skew in the direction of the predictor's current visceral states. In contrast, it's less obvious how theory-based views would explain this result, since there is evidence that subjects' broader theories about people's preferences do not explain empathy gap results.[19]

1.2 The limits of first-personal projection: When simulating others' states is not possible

So far, we've suggested that hot-cold gaps emerge by way of projection, wherein subjects treat their current visceral states as default inputs for a simulation. In some cases, these predictions are inaccurate. We now turn to evidence that subjects typically cannot correct their simulations so as to avoid these erroneous predictions. This isn't to suggest that there is *no* way by which they might correct the predictions, but the possibility of doing so is at least seriously constricted. Moreover, there is a lack of evidence to indicate that subjects can systematically overcome these gaps.

On its face, one might think that predictive gaps emerging from a process of simulation *should* be surmountable by a relevant process of counter-simulation; after all, one might think subjects can simply import *a different hot or cold state than their own* into the simulation to yield a better prediction. But several results tell against this suggestion. In particular, for especially intense states and for states of certain kinds (e.g. pain), it is not obviously possible for subjects to simulate visceral states that are incongruent with their current actual states. For instance, it is not obviously possible to fully simulate intense pain when one is currently not in any pain at all.

A first result which suggests that subjects cannot use simulation to correct their erroneous hot-cold predictions comes from Nordgren et al. (2006). In this study, fatigued participants were asked to assess some non-fatigued person's behavior by pretending that they were also non-fatigued; this instruction did not alter the predictive gap. The same result held of non-fatigued participants asked to assess a fatigued individual by pretending they were fatigued. These results led the study's authors to

[16] Berntsen and Jacobsen (2008), Depow et al. (2021), Van Boven and Loewenstein (2003), Van Boven et al. (2013)

[17] Davis et al. (2004).

[18] Pronin and Ross (2006), Gilbert and Wilson (2007), Van Boven et al. (2013).

[19] Steinmetz et al. (2018).

conclude that "… efforts designed to help people overcome empathy gaps are likely to be unsuccessful." (Nordgren, p. 638).[20]

A second set of results which suggests that subjects struggle to correct their simulation-based predictions comes from results concerning pain, especially extreme pain. Some evidence suggests that those who have undergone extreme pain but are not currently experiencing such pain—such as those who have previously given birth—struggle to accurately predict the behaviors of those currently in extreme pain. Montero (2020) argues on the basis of this and other evidence that the 'felt' component of pain is in principle inaccessible in memory.[21] If Montero is right, and if simulations of pain would require the reconjuring of pain on the basis of phenomenal pain memories, then predictive gaps related to extreme pain cannot be bridged through a process of simulation, as Montero herself argues.[22]

Why would it be impossible for subjects to fully close intrapersonal empathy gaps, e.g., why would it be impossible for someone who is in the throes of an intense nicotine craving to simulate a feeling of not craving at all? We take this to be an open empirical matter, but one provisional answer would draw partly on the claim that humans have little way of controlling their visceral states without altering the relevant biological state (e.g., by smoking a cigarette). If imagined visceral states are additionally constrained by actual visceral states, then we should expect that hot-cold gaps cannot be closed without changes to the relevant biological state. Whether, or to what extent, we have such control may depend on the type of state involved. A more complete explanation requires further empirical research.[23]

We conclude that there are good reasons to think that, in at least some cases, it is not possible for subjects to simulate hot or cold states different than their own, perhaps especially when therelevant states are dramatically different than one's current state. To say that empathy gaps cannot be overcome by a process of counter-simulation is not, of course, to deny that the *behavioral* consequences of empathy gaps cannot be altered, for instance, by an agent's choosing to abide by certain time-tested heuristics or rules. Consider: the person who finds herself very hungry while at the grocery store, doing her food shopping for the week, might choose to stick to a premade shopping list, perhaps one made by looking at her average food consumption in the past, rather than relying on her current visceral state to make predictions about how much food she will need for the week. This choice might well prevent this agent

[20] Results from Loewenstein, Prelec, & Shatto (1998) further suggest that subjects cannot correct empathy gaps even when motivated to do so.

[21] See Christensen-Szalanski (1984) & Morley (1993) for examples of corroborating evidence.

[22] Montero (2020: 119–122). See also Read and Loewenstein (1999). An important challenge for our view is from Steinmetz et al. (2018), which might initially seem to suggest that subjects can counter-simulate in some cases. We think that closer inspection shows that these results do not suggest that this is possible, especially for intense states. For one thing, the Steinmetz et al. (2018) result involved small effect sizes which were not measured against subjects' baseline states, which opens up other interpretations of the results.

[23] There is a separate, significant literature on emotion regulation that suggests that cognitive reappraisal of events can aid emotion regulation, suggesting that there are some cognitive mechanisms by which we can control our emotional arousal. See, e.g., McRae and Gross (2020) for a recent review. We are not presuming that what holds for visceral states holds for emotional states, but there might be helpful parallels between the kinds.

from buying more food than she needs, but *it doesn't alter her first-personal predictive process*; it merely insulates her food decisions from that process, by treating an alternative source of evidence, one rooted in third-personal evidence about her past behavior, as a superior basis for action.

2 Hot-cold empathy gaps and knowledge

We turn, in this section and the next, to considering some of the philosophical implications of hot-cold empathy gaps. In this section, we will argue that the predictions which figure in hot-cold gaps do not typically amount to knowledge, even when they happen to be accurate. Ultimately, in the next section, we will draw on this result to sketch a challenge for L.A. Paul's view that authentic action requires a form of experience-based knowledge of one's values and how those values map onto relevant outcomes.

In some cases, the predictions generated by projecting one's own hot or cold states into a simulation will not generate true beliefs; indeed, the psychological literature focuses on these cases. For instance, if Nadia is not currently craving a cigarette, she might underestimate the extent to which later, when she is craving one, it will be hard for her to resist lighting up, even if she has a long-term goal of quitting smoking. Or, if Nadia is exhausted from work, she might underestimate how much her well-rested partner will want to go out for dinner later.

However, in some cases, the process which underpins hot-cold gaps will result in *accurate* predictions. This will happen whenever the target of the prediction happens to be in the same visceral state as the predictor. For instance, suppose Claire is feeling energetic, i.e., extremely non-fatigued, when deciding whether to sign up for a mountain trek later, while on vacation. As a result of importing her current energetic state into a simulation of going on the trek, she concludes that she would very much enjoy going on the trek. Further suppose that, as it turns out, Claire *will* happen to be feeling energetic during the scheduled trek, so she will enjoy the excursion. In this case, the strategy of importing one's current hot state into a simulation of a future event yields a correct verdict.

