Penultimate Draft, forthcoming in Oxford Studies in Philosophy of Mind, edited by Uriah Kriegel. Please cite published version.

Idealism and the Interface Theory[1]

Geoffrey Lee

Metaphysical idealists in the Berkeleyian tradition advocate for the global metaphysical priority of the mental over the physical. This kind of idealism is widely regarded as highly revisionary and difficult to develop in a coherent and plausible way, although occasionally theorists take on the challenge[2]. Kantian Metaphysical Idealists, on the other hand, hold that there is a mind-independent "noumenal" world, but that it is unknowable to observers such as ourselves. Moreover, what is accessible is the "phenomenal world", a kind of mental construction that does not reflect the true nature of the noumenal world, but rather reflects the nature of the perceiving mind itself (for Kant, the a priori organizational constraints on perceptual experience).

Views with a Kantian-idealist flavor often pop up in contemporary philosophy[3]. For example, recently some analytic metaphysicians (Langton (1998), Lewis (2001)), have developed a Kantian view on which it is the intrinsic nature of fundamental physical properties that constitutes the inaccessible noumena. And views of this kind have proliferated among so called Russellian Monists (including Panpsychists), who hold that the hidden intrinsic nature of the fundamental world could be the ingredient we are missing in understanding how conscious experience could be grounded in physical events[4]. On these views, we can access the abstract structure of the physical world – both at the fundamental level and macroscopically – through mathematical modelling. But such abstract structural descriptions do not tell us which properties and relations actually fill out this structure[5].

Here I'm also interested in a kind of Kantian view, but one that instead focuses precisely on the kind of structural knowledge that is *not* challenged on Russellian Monist views. On the *Structural Idealist* view, it is the structure of the noumenal world that is inaccessible to us. For the purposes of the present paper, I will give this a *perceptual* reading in terms of two theses. First, according to *Structural Opacity* (as opposed to *Structural Transparency*), the structure of

[1] Many thanks to Uriah Kriegel, Robert Prentner, Jonathan Simon and Galen Strawson and for helpful comments on an earlier draft, and to participants in the 2023 workshop at Rice university for helpful discussion and feedback.

[2] see Pelzcar (2015) for a recent version and Lee (2016) for a response. For a more detailed taxonomy of idealist views see Chalmers (2017).

[3] In addition to Langton and Lewis style "Kantian Humility", views that are sceptical of stronger forms of 'metaphysical realism' (whatever that amounts to), such as Putnam's internal realism [1981, 1983], and the pragmatist views of thinkers like Dewey, Carnap, Rorty, or Brandom can be seen as at least having a Kantian-idealist flavor to them.

[4] Chalmers (2015), Goff (2017) Morch (2014) Roelofs (2019), Strawson (2009) are some recent examples.

[5] To put it another way, two universes could be structurally identical, and appear to function the same way to observers mathematically modelling these universes, but differ in the properties and relations that actually fill out the structural models. Observers have no way of distinguishing these situations, beyond conceiving of them indexically as the one that *actually obtains* (the exception being the qualitative character of their own minds, of which they do have non-structural knowledge).

the external world as presented in perceptual experience does not reflect its objective physical structure. Perception is not a transparent window onto the world, but more like a highly distorting filter that changes the appearance of the physical world beyond all recognition. Second, according to a thesis I'll call *Irreversibility,* we lack the epistemic means to undo the distorting filter and recover the true structure of the world. Our attempts to "get behind appearances" and model the structure of the physical world at best enable us to understand features of the physical to phenomenal transforming function (for example, certain kinds of invariances in it), but not to actually reverse it and know what the world is really like (structurally) in itself. We are forever "stuck inside the headset".

A version of Structural Opacity has been defended recently under the guise "The Interface Theory", by Hoffman, Singh and Prakash (henceforth "HSP") (Hoffman et al (2015a,b), Hoffman (2019), Prakash et al. (2020), Prakash et al (2021)). Theirs is a Darwinian spin on the idea. They offer an evolutionary debunking argument against Structural Transparency. Our perceptual systems are not tuned (or *primarily* tuned) to present to us the objective physical structure of the environment around us, but rather, qua results of natural selection, are tuned to present the *fitness payoff structure* of the environment. Moreover, there is no happy alignment between these structures. They are *uncorrelated,* and this means that in an important sense, perception is not a "veridical" presentation of the environment, but rather a pragmatically motivated *interface*, akin to a desktop environment on a computer (a metaphor familiar from Dennett's work on consciousness (e.g. Dennett (1993)). Moreover, the physical-to-phenomenal function is *non-monotonic* (i.e. not order-preserving) in a way that gives us a version of Irreversibility. Evolution has not given us a perspective on the world that allows knowledge of its noumenal structure, even in principle.

The following quote makes very clear the Kantian flavor of the view:

> "When we compare psychophysical measurements of shape to spatial measurements in the physical (or a simulated) environment, we are simply evaluating the degree of coherence between two different levels of description within our perceptual interface. This can indeed be an informative evaluation. But we are not somehow getting outside of our own interface in order to compare perceptual experience with objective reality." (HSP (2015b) p. 1573)

It's true that Hoffman himself (2019 ch.10) has developed a decidedly non-Kantian version of the interface theory, supplementing it with a view of noumenal world he calls "conscious realism". On that view conscious subjects are the fundamental entities whose properties and relations constitute the objective world (the view is similar in spirit to Millian phenomenalism). But the interface theory (as I will read it) is not committed to conscious realism, and here I do not engage with that aspect of Hoffman's view[6]. For example, it is consistent with the interface theory that the realizer of unknowable noumenal structure is wholly non-mental.

My goal in this paper is to explain and evaluate the interface theory *qua structural idealist view*. In section 2, I argue that it is clearly intended as a structural idealist view, and that this is

[6] Since the interface theory (at least as I understand it) is committed to unknowable noumenal structure that grounds the phenomenal world, it's not even totally clear how conscious realism, which grounds everything in the (knowable?) states of conscious agents, is *compatible* with the interface theory.

what makes it interesting – a point that I think has been lost in translation in the (limited) critical reaction it has received from philosophers. In section 3, I unpack and partially evaluate HSP's case for the view. In section 4, I further develop the view and explain how I see the shape of the debate with realist opponents. My overall goal is to make a tentative case for structural transparency: our perceptual experience does provide access to the objective physical structure of the environment. Despite realism itself involving some ambitious and questionable commitments, the objections to structural idealism and the weakness of the positive case for it make realism the more attractive position.

2. Three notions of Veridicality

The interface theorists make the claim that *perception is typically non-veridical* the centerpiece of their view. They sell this as an iconoclastic debunking of the standard views of both philosophers and cognitive scientists, liberally quoting figures in both camps they see themselves as at odds with, even going so far as to claim opposition to the correspondence theory of truth (!) (e.g. Singh et al (2020) p.3). Although I think their view is interesting, I see this as a mistake that invites misunderstanding. What they should have said is : "There is an important notion of veridicality that has been missed by recent cognitive scientists and philosophers. Even if perception is typically veridical in the sense that recent theorists have claimed, it could be non-veridical in an important (and disturbing) further sense; moreover it *is* non-veridical in this further sense".

In fact I think there are three important notions of veridicality that are in play here : the content-based notion, the relational notion, and the structural-transparency notion.

On the content-based notion, perceptual experience has a content: a proposition that specifies how the environment is arranged around the subject. Perception is veridical provided the proposition is true. For example, my experience might represent that an orange sphere is in front of me, and is veridical provided an orange sphere is in fact in front of me. This is the standard notion of veridicality in the literature, although for reasons I'll get to, it doesn't really figure in the interface theory.

To say that experience is veridical in the *relational* sense presupposes a relational view of the phenomenal character of experience. On that view, having an experience with a certain phenomenal character consists in standing in a special sensory relation (e.g. awareness or acquaintance) to the *objects* of experience (facts/events/objects/properties/propositions)[7]. Perception is *relationally-veridical* provided the phenomenal character of experience relates the subject to an actually-obtaining mind-independent arrangement of entities in their environment. The classic version of a relational view is naïve realism, where the character of experience is constituted by acquaintance with mind-independent items in the physical environment around the subject: experience is something like a "transparent window" onto the external world. A sense-datum view would also be a kind of relational view, but it doesn't give

[7] On some relational views, these are not "objects of experience" in the ordinary sense, such as tables and trees, and the relevant sensory relation is not "awareness" or "perception" in any ordinary sense. For example, on representationist views, the object is a proposition or property, and the sensory relation is a special representational relation (see e.g. Pautz (2020) ch. 3).

us veridical experience in the relevant sense, because the relata are mind-dependent sense-data which merely stand in for external items.