So, the process which underpins hot-cold empathy gaps sometimes yields accurate predictions, and sometimes does not. But we can further ask: when the process of projecting one's own states into a simulation generates accurate predictions, do those accurate predictions tend to amount to knowledge?

Here are two reasons to doubt that predictions generated from projection in a simulation typically constitute knowledge, even when those predictions are accurate. First, there is something inherently odd about the process, in the sense that there is no obvious reason why (say) the fact that one is currently hungry should increase the odds that one will oneself be hungry at some arbitrary later time. Nor does it obviously increase the odds that some arbitrary other person is currently hungry. Consider that whether someone is in one of the states in question—hunger, thirst, pain, fatigue, sexual arousal, and the like—tends to vary over the course of a day in the same person and tends to also vary between people at a given time. Our suggestion is not that the *felt* aspect of these states is highly variable across persons—these might well bear

important similarities between agents—but rather that the fact of whether one is (say) hungry versus sated will naturally tend to vary in cyclical ways. The result is that the transition from (say) "I'm hungry" to "you are hungry" is an extremely odd one; were it an inference, for instance, it would not be a valid inference. So, there is something odd about the internal 'logic' of the process. At least on its face, it's not clear why we should expect this process to be truth-preserving, let alone knowledge-conferring. Someone can be hungry now without this fact being a guide to whether that same person will be hungry later; someone can be tired now without this fact being a guide to whether their toddler is also tired now.

Second, the process itself will not typically be reliable, despite the fact that it sometimes yields accurate predictions. On some views, knowledge can *only* be generated by a reliable method. But even those who reject this view might well grant that the reliability of some process is a *defeasible indicator* of whether that process generates knowledge.[24]

To illustrate that the process which underlies hot-cold prediction is often unreliable, consider in particular *safety*, where some process is safe just in case: in most or all near worlds where you form that belief on the basis of that method, that belief is true (Williamson, 2000; Pritchard, 2009; Sosa, 1999). In contemporary work on the connection between knowledge and reliability, safety is the most often-discussed variant of reliability. So, we will take it that if the method of first-personal simulation is not safe, this is a good if defeasible reason to think that this method does not satisfy whatever kind of reliability (if any) which is either a condition on knowledge or a defeasible indicator of knowledge.

To see that the process of first-personal simulation is not safe, consider again Claire's true belief that she will enjoy the mountain excursion during her future vacation. Claire's judgment is formed because she imports her current energetic state into a simulation about what it would be like for her to undertake the excursion, and she exploits that simulation to derive the judgment that she would enjoy the excursion. Does this method produce true beliefs in most or all near worlds, as safety requires? Assuming Claire's states of energy and fatigue ebb and flow much like the rest of ours, based on a complex of biological and situational factors, the answer must be 'no.' For, in some near worlds, Claire is exhausted when she considers whether she would enjoy the trip even though—in that world—she would happen to be in a well-rested state before the excursion itself. In other near worlds, Claire is in an energetic state when she considers whether she would enjoy the trip even though she will be worn out from travel and work obligations just prior to her trip.

We claim that the process of projection is typically unreliable, but we do not deny that it is ever reliable. Certainly, this process might be reliable for some individuals and some kinds of visceral states. For instance, two people who tend to eat together might tend to have similar cycles of hunger and satiation and thus, their first-personal-based predictions about the other's current hunger might be reliable. But in any context where such visceral states are not synced, the process of projection will

[24] One of the authors is a theorist of this stripe, having argued that safety is not necessary for knowledge (Helton & Nanay, 2019). This view is consistent with the view that safety is a good if defeasible proxy for whether some process is knowledge-conferring.

not be reliable, and these contexts are extremely common. So, at least many such processes will be unreliable. This is a reason to think that the process of simulation does not typically generate knowledge pertinent to what one's visceral states will be at a different time.

We conclude that the process which underpins hot-cold empathy gaps is (at least often) not safe. We take this to be a good if defeasible reason to doubt that this method is generally knowledge-conferring, even when its predictions happen to be accurate. Certainly, there is more to say on this matter. For one thing, there are other ways a process might be knowledge-conferring which have to do neither with its reliability nor with its internal logic.[25] But, we take the preceding considerations to suggest that the process of hot-cold simulation is not typically knowledge-conferring.

One might object to our claim that hot-cold simulation is not typically reliable that this process could be *made* reliable, so long as one builds into the relevant simulation enough pertinent details. For instance, perhaps Claire can imagine rather vividly the likely weather and mountain conditions of the proposed trek and in this way bypass the documented tendency to import one's current visceral state into simulations about what one would choose to do at a different time.[26] There are two things to say about this. First, we take it that what matters for knowing what one would want in the relevant cases is at least significantly to do with what one's relevant visceral state will be like, and not just to do with what environmental conditions would be like. For instance, what matters for Claire getting it right is whether she imports into the simulation the fatigue she will in fact feel after travel, and not (just) what the environmental conditions will be.

Second, we take the empirical evidence discussed in Sect. 1.2 to powerfully suggest that it is at least extremely difficult (and perhaps impossible) for subjects to vividly counter-simulate visceral states different than the ones they're currently in. So, for instance, it will be at least extremely difficult for Claire to vividly imagine that she is fatigued when she is in an energetic state, with the result that her attempt to use simulation itself to make the decision will tend to be unreliable. This is precisely why hot-cold gaps arise and persist even in the face of subjects' efforts to overcome them.

Here is a different objection one might make to our claim that hot-cold predictive processes are not typically knowledge-conferring. Perhaps someone who is aware of the empirical reality of hot-cold gaps could exploit this theoretical knowledge to render her own predictions more reliable in the following way: She could decide to use this kind of predictive strategy only when she independently knows that she is in roughly the same visceral state as the one which is pertinent to some future choice.[27] For instance, consider again Claire, who is deciding whether to book a mountain excursion on her next trip. She might strategically decide to make this decision when she's very tired, as she anticipates that, after traveling to her destination, she will be

[25] For instance, the process might have a kind of *presentational phenomenology* which helps render it potentially knowledge-producing. See, e.g., Bengson (2015) and Chudnoff (2012). Or it might disclose truth-makers in a way that renders it especially epistemically valuable (Johnston, 2006, 2011). Or it might confer certain relevant capacities or competences, see, e.g., Schellenberg (2018) and Miracchi (2015).

[26] We thank an anonymous referee for this objection.

[27] We thank Olivia Bailey for this suggestion and for helpful discussion on this point.

very tired. Deliberately waiting until she's in the 'right' state might help her make a good choice, even using the predictive process of simulation.

We accept that this strategy of employing higher-order knowledge to exploit one's first-order predictive process is likely to increase the reliability of one's predictions, perhaps to the extent that these predictions qualify as knowledge.[28] Indeed, we think that the platitude mentioned in the paper's introduction, *never go grocery shopping while hungry*, likely functions as a specific version of this broader epistemic strategy. By making sure you are in the relevant state (relatively satiated) while making a choice about what to eat later, you ensure that your natural predictive tendencies come closer to what you will want to eat in the future.

At the same time that we think it likely that using higher-order knowledge about hot-cold gaps might, in some cases, permit one to exploit them in the suggested way, thus potentially producing knowledge, we do not take this result to be in tension with our claim. Our claim is that first-personal prediction *itself* is not typically knowledge-conferring. We do not deny that this process, *properly embedded within some broader process*, might be knowledge-conferring. Moreover, and more relevantly to our later points in the paper, we think this higher-order strategy will often be difficult or impossible to implement, such that ordinarily, hot-cold predictions will not be knowledge-conferring. Sometimes one must go grocery shopping while hungry, and sometimes one must make a choice about what to do on a trip when one is well-rested.

Going forward, we will help ourselves to the assumption that first-personal simulation across hot-cold gaps is at least often not knowledge-conferring. We turn to drawing out what, if anything, this result tells us about the scope of an important notion of authenticity developed and defended by L.A. Paul.

3 Hot-cold empathy gaps, authenticity, and the limits of subjective knowledge

Intuitively, some of our actions reflect our values and some do not, either because those actions are neutral with respect to our values or else because they fly in the face of our values. For instance, if Calvin values honesty and, guided by this value, discloses something embarrassing but important about his past to his new romantic partner, his disclosure reflects his values in at least some way. Or, if Nadia values being a reliable friend but, due to an ongoing dependence on alcohol, struggles to keep promises to her friends, Nadia's failure to keep her promises does not reflect the value she places on being a reliable friend.

While acting in a way that reflects one's values is not the only standard by which actions might be assessed, we take it to be an important one and one that many of us care about.[29] Following L.A. Paul, we will dub actions which reflect one's values

[28] We leave it open whether the internal structure of this strategy permits that it might be knowledge-conferring, given the odd 'logic' or inferential structure it exploits.

[29] In addition to assessing an action along the familiar lines, such as with respect to its morality, its aesthetic qualities, or its practical rationality, we can also assess it on any other number of grounds, such as whether it exhibits spontaneous freedom (Gingerich, 2022) or whether it exhibits shared improvisational

in the right way *authentic* ones and those which do not reflect one's values in the right way *inauthentic* ones.[30] In making this terminological point, we do not mean to suggest that other views of 'authenticity' are inapt. What we care about is value-reflectance, and the choice of terminology is arbitrary.

What is required for an action to be authentic in the sense of reflecting one's values? More particularly, what epistemic situation must one be in to carry out authentic action? Paul has argued that in at least some cases, authentic action requires knowledge of one's values obtained from first-personal experience. These experiences permit a kind of grasp of one's own values, a way of understanding them which is importantly different than other ways by which one might come to know about one's own values, such as via testimony from others about what one values. The kind of grasp Paul is concerned with is an experiential kind, and Paul uses the 'Mary' thought experiment to draw the distinction: Before leaving the black-and-white room, Mary has a purely cognitive understanding of color experience, but after leaving it and seeing color for the first time, she attains an experience-based grasp of color experience, an understanding that is somehow distinct from her purely cognitive understanding of color vision, even if the facts which are understood are the same.

As already mentioned, we employ the term *subjective knowledge* as short-hand for first-personal-experience-based knowledge of one's own values. By employing this term, we do not mean to suggest that the knowledge *itself* is somehow subjective. For instance, we are neutral on the question of whether the pertinent form of knowledge is itself influenced by interests in some way.[31] Rather, we reserve 'subjective knowledge' to refer to the kind of knowledge that is acquired via first-personal modes of experience and is to do with a subject's *own* values (whether or not those values are generally held or objective).

These remarks from Paul are representative of her view about the connection between authenticity and subjective knowledge:

> Authentic decision-making can require imaginative knowledge of what my future circumstances will be, where such imaginative knowledge carries with it a direct affective, emotional engagement that allows me to cognitively and emotionally empathize with my possible future selves.[32]

Notice two things about this passage. First, Paul suggests (by pragmatics) that authentic choice needn't require 'imaginative knowledge,' though it 'can require' it. (For Paul, 'imaginative knowledge' is a species of subjective knowledge). That is, this form of subjective knowledge is necessary for authentic choice but only *in some range of cases*. Second, on Paul's conception, it is *knowledge* produced by imagina-

agency (Bagley, 2013, 2015). We pursue a thoroughgoing pluralism about the standards along which action might be assessed. Thanks to Sophie Dandelet for discussion on this point.

[30] Paul (2014: 105–107) and Paul (2015a: 761–762).

[31] That is, this terminology is meant to be neutral about pragmatism about knowledge. See Kim (2017) for a helpful overview.

[32] Paul (2015b: 810). As Paul uses the terms, the relevant 'imaginative knowledge' is knowledge of the subjective value of an outcome that involves lived experience, where the knowledge is attained by imaginatively prefiguring that experience. That is, 'imaginative knowledge' is a kind of subjective knowledge.

tion (or experience) that is, in at least some cases, required for authentic decision-making.[33] As we are interpreting Paul on the basis of this and other passages, her total view of the relation between authenticity and experience is this:

> AUTHENTICITY FROM SUBJECTIVE KNOWLEDGE.
> All else being equal, authentic action requires first-personal-experience-based knowledge of one's values, i.e., subjective knowledge.

For short, we sometimes refer to this view as *the subjective knowledge view*. We will say more shortly about Paul's view of what else must be equal for this requirement to be triggered. For now, it is important that while Paul often speaks of subjective knowledge being achieved imaginatively, her fuller view is rather that this subjective knowledge can include knowledge of memories and also cognitive encodings of information, so long as this information is first-personally indexed and derived from some sampling of one's total experience during some duration.[34]

Paul motivates the subjective knowledge view of authentic decision-making in several ways, just some of which are these: First, she draws on certain compelling thought experiments to motivate the view. For instance, she suggests that were a person to decide whether to have a child by relying solely on the testimony of others about what parenthood is like, this would be odd and intuitively inauthentic.[35] Second, she suggests that the kind of grasp conferred by subjective knowledge of one's values permits a 'sense of control' in one's choices; the link seems to be that better control over one's actions would tend to increase the odds that one's actions will reflect one's values.[36] Third, Paul suggests that authentic preferences should be formed on the basis of subjective values, where these values are partly constituted by, but not exhausted by aspects of experience.[37] Together, these claims characterize an intuitive ideal of knowledgeably navigating the world guided by one's own first-personal evaluative perspective.