Although HSP don't give the kind of theoretical formulation that would please philosophers, it is fairly clear they do not think experience has relational veridicality. For them, the mind-independent world is (in some sense) hidden from us behind a veil of appearance[8]. However, I don't think this is the big fish they have to fry, so I set it aside here.

Their big fish is the idea that experience is not veridical in a *structure-preserving* sense. The intuition that matters here, I think, is that the manifest image does not distort or scramble the fundamental physical image beyond all recognition. In some sense, the manifest features we perceive are reasonably "natural" in the way they derive from the fundamental, and as a result the project of reverse engineering the world and trying to infer what fundamental image looks like, is not hopelessly misguided. A loose comparison might be with a photograph of a detailed scene. The photograph will inevitably have limited resolution, and as a result much detail will be lost. Still, some of the broad overall features of the original scene are straightforwardly available in the photo, in a way that they wouldn't be if the camera totally scrambled the image. Similarly, although perception is not a direct window onto the fundamental physical world, we think that it gives us a summary that is not completely unrelated to the fundamental layout, but rather tells us some interesting broad features of how fundamental reality around us is arranged. This means that "filling in the details" theoretically is not a misguided project.

This intuitive formulation stands in need of much clarification, and I will say more later about how I think we should develop it. But I hope this is enough to make it intuitive that there might be an important notion of veridicality in this ballpark that we should pay attention to. This is the primary sense in which I am in complete agreement with the interface theorists (the difference is in our attitude to whether experience *is* veridical (i.e. structurally transparent)).

I will also immediately note that "representation through resemblance" has, of course, always been a central idea in approaches to representation[9]; so there is nothing new in suggesting an approach in that category. What I think is interesting is a specific *version* of that idea ("veridicality as non-scrambling"), which I see embodied in HSP's approach. As I will explain below, notions of representation-by-resemblance in play in current debates in the philosophy of cognitive science typically do *not* capture the kind of "non-scrambling" I am interested in here.

The way the idea is developed by the interface theorists themselves requires a little set up. Following the classic psychophysics tradition, they believe in what I will call the

---

[8] In an attempt to clarify their position, McLaughlin and Green (2015) understandably read them as holding a sense-datum theory. In response, they vigorously deny this and seem to adopt a representationalist view on which the phenomenal character of experienced is to be understood intentionally – it is constituted by the sensory representation of a proposition concerning the layout of the external world – but on which this proposition is typically *false* – it represents objects and property instantiations that *don't exist* (see HSP 2015b pp 1569-1572). That discussion suggests their view is similar to the Chalmers/Pautz "Edenic" view of perceptual experience (Chalmers (2006), Pautz (2020)) .

[9] For example, the idea plays an important role in early modern debates about mental representation, for example in the debate about realism between Locke and Berkeley. Shepard's notion of second-order isomorphism is foundational to modern accounts of representation-by-resemblance – see Shepard and Chipman (1970) (Shea (2019) and Neander (2017) are examples of recent accounts along these lines).

psychometric mapping thesis : for a given organism, there is a relevant[10], well-defined psychometric mapping from objective physical structure to phenomenal structure (it could be probabilistic rather than deterministic). So for example, a component of this mapping function could be a map from spatial distances into an internal experiential quality space that we would intuitively describe as "experiences of distance (or length)". The map tells us (something like) the typical experience (or experience*s* if the map is non-deterministic) that the subject would have in response to an item with a certain length impinging on their retina.

Now, as this example may immediately makes vivid, our experiences of features like spatial distance can be quite context sensitive (even under optimal, ecologically valid conditions etc), which could make trouble for an overly simplistic mapping thesis that does not contain contextual parameters as variables. I ignore this important complication for now, however (we can perhaps imagine these variables to be absorbed by an appropriate choice of what the physical stimulus parameter is). One might also wonder: what determines the choice of physical stimulus parameters? Are there specific "optimal" conditions where the functioning of perceptual systems is particularly relevant to the determination of the function? I briefly return to these issues below.