We find these motivations to be extremely compelling ones, such that we think the thesis of authenticity from subjective knowledge merits serious consideration.[38] At the same time, we will suggest that considerations from hot-cold gaps show that the requirement that authentic action be governed by subjective knowledge is triggered

[33] On the first point, see, e.g., Paul (2020). For complementary discussion by Paul on the connection between knowledge and authentic experience, see, e.g., (Paul, 2015b: 808, 810). For a metaphysics of the phenomenal feel which figures in simulation and experience, seePaul (2017b).

[34] Paul (2014: 106).

[35] For related points, see, e.g., Paul (2014: 75) and Paul (2015b: 809, 811–813).

[36] Paul (2014: 107).

[37] Paul (2014, 2015b, 2017a). See also Paul (2015b: "Reply to Campbell") for the view that a subject's values are partly individuated by worldly features.

[38] We are sympathetic to the thought–helpfully raised to us by Jane Friedman–that it isn't knowledge at all which ought to ground authentic action, but is rather some other kind of state altogether, such as belief, justification, or something else. The rival view we will sketch appeals to understanding, and we are neutral on the controversy over whether understanding is a form of knowledge. So, for present purposes, we are neutral on the question of whether the epistemic grounds of authenticity ought to be knowledge or some other epistemic good.

in far fewer circumstances than one might have thought. To be clear, Paul does not suggest that subjective knowledge is always required for authentic decision-making, and relatedly, we do not take the points we will make shortly to suggest a counter-example to her claim. Rather, we will argue that there are few cases in which Paul's requirement will be triggered. This result in turn motivates a re-evaluation of Paul's theory, with an eye to whether some broader theory might subsume hers. We will ultimately sketch such a theory.

To mount this argument, we will draw on three claims previously defended: First, hot-cold empathy gaps derive from a first-personal predictive process. Second, this process does not confer knowledge, even when the judgments it produces are accurate. Third, in at least some cases, subjects cannot 'correct' this process via a process of counter-simulation. As a result, in at least some cases, subjects cannot make choices involving hot or cold states that are based on subjective knowledge, even where experience is broadly construed to include the sum of one's present perceptual, emotional, cognitive, and other processes. For, as a psychological matter, subjects often simply cannot simulate in a way that would produce this sort of knowledge.

For the sake of keeping things concrete, consider how these claims apply in specific cases. Recall Nadia, the smoker who is trying to quit. With respect to Nadia, our claims are these: first, when Nadia is in this non-craving state, she is likely to rely on this state in making a prediction about her later preferences. Second, this prediction does not constitute knowledge, whether or not it happens to be correct in this particular instance. Third, while Nadia is in this non-craving state, it is not psychologically possible for her to simulate otherwise, i.e., to simulate what it would be like for her to (say) walk past her cigarettes when she is craving a cigarette. So, Nadia's choice to leave her cigarettes where they are, lying on the coffee table, is not based on knowledge from simulation.

The same claims apply to Claire, who, while in an energetic state, considers whether to book a mountain excursion for a later vacation. Because Claire's assessment of the trek is influenced by her energetic state, her choice is not based on subjective knowledge of her values. This is because the kind of simulation she employs isn't knowledge-conferring, even if accurate. Moreover, it is not psychologically possible for Claire to make this choice in a way that is based on knowledge from simulation. She cannot correct the first-personal predictive process to make it a knowledge-conferring process because, while in a highly energetic state, she cannot fully simulate the state of being fatigued.

Consider what these results mean for Paul's view of authenticity, on which, *ceteris paribus*, authenticity requires subjective knowledge of one's values. On this view, at least one of the following claims must be true of choices such as Nadia's choice concerning her cigarettes: These choices are not assessable with respect to authenticity; these choices do not trigger the requirement of subjective knowledge; and/or these choices are systematically inauthentic.

We will argue that that these choices are assessable with respect to authenticity and also, more briefly, that they are not systematically inauthentic. This leaves the result that these kinds of choices do not trigger Paul's requirement of subjective knowledge. We think this result can be defended on non-ad hoc grounds, ones Paul herself develops. But we will ultimately suggest that this result should motivate us

to reconsider the subjective knowledge view itself, since the view turns out to be of limited scope.

First, to defend the thought that choices about leaving cigarettes out or booking a mountain excursion are assessable with respect to authenticity. One might reasonably wonder whether these micro-decisions can reflect one's values in whatever way authenticity requires. So, one might think these decisions are not assessable with respect to authenticity, perhaps because they are comparatively insignificant.

Certainly, these are not the kinds of choices which tend to animate discussions about authenticity. More commonly, this literature focuses on dramatic, one-off decisions, such as whether one should move to another country or whether one should have a child. Indeed, many of the common examples involve what Paul dubs *transformative experiences*, where these are experiences which alter oneself in a deep way and which are such that one cannot truly know what they are like until one has had them. The decision about whether to leave the cigarettes out or whether to go on a mountain hike are extremely unlikely to deeply change oneself in the way transformative experiences do.[39]

There are two things to say about the concern that the kinds of prosaic choices which often figure in hot-cold decisions are not significant enough to be evaluable with respect to authenticity. First, while some hot-cold decisions are not likely to be life-changing, others are potentially transformative. For instance, the choice about whether to enter rehab to treat one's addiction is potentially a choice made at least partly on the basis of hot-cold simulations, insofar as it is much easier to decide to enter rehab when not currently craving a substance. This choice is also potentially transformative, since it might alter a person's deep self and might involve experiences one cannot fully grasp in advance.

Second, we would resist the suggestion that even prosaic, non-transformative hot-cold decisions are not sufficiently tied up with one's values to qualify as either authentic or inauthentic. For arguably, an authentic *life* is no less made up of the countless micro-decisions one makes throughout the day than of those large decisions which keep us up at night. These small decisions are on their own of little consequence, but together they make up much of what a life *is*: what kind of diet you and your loved ones have, whether you travel, whether you smoke, whether you make plans with friends—with all of the attendant sequelae these choices have for susceptibility to depression and other illness, the quality of one's social relationships, the nature of one's accomplishments, and on. When it comes to an authentic life, these micro-decisions are not minutiae; they help make up the fabric of an authentic life, if not the whole of it.

Our first conclusion then, is this: Decisions made at least partly on the basis of hot or cold states *are assessable with respect to authenticity.* Given this result, and given that such choices are not systematically inauthentic (a point to which we return shortly), Paul's view of authenticity forces us to the view that these choices are not the kinds of choices which trigger the requirement of subjective knowledge.

At this point, Paul might reasonably object to our suggestion that hot-cold decisions cannot be based on subjective knowledge. While Paul's view is that knowl-

[39] Paul (2014).

edge obtained by a sampling of one's experience is sometimes required for authentic choice, recall that this sampling involves *all* subjective aspects of one's experience; it isn't restricted to one's visceral states. It can extend, for instance, to subjectively encoded memories or to other first-personally presented forms of knowledge. So, Paul might maintain that, even if the hot-cold predictive process alone cannot generate knowledge of another person or one's future self, other elements of one's psychology can serve this role, even when it comes to hot-cold decision-making. For instance, Nadia might remember that whenever she has left out her cigarettes in the past, she tends to light up shortly thereafter. So, her memory might serve as a form of subjective knowledge which can ground her authentic choice not to leave her cigarettes out, even if the hot-cold predictive process itself cannot play this role.[40]

We acknowledge that the epistemic limitations of hot-cold gaps do not, in their own right, create a difficulty for Paul's subjective knowledge view, precisely because Paul's view encompasses forms of knowledge which go beyond simulation from visceral states. At the same time, we think there are additional reasons to think that in many hot-cold cases, these additional forms of knowledge are either not subjective or else are inadequate to ground knowledge about the relevant choice. As a result, in at least some such cases (and perhaps most), the subject will lack subjective knowledge which might ground her decision.

For instance, consider the suggestion that memories which might inform one's hot-cold decision, such as Nadia's memory that whenever she leaves the cigarettes out, she tends to smoke later. If this memory is encoded third-personally—for instance, if Nadia merely remembers *that* she does this, without accessing the values which explained her behavior in a subjective way—then the knowledge is not subjective in the relevant way.

If, on the other hand, Nadia's memory encodes in a 'felt,' first-personal way the values which explained why she smoked in those cases, e.g., if she remembers the intense craving which accompanied that action, then that knowledge *is* subjective and *is* potentially the sort of knowledge which might ground an authentic decision in the way Paul's view requires. But, we think it empirically implausible that, whilst in a non-craving state, Nadia will be able to access this memory of the vivid, first-personal feeling of craving. For, if the typical subject in this case were able to access this feeling of craving, it would be puzzling why she would tend to act in accordance with her current non-craving visceral state. And the literature on hot-cold gaps suggests precisely that she will tend to act in accordance with her current visceral state. This isn't to suggest that it is impossible that she access this memory or that there is no variance across subjects or cases, just that there are independent grounds for thinking that this is typically not the case.[41]

We conclude that choices involving hot-cold states are often made on the basis of simulations which are not themselves knowledge-producing. We also think that in at least many (and perhaps most) such cases, subjects lack access to other subjec-

[40] We thank Laurie Paul for helpful discussion on this point.

[41] See Montero (2020) for an argument that these memories are deeply inaccessible. See Bailey (2023) for a discussion of how one's past *sensibilities*, where these are broad emotional orientations, can become inaccessible to one.

tive grounds for making the choice. Together, we take these claims to force Paul to the view that either: the requirement that authentic action be grounded in subjective knowledge is often not triggered, as in many hot-cold decisions, or: these decisions are systemically inauthentic, as a matter of something like psychological necessity.

We take the result that the relevant decisions are systematically inauthentic, as a matter of something like psychological necessity, to be a highly unattractive one. Briefly, here is why: The standard of authenticity, qua value-reflectance, is presumably constrained by something like psychological ability. That is, in order for some choice to be assessable with respect to authenticity, it must be the kind of choice which, psychologically speaking, one can make in a way which reflects one's values. This is why it is might *not* be inauthentic of Travis to fail to write a deeply moving country song even if he values country music and wishes to express his feelings in this way; perhaps, he simply can't do this, consistent with his current psychological mechanisms and skills. But, this is why it *might* be inauthentic of Travis to stop listening to country music because he fears what others think of his taste in music, despite the fact that he greatly values the art form. It is well within Travis' ability to listen to country music. Psychologically speaking, Travis can do this.

Our suggestion is that there is a kind of 'authenticity' implies 'psychological ability' principle. In order for some action to be the kind that triggers a requirement of authenticity, it must be the kind of action one can, psychologically speaking, make. On this view, it is incoherent that an entire class of choices are systematically inauthentic of something like psychological necessity. For, if these choices are not the kinds of choices which, psychologically speaking, can be authentic, they are not assessable with respect to authenticity.[42]

We conclude that considerations from hot-cold gaps suggest that there are surprisingly few cases in which authentic action requires subjective knowledge. Specifically, no choices driven by hot or cold states require subjective knowledge, and such choices are extremely common. Since Paul's view is that authenticity requires subjective knowledge in only some cases, her view is consistent with this result. Moreover, Paul explicitly considers some cases where subjective knowledge is not available and explicitly argues that in such cases, authentic action can be grounded in third-personal knowledge, such as in knowledge from expert testimony.[43]

Thus, our complaint with Paul's view is not that hot-cold decisions constitute a counterexample to it; by Paul's own lights, the requirement that authentic decisions be based on subjective knowledge is not always triggered. But, we think that hot-cold decisions nevertheless raise a *puzzle* about Paul's view, a puzzle deriving from the ubiquity of hot-cold decisions in human experience: If a great many decisions can be authentic which do not trigger the requirement of subjective knowledge, should we

[42] Needless to say, there is far more to say about this principle, which is a species of 'ought' implies 'can.' See Helton (2020: 513–16) for the defense of a relevantly similar 'epistemic ought' implies 'psychologically can' principle. For an argument that this principle is neutral on the question of whether the relevant psychological ability is under voluntary control, see Helton (2020: 509–12). For an argument that such principles should be construed in terms of an agent's *individual* ability, see Ryan (2003). For recent discussion of 'ought' implies 'can' in an experimental context, see Henne et al. (2016), Hannon (2018), Semler & Henne (2019).

[43] Paul (2015b: 811).

be confident that it is subjective knowledge per se which bears the relevant connection to authentic choice? Or rather, and despite the suasive motivations for the subjective knowledge view, might the epistemic grounds of authenticity be something other than subjective knowledge? More specifically, might there be some more general epistemic condition on authenticity which both explains why subjective knowledge is required in certain contexts and also explains why it is not needed in a great many others?