Assuming we have these functions in hand, HSP (2015, p.1482) define a *perceptual strategy* as the psychometric mapping that an organism in fact uses. They then distinguish various different kinds of strategy. For current purposes, two are particularly relevant: a *hybrid realist* strategy and an *interface* strategy. The issue here is the extent to which the psychometric map is structure-preserving (if it is, they call it "realist" or "critical realist"[11]). HSP typically understand structure-preservation specifically in terms of whether the function is *monotonic* (i.e. order-preserving), although I think it should be an open question what kinds of structure-preservation are theoretically important here (more on this below).

The "hybrid realist" view is that perception is realist with respect to "primary" qualities like duration, spatial distance, and mass, and is non-realist (i.e. scrambling) with respect to "secondary" qualities like color or smell (I think it's fair to say that this is both a non-standard but also theoretically attractive and intuitive way to draw this distinction). Whereas on an "interface strategy" the mapping is non-monotonic with respect to *every* stimulus parameter: i.e. distance, duration etc. are not really different from color and smell after all! The *Interface Theory* is the view that we humans use an interface strategy[12].

---

[10] By which I mean: the function captures a theoretically important relationship between experience and reality – the kind of relationship we have in mind when we talk about what the experience is *of*, what it *represents*, or what it has as a *content*. How exactly to understand this relationship is of course a central problem in philosophy of perception.

[11] They distinguish two kinds of realism, *naïve realism* and *critical realism*. They are hard to interpret here, but on my reading critical realism is intended as an umbrella category that includes naïve realist views in the philosopher's "phenomenology as acquaintance with the world" sense, but also includes views like sense-datum views, qualia views and representationalist views. That is, the critical realist only requires a structure-preserving psycho-physical mapping, and does not further require that phenomenal properties are individuated in terms of relations to mind-independent stimulus properties, whereas the "naïve realist" does include this requirement (in addition to the structure-preservation requirement).

[12] Cf Berkeley's argument in his *Three Dialogues* for assimilating allegedly primary qualities like shape and size with mind-dependent secondary qualities like taste and color (Berkeley (1713/1979)).

Some clarifications. First, to preserve order (monotonicity) is to preserve structure only in a fairly weak sense. In particular, the class of order-preserving mappings between metric spaces includes maps that distort distances in strange ways, warping the world in a way that (intuitively) gives us a kind of non-veridicality which we might want a notion of "structural transparency" to capture. But this is no real objection. Considering monotonicity is motivated in part by a desire not to make strong assumptions about the structures of physical and phenomenal variables (e.g. that they have a well-defined metric), but also (I assume) because monotonicity is typically *necessary* for structure-preservation in other senses we might be interested in (e.g. preserving metric structure). Attacking the monotonicity of perception is therefore strong dialectically for HSP, and is consistent with considering stronger kinds of structure-preservation in other contexts.

Second, we can distinguish between strongly and weakly monotonic functions. The difference is that a weakly monotonic function need not be strictly increasing or decreasing, but can map points in an interval onto the same point. HSP are arguing that our perceptual functions are not even weakly monotonic. One way in which this is significant is that dimensionality reducing mappings (e.g. a 3D to 2D linear projection), which are often characteristic of human perception, are only weakly monotonic : but that is consistent with them being structure-preserving in the relevant sense.

Third, in real life when we are presented with some kinds of non-monotonic maps (e.g. a jigsaw puzzle), we *can* figure out how to undo them. But that's because we have priors about the true structure of things. For example, imagine taking a picture of a face and cutting it into horizontal strips and scrambling their order – most of us can solve this kind of puzzle. But imagine you had only ever seen faces scrambled in a certain systematic way (or more saliently, the *whole world* scrambled in a certain systematic way)!! Then (perhaps) you would not have priors that would enable you to "solve" for the true structure. In fact, the world might not even seem incongruous or surprising or in need of solving or rearranging in the way that a jigsaw puzzle does. So the mere fact that "the world looks normal" is not in itself evidence against perception being non-monotonic (although there may be more sophisticated objections in this ballpark that are telling (more later)).