Several possible answers present themselves. We will briefly sketch and tentatively defend just one of these, what we dub *the understanding view of authenticity*. On this view, authentic decisions should be informed by a kind of *understanding* of one's values and how those values fit with possible outcomes. At a minimum, the relevant form of understanding will involve an appreciation of the explanatory relations between one's values and relevant outcomes.[44]

It seems likely that in at least some cases, the relevant form of understanding can be achieved by subjective forms of knowledge. But we are presuming that, in at least some cases, understanding can be achieved elsewise. For instance, we think it likely that one might come to appreciate what one's values are and how they connect to relevant outcomes via third-personal means, as when a close friend helps one better appreciate what one's values are by pointing out patterns in one's past behavior suggesting what kinds of values might underpin those patterns.[45] If a close friend tells you that you should move to Los Angeles because you love the outdoors, this can help ground not only knowledge of what you should do but knowledge of what *your* values are and how those values connect to some outcome; this is understanding in the relevant sense.[46] To bring this point back to hot-cold decisions, such as decisions about whether to leave the cigarettes out or whether to hike later: Our view is that, as a matter of human psychology, such decisions can only be authentic when grounded by non-experiential sources of understanding, such as through testimony from oneself or others about what one values.

While we think testimony from a friend might, in some cases, generate the relevant form of understanding, we would deny that expert testimony typically generates this form of understanding. For, an expert's information would presumably derive from population-level statistical information, whereas your close friend will have non-statistical evidence about *your* values and how those values relate to your pros-

[44] Some theorists treat understanding as a species of knowledge, whereas others treat knowledge and understanding as disjoint kinds. In construing understanding in terms of an *appreciation* of explanatory relations, we are meaning to be neutral on the question of whether the relevant form of appreciation is knowledge, grasp, or something else. We are also neutral on the question of whether understanding is a kind of cognitive map (Gopnik, Glymour, & Sobel, 2002; Gopnik et al., 2004; Grimm, 2016), a skill (Hills, 2016; Khalifa, 2017; Cf. Sullivan, 2018), a kind of relevance matching (Roush, 2016, 2017), or a propositional attitude. For helpful overviews on recent work on understanding, see: Baumberger, Breisbart, and Brun (2017), Grimm (2021), and Hannon (2021). See McSweeney (2023) for a view of understanding on which it involves a kind of phenomenal experience and a belief. In contrast to this usage, we are here employing 'understanding' in a factive way.

[45] This process might be deeply dynamic, as in the kind of reciprocal form of conversation described by Dover (2022).

[46] See Harman (2009, 2015) for the suggestion that testimony can ground rational action.

pects.[47] In this much, we are in agreement with Paul that testimony from an expert of the kind that does not confer knowledge about one's own values is insufficient to ground authentic choice. For a similar reason, the understanding view of authenticity goes against a view of authenticity inspired by expected utility theory. On such a view, authentic choice only requires knowing the probabilities and the utility values of the outcomes of a choice, such as on the approach taken by Pettigrew (2015).

Given our assumption that understanding can be achieved via routes that don't have to do with subjective knowledge, such as through testimony from a close friend, our view further suggests that: There is a kind of parity between subjective knowledge and other forms of understanding. All forms of understanding are on a par when it comes to grounding authentic action. This is so even though different forms of understanding differ in other respects, such as in the psychological mechanisms which produced them.

One might object to the understanding view on the grounds that it flies in the face of the powerful intuition that subjective knowledge is invariably a superior grounds for authentic action. Paul presses this point in many contexts, often returning to the example of Mary in her black-and-white room. Before leaving the room, Mary the vision neuroscientist might understand color experience in *some* sense, but the phenomenal grasp she gains upon leaving the room and seeing red for the first time is intuitively better, both on epistemic grounds and for some action-guidance purposes, than anything she might have learnt from a textbook.

We understand the appeal of the suggestion that the kind of understanding which is generated by first-personal experience is intuitively superior to other forms of understanding one might have, such as understanding derived from a close friend's testimony about one. At the same time, we think it is worth taking seriously that this intuition might be incorrect. In a conjectural mode, we suggest how this intuition might be debunked: Perhaps there is something 'seductive' about evidence gotten from one's own experience, something which makes subjects trust them even when they are not trustworthy. If this is so, then perhaps in making judgments about the value of subjective knowledge, we ourselves are hostage to this seductive quality of experience.

Along these lines, there is evidence that subjects treat their own psychological simulations as good sources of information, even in contexts where they are not reliable.[48] For instance, one study showed that participants' memories were superior as a basis for certain predictions than their psychological simulations but that subjects nonetheless relied on their simulations in forming their predictions.[49] It turned out that what explained this reliance on simulation was the phenomenal vividness of subjects' simulations as compared to their memories. Likewise, subjects routinely make hot-cold predictions on the basis of simulations despite the fact that these simulations are, we have argued, not typically knowledge-producing.

[47] The difference between statistical and non-statistical evidence is often held to be normatively significant in other epistemic contexts. See, e.g., Buchak (2014) and Moss (2022).

[48] Chituc et al. (2021); Kappes and Morewedge (2016); Levine et al. (2020).

[49] Levine et al. (2020).

Together these results suggest that perhaps psychological simulations confer a *feeling* of understanding, one which results in subjects' relying on those simulations, whether or not the simulations are reliable. When this feeling of understanding is spurious, for the reason that the simulations are inaccurate and thus cannot confer understanding, simulations generate what C. Thi Nguyen has termed a 'seduction of clarity.' They give their subjects a feeling of understanding which is, in at least some cases, not accompanied by actual understanding. On Nguyen's view, what is 'seductive' about this feeling is that it can prompt subjects to terminate a process of inquiry, potentially prematurely, and to trust the source of the feeling, whether or not the source is an epistemically good one.[50]

Notably, the suggestion that one's own experience might be 'seductive' in the relevant sense is consistent with the view that subjective experience can reveal the deep nature of at least certain aspects of its objects.[51] For instance, consider the case in which Claire is currently well-rested and attempts to imagine what it will be like to go on a mountain trek later. The hot-cold results predict that she will imagine that this trek will be refreshing and invigorating. In reality, (let's suppose) she will be exhausted from an international flight and the trek will be trying and miserable. To say that Claire's simulation gets some things wrong about this case isn't to deny that it might also get some things right. For instance, Claire's imagining might well reveal something about the deep nature of (say) *feeling* invigorated. In this way, the simulation might have a kind of undeniable epistemic value. But this simulation will also get something wrong, insofar as it will suggest that this trek *will be* invigorating, instead of excruciating. And when it comes to Claire's choice, *this* prediction is the relevant one. Our suggestion is that Claire's simulation is 'seductive' in Nguyen's sense if she is inclined to trust it over other, better sources of evidence about what her experience of the trek will be, such as (say) a note she wrote to herself after her last trip, reminding herself not to do so many activities on vacation.