Fourth, how we should understand the "structure-preservation" in structural transparency of course depends on the theoretical role we want it to play. From my point of view (which I believe is similar to HSP's on this point), there are (at least) three important roles to consider. First, I think I think veridical experience can reasonably be understood as an *epistemic end-in-itself*[13], and that structural-transparency gives one reasonable gloss on this. We want experience to "tell us what the world is like", and if structural transparency fails, then there is one way in which it does not live up to this epistemic aspiration[14]. Second, veridical experience (understood in any of the three ways discussed here) potentially helps explain the functioning and success of the organism; that is, the notion has a (potential) high-level *causal-explanatory role*. When it comes to structural transparency, we can ask whether veridical

[13] For example, creating photos that allow us to view previously unseen objects in deep space (e.g. the 2022 images of a black hole from the event horizon telescope), enables what feels to us like an epistemic achievement with respect to these objects, even if we already knew they exist and what features they have.

[14] Although it may live up to this in other ways, e.g. by having an accurate representational content (see below).

experience (in that sense) might help explain how an organism succeeds in surviving and reproducing, or otherwise functions psychologically and behaviorally (wouldn't scrambled perception be useless and lead the organism to quickly die?). Third, there is an *epistemic-theoretical role* for the notion, with respect to the scientific project of *downwards inference:* inferring the deeper structure of the world from the manifest appearances. Is perception structure-preserving in a sense that would make the project of downwards inference epistemically tractable?

HSP's Darwinian debunking argument (discussed below in section 3) can be seen as an argument that the causal-explanatory role of experience (surprisingly) supports the view that experience is *not* structurally transparent in the ways that matter for the two epistemic roles just mentioned. In the case of downwards inference / reversibility (which I will be particularly focused on), they might therefore be read as saying that, in general, monotonic perception is necessary for the underlying structure of the world to be knowable. But examples like jigsaws already show that this isn't quite right. We might also mention cases like color and smell which may violate monotonicity in their relationship to underlying physical properties, but in a way that we can figure out, because it is merely local. What HSP have in mind is that a *global* kind of non-monotonicity might make for unknowable noumenal structure. But how exactly do we characterize this epistemically problematic kind of scrambling? On reflection, it's a tricky substantive problem to figure out what kinds of structure-preservation we implicitly assume in making downwards inferences. I'll return to this issue in section 4, discussing monotonicity for now as a place-holder for the kinds of "non-scrambling" we might ultimately be interested in if downwards inference is our focus.

Fifth, the notion of structural transparency presupposes we have some grip on the idea of the "objective structure" of the environment – an idea that of course can be challenged. As I discuss in section 4, a crucial idea here is that the world has *fundamental* physical (or non-physical) structure. I think of this in terms of a commitment to fundamentality or naturalness as a basic piece of metaphysical ideology (Lewis 1983, Sider 2013), although the discussion here is probably compatible with other approaches. Further, in section 4 I discuss one way of developing the idea of "real patterns" or "objective structure" in the grounded non-fundamental world. In the end (section 4), I think perceptual transparency is best understood as the view that we perceive such real patterns.

Sixth, I think it's crucial to understand that non-veridicality in the structural-transparency sense is perfectly compatible with veridicality in the standard content-theoretic sense[15]. I think HSP are rather illicitly side-stepping the debates about experiential content by granting themselves a *relevant* psychometric function. In particular, we might think that part of what makes a function "relevant" is that the stimulus parameter is *represented* (in some theoretically relevant sense) by the experience it maps to. And I suspect that the spadework that would go into spelling out what determines the psychometric function as relevant would be fairly similar to the spadework that goes into theorizing the notion of experiential content or representation. Regardless, the important point here is that non-monotonic perceptual mappings are prima facie consistent with experience veridically representing real features of

[15] It's also consistent with veridicality in the *relational* sense, assuming I can be perceptually related to macroscopic properties that are grounded in a structurally opaque way.

the environment. For example, it is common-place to hold that color experiences represent *spectral reflectance profiles*, even if the psychometric color map is massively dimensionality reducing and non-monotonic. On this framework, a color experience is veridical provided the surface has the kind of reflectance profile the experience represents, where e.g. this might be cashed out as the kind of profile the experience is designed (by evolution or learning) to detect. So we have veridicality *plus* non-monotonicity.