Now for the speculative bit: Perhaps the fact that first-personal imagination in general tends to produce a strong feeling of understanding, one which is nevertheless at least sometimes not accompanied by actual understanding, explains the powerful intuition that subjective knowledge is invariably superior to non-subjective knowledge as a grounds of authentic action. If first-personal imagination produces a 'seduction of clarity' in Nguyen's sense, perhaps this felt sense *also* drives an intuition that (say) the only authentic way for Claire to decide whether to go on the mountain trek is to vividly imagine the experience in detail, noting how she would feel in such a situation. We, being humans, find this thought extremely compelling for the same reason that we find our own experiences trustworthy: they generate a powerful feeling of understanding, one rooted in the vividness of those experiences, whether or not the relevant experience is accurate, reliable, or otherwise epistemically worthy. If this is the case, then the total theory of the epistemic grounds of authenticity needn't accommodate the intuition that where available, subjective knowledge is invariably the best grounds of authenticity. For, this intuition might be inaccurate for reasons explicable in terms of established psychological mechanisms.

[50] Nguyen (2021).

[51] See, e.g., Johnston's (1992) discussion of this thesis in the context of color perception.

We present these remarks concerning the epistemic grounds of authenticity as invitations to further inquiry.

Acknowledgements For helpful discussions on this paper, we are indebted to: Josh Armstrong, Olivia Bailey, Molly Crockett, Sophie Dandelet, Daniela Dover, Jane Friedman, Ram Neta, Laurie Paul, Vida Yao, and the participants of the Transformative Experience workshop at Yale University.

Declarations

Conflict of interest We, the authors, have no conflicts of interest to report.

References

Bagley, B. (2013). *Improvisational agency* (Doctoral dissertation, The University of North Carolina at Chapel Hill).
Bagley, B. (2015). Loving someone in particular. *Ethics*, *125*(2), 477–507.
Bailey, O. (2018). Empathy and testimonial trust. *Royal Institute of Philosophy Supplements*, *84*, 139–160.
Bailey, O. (2021). Empathy with vicious perspectives? A puzzle about the moral limits of empathetic imagination. *Synthese*, *199*(3), 9621–9647.
Bailey, O. (2023). What must be lost: On retrospection, authenticity, and some neglected costs of transformation. *Synthese*, *201*(6).
Barlassina, L., & Gordon, R. (2017). Folk psychology as mental simulation. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation/.
Baumberger, C., Beisbart, C., & Brun, G. (2017). What is understanding? An overview of recent debates in epistemology and philosophy of science. *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, 1–34.
Bengson, J. (2015). The intellectual given. *Mind*, *124*(495), 707–760.
Berntsen, D., & Jacobsen, A. S. (2008). Involuntary (spontaneous) mental time travel into the past and future. *Consciousness and Cognition*, *17*, 1093–1104.
Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, *169*(2), 285–311. http://www.jstor.org/stable/42920418.
Chituc, V., Paul, L., & Crockett, M. (2021). Evaluating Transformative Decisions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43.
Christensen-Szalanski, J. J. (1984). Discount functions and the measurement of patients' values. Women's decisions during childbirth. *Medical Decision Making*, *4*(1), 47–58. https://doi.org/10.1177/0272989X8400400108.
Chudnoff, E. (2012). Presentational phenomenology. *Consciousness and Subjectivity*, 51–72.
Davis, M. H., Soderlund, T., Cole, J., Gadol, E., Kute, M., Myers, M., & Weihing, J. (2004). Cognitions associated with attempts to empathize: How do we imagine the perspective of another? *Personality & Social Psychology Bulletin*, *30*(12), 1625–1635. https://doi.org/10.1177/0146167204271183.
Depow, G. J., Francis, Z., & Inzlicht, M. (2021). The experience of empathy in everyday life. *Psychological Science*, *32*(8), 1198–1213.
Dover, D. (2022). The conversational self. *Mind*, *131*(521), 193–230.
Fisher, G., & Rangel, A. (2014). Symmetry in cold-to-hot and hot-to-cold valuation gaps. *Psychological Science*, *25*(1), 120–127.
Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 515–547). Oxford University Press.
Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, *317*(5843), 1351–1354. https://doi.org/10.1126/science.1144161.
Gilbert, D. T., Gill, M. J., & Wilson, T. D. (2002). The future is now: Temporal correction in affective forecasting. *Organizational Behavior and Human Decision Processes*, *88*(1), 430–444.
Gingerich, J. (2022). Spontaneous freedom. *Ethics*, *133*(1), 38–71.

Gopnik, A., Glymour, C., & Sobel, D. (2002). Causal maps and Bayes nets: A cognitive and computational account of theory-formation. *The Cognitive Basis of Science*, 117–132.
Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*(1), 3.
Grimm, S. (2016). Understanding and transparency. *Explaining understanding* (pp. 228–245). Routledge.
Grimm, S. (2021). Understanding. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2021/entries/understanding/.
Hannon, M. (2018). Intuitions, reflective judgments, and experimental philosophy. *Synthese*, *195*(9), 4147–4168.
Hannon, M. (2021). Recent work in the epistemology of understanding. *American Philosophical Quarterly*, *58*(3), 269–290.
Harman, E. (2009). I'll be glad I did it reasoning and the significance of future desires. *Philosophical Perspectives*, *23*, 177–199.
Harman, E. (2015). Transformative experiences and reliance on moral testimony. *Res Philosophica*.
Helton, G. (2020). If you can't change what you believe, you don't believe it. *Noûs*, *54*(3), 501–526.
Helton, G. & Nanay, B. (2019). Amodal completion and knowledge. *Analysis*, *79*(3), 415–423.
Henne, P., Chituc, V., De Brigard, F., & Sinnott-Armstrong, W. (2016). An empirical refutation of 'ought' implies 'can'. *Analysis*, *76*(3), 283–290.
Hills, A. (2016). Understanding why. *Noûs*, *50*(4), 661–688.
Hirji, S. (2022). Outrage and the Bounds of Empathy. *Philosophers' Imprint*. 22.
Johnston, M. (1992). How to speak of the colors. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *68*(3), 221–263.
Johnston, M. (2006). Better than mere knowledge? The function of sensory awareness. *Perceptual Experience*, 260–290.
Johnston, M. (2011). On a neglected epistemic virtue. *Philosophical Issues*, *21*, 165–218.
Kappes, H. B., & Morewedge, C. K. (2016). Mental simulation as substitute for experience. *Social and Personality Psychology Compass*, *10*, 405–420. https://doi.org/10.1111/spc3.12257.
Khalifa, K. (2017). *Understanding, explanation, and scientific knowledge*. Cambridge University Press.
Kim, B. (2017). Pragmatic encroachment in epistemology. *Philosophy Compass*, *12*(5), e12415.
Levine, L. J., Lench, H. C., Stark, C. E., Carlson, S. J., Carpenter, Z. K., Perez, K. A., Stark, S. M., & Frithsen, A. (2020). Predicted and remembered emotion: Tomorrow's vividness trumps yesterday's accuracy. *Memory (Hove, England)*, *28*(1), 128–140.
Loewenstein, G. (2005a). Projection bias in medical decision making. *Medical Decision Making*, *25*(1), 96–105.
Loewenstein, G. (2005b). Hot-cold empathy gaps in medical decision making. *Health Psychology*, *24*, S49–S56.
Loewenstein, G., & Adler, D. (1995). A bias in the prediction of tastes. *The Economic Journal*, *105*, 929–937.
Loewenstein, G., Prelec, D., & Shatto, C. (1998). *Hot/cold intrapersonal empathy gaps and the underprediction of curiosity*. Unpublished manuscript, Carnegie Mellon University.
Manne, K. (2017). *Down girl: The logic of misogyny*. Oxford University Press.
McRae, K., & Gross, J. J. (2020). Emotion regulation. *Emotion*, *20*(1), 1–9. https://doi.org/10.1037/emo0000703.
McSweeney, M. M. (2023). Metaphysics as essentially imaginative and aiming at understanding. *American Philosophical Quarterly*, *60*(1), 83–97.
Miracchi, L. (2015). Competence to know. *Philosophical Studies*, *172*(1), 29–56.
Montero, B. G. (2020). What experience doesn't teach: Pain amnesia and a new paradigm for memory research. *Journal of Consciousness Studies*, *27*(11–12), 102–125.
Morley, S. (1993). Vivid memory for everyday pains. *Pain*, *55*(1), 55–62. https://doi.org/10.1016/0304-3959(93)90184-Q.
Moss, S. (2022). Knowledge and legal proof. In Tamar Szabó Gendler, John Hawthorne, and Julianne Chung (eds), *Oxford Studies in Epistemology* 7, 176–213.
Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences*, *364*(1521), 1273–1280. https://doi.org/10.1098/rstb.2008.0314.
Nguyen, C. T. (2021). The seductions of clarity. *Royal Institute of Philosophy Supplements*, *89*, 227–255.
Nordgren, L. F., van der Pligt, J., & Van Harreveld, F. (2006). Visceral drives in retrospect: Explanations about the inaccessible past. *Psychological Science*, *17*(7), 635–640.