It's also important to note here that we can take the form of the psychometric mapping function, and use it to define a mind-independent high-level property of surfaces, such that we can think of *this* property as the stimulus property, and think of the psychometric map as completely structure-preserving. For example, if experienced brightness is a power function of luminance, we can apply this function to luminance to get a mind-independent quantity defined in terms of luminance ("physical brightness"). Call this the *induced* stimulus property. Even if perception scrambles the world, we can legitimately think of it as veridically presenting the subject with these induced worldly features (although typically it's more theoretically illuminating to take the underlying physical parameter like luminance to be what's represented). Again, this is perfectly consistent with it being *non*-veridical in the important structural-transparency sense.

When theorists in philosophy and cognitive science say that unless an organism's perceptual experience is typically veridical they would quickly die, they are talking about this content-theoretic sense of veridicality. For example, if the function of experiencing the color red is to distinguish fruit from foliage, then if an organisms' perceptual system started misfiring and randomly assigning red to non-fruit parts of trees instead, then the organism might quickly die because it could not find fruit to eat. All of this is perfectly compatible with color experience being non-monotonic. (Cohen (2015) makes a similar point in his critique of HSP, but I think he fails to pick up on the fact that HSP are really interested in veridicality in a different sense).

Finally, the difference with notions of structural / imagistic / iconic / analog representation that have been theorized in the recent literature is this[16]. Those notions typically allow that the mapping from lower-level physical structure onto representational structure could be fairly complex and gruesome (think of how color experience can be a structural representation of surface reflectance). One way to put this is that on these views, we could happily construe the external structure being represented as the *induced* structure, which can be veridically represented even under "scrambling" conditions. Of course, if the function of internal structural representations is to mirror external structure inside the organism in a way that is computationally and behaviorally useful, we will want to know *why* a structure that is only gruesomely related to physical structure *is* useful to structurally represent. But that is conceivably answerable (e.g. it might be fitness payoff structure (see below)!).

In section 4 I will discuss in more detail the definition and theoretical importance of structural transparency. First, I want to briefly consider HSP's arguments for the interface theory.

3. The case for the Interface Theory (aka "The Case Against Reality")

[16] See again Shea (2019), Neander (2017); also Beck (2018) on analog representation.

On my reading, HSP's argument is structured in the following way (I'll call this "the master argument"):

(1) **Fitness perception thesis** : Perception presents us (only) with the structure of fitness payoffs.

(2) **Payoff distribution thesis** : Fitness payoff structure is non-monotonically related to physical structure

Therefore:

(3) **Perceptual Non-monotonicity** : Our perceptual strategy is globally non-monotonic

Therefore:

(4) **Non-veridicality :** Perception is non-veridical

As we will see, although HSP sell their argument as a Darwinian debunking of perceptual veridicality, one of their strategies for motivating the premises of the argument ("the indifference argument") gives us an argument that in fact has nothing to do with Darwinian considerations.

On my reading, the move from (3) to (4) is trivial once we get clear on what "veridical" is supposed to mean (I'm assuming for now that monotonicity is an adequate way to theorize "non-scrambling"). So the main issue is the motivation of the premises. Let's consider them in turn.

3.1 The Fitness Perception Thesis

A fitness payoff is the expected change in fitness from performing an action. So we might also call the thesis "strong perceptual pragmatism" : we only perceive the expected payoffs/costs of possible actions we could perform. This is quite prima facie counterintuitive, and can be usefully contrasted with a realist view on which we perceive action-independent features of how the world is arranged, which, *together with an independent sense of goals and priorities*, we use to *compute* an optimal course of action.

That said, there are cases such as *tastiness* and *attractiveness* where something like perceptual pragmatism is fairly plausible. For example, it might be that the function of tastiness is to make a recommendation to eat or not eat the food, so there is (at least) an immediate *connection* between the perceived property and a pragmatic implication. So we can read the fitness perception thesis as saying that all perceived features are like tastiness and attractiveness in this way. When we think about properties like spatial distance and duration this might seem a surprising view (what in general is the expected cost of an item (*any* item) having a certain size, or being a certain distance away?), but perhaps there is a good argument for it.

HSP's argument for the view is that in toy evolutionary models, perceptual strategies that are monotonic in fitness outcompete those that are not.