Nordgren, L. F., van der Pligt, J., & van Harreveld, F. (2007). Evaluating eve: Visceral states influence the evaluation of impulsive behavior. *Journal of Personality and Social Psychology*, *93*(1), 75.
Nordgren, L. F., Banas, K., & MacDonald, G. (2011a). Empathy gaps for social pain: Why people underestimate the pain of social suffering. *Journal of Personality and Social Psychology*, *100*(1), 120.
Nordgren, L., McDonnell, M., & Loewenstein, G. (2011b). What constitutes torture? Psychological impediments to an objective evaluation of interrogation tactics. *Psychological Science*, *22*, 689–694.
Paul, L. A. (2014). *Transformative experience*. Oxford University Press.
Paul, L. A. (2015a). Précis of *transformative experience*. *Philosophy and Phenomenological Research*, *91*(3), 760–765.
Paul, L. A. (2015b). *Transformative experience*: Replies to Pettigrew, Barnes and Campbell. *Philosophy and Phenomenological Research*, *91*(3), 794–813.
Paul, L. A. (2017a). *De se* preferences and empathy for future selves. *Philosophical Perspectives*, *31*(1).
Paul, L. A. (2017b). Phenomenal feel as process. *Philosophical Issues*, *27*(1), 204–222.
Paul, L. A. (2020). The first time as tragedy, the second as farce. *Journal of Consciousness Studies*, *27*(11–12), 145–153.
Paul, L. A. (2021). The paradox of empathy. *Episteme*, *18*(3), 347–366.
Pettigrew, R. (2015). Transformative experience and decision theory. *Philosophy and Phenomenological Research*, *91*(3), 766–774.
Pickard, H., & Pearce, S. (2013). *Addiction in context* (pp. 165–189). OUP.
Poggiolini, C. (2019). High self-efficacy regarding smoking cessation may weaken the intention to quit smoking. *Cogent Psychology*, *6*, Article 1574096. https://doi.org/10.1080/23311908.2019.1574096.
Pritchard, D. (2009). Safety-based epistemology. *Journal of Philosophical Research*, *34*, 33–45.
Pronin, E., & Ross, L. (2006). Temporal differences in trait self-ascription: When the self is seen as an other. *Journal of Personality and Social Psychology*, *90*(2), 197–209. https://doi.org/10.1037/0022-3514.90.2.197.
Read, D., & Loewenstein, G. (1999). Enduring pain for money: Decisions based on the perception and memory of pain. *J Behav Decis Making*, *12*, 1–17.
Roush, S. (2016). Simulation and understanding other minds. *Philosophical Issues*, *26*(1), 351–373.
Roush, S. (2017). The difference between knowledge and understanding. In R. Borges, de C. Almeida, & P. Klein (Eds.), *Explaining knowledge: New Essays on the Gettier Problem* (pp. 384–407). Oxford University Press.
Ryan, S. (2003). Doxastic compatibilism and the ethics of belief. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *114*(1/2), 47–79.
Schellenberg, S. (2018). *The unity of perception: Content, consciousness, evidence*. Oxford University Press.
Semler, J., & Henne, P. (2019). Recent experimental work on ought implies can. *Philosophy Compass*, *14*(9), e12619.
Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives*, *13*, 137–149.
Spaulding, S. (2018). Mindreading beyond belief: A more comprehensive conception of how we understand others. *Philosophy Compass*, *13*(11), e12526.
Spaulding, S. (2020). What is mindreading? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(3), e1523.
Steinmetz, J., Tausen, B. M., & Risen, J. L. (2018). Mental simulation of visceral states affects preferences and behavior. *Personality and Social Psychology Bulletin*, *44*(3), 406–417.
Sullivan, E. (2018). Understanding: Not know-how. *Philosophical Studies*, *175*(1), 221–240.
Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, *29*, 1159–1168.
Van Boven, L., Dunning, D., & Loewenstein, G. (2000). Egocentric empathy gaps between owners and buyers: Misperceptions of the endowment effect. *Journal of Personality and Social Psychology*, *79*, 66–76.
Van Boven, L., Loewenstein, G., Welch, N., & Dunning, D. (2012). The illusion of courage in self-predictions: Mispredicting one's own behavior in embarrassing situations. *Journal of Behavioral Decision Making*, *25*, 1–12.
Van Boven, L., Loewenstein, G., Dunning, D., & Nordgren, L. F. (2013). Changing places: A dual judgment model of empathy gaps in emotional perspective taking. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (pp. 117–171).
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.