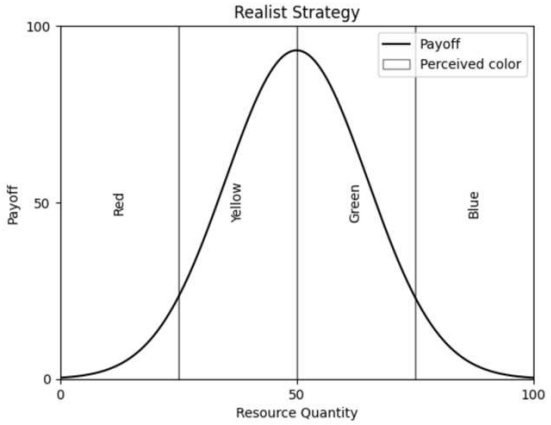

Fig.1 (based on HSP (2015a fig 2). Realist psychometric function mapping resource quantities to perceived colors, with payoff also shown. This discrete mapping is (weakly) order preserving (note : colors are assumed to be ordered in hue circle in standard way).

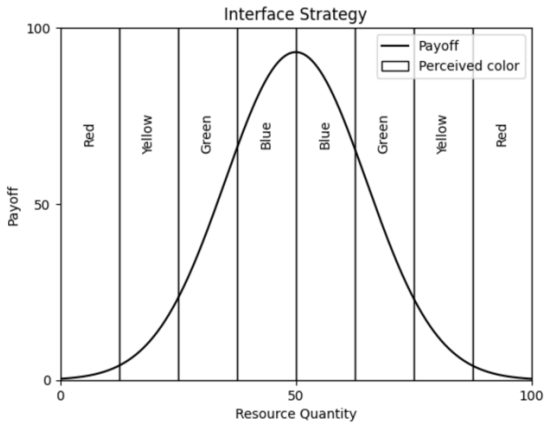

Fig.2 (based on HSP (2015a fig 3). Interface psychometric function, mapping resource quantity to perceived colors, with payoff also shown. This discrete mapping is not order-preserving.

In these toy models, there is an environment consisting of a grid with resources distributed in different quantities across grid squares. Fitness payoff is taken to be a normal (and therefore non-monotonic) distribution of resource quantity – both taking too little and taking too much is non-optimal. Creatures grab resources from squares based on their perceptual state. In one kind of model, populations with different perceptual strategies engage in an evolutionary competition. The result is that an interface strategy (i.e. a non-monotonic psychometric function) for resource quantity (fig. 2) dominates a critical realist strategy (fig. 1). That is, it is advantageous to perceptually group together resource quantities with similar payoffs, even though they do not form a contiguous grouping (consider e.g. the yellow grouping in fig.2). In another kind of model, the strategy itself is allowed to evolve through combining strategies of mates in a quasi-genetic way. This genetic algorithm ends up optimizing towards an interface strategy for resource quantity[17].

Does this support perceptual pragmatism for *all* perceived properties? Note that the argument just assumes the payoff distribution thesis – that payoffs are non-monotonic with respect to physical stimulus properties. If they were, e.g. linearly related, then there would be no competition between payoff and physical quantity to be perceptually represented. It may seem unsurprising that if we assume payoff non-monotonicity, the interface strategy ends up dominating: we'll come back to this momentarily. But actually there is another problem here, which is that real systems are subject to computational constraints that (arguably) aren't

[17] Prakash et al. (2021) provide a proof that "fitness beats truth" in a class of such toy models.

adequately modelled here. The optimal course of action for an organism is a very complicated function of stimulus parameters, and to even approximate it requires a complex multi-stage computational process. Perceptual transducers typically only modestly and monotonically transform a perceptual parameter – e.g. representing a power function. This inputs into a complex multi-layered computational system that outputs a motor instruction only after many layers of processing, layers which may represent stimulus features at increasing levels of abstraction from the transduced properties. This means that even if a complex non-monotonic transformation of the stimulus property occurs to determine action, there may also be, in earlier layers, properties represented which are fairly simple monotonic functions of stimulus quantities. And these earlier layers might correspond to conscious experiences. For example, it could be that a creature experiences resource quantity in a monotonic way, and then this is transformed non-monotonically *post-perceptually*, to determine optimal course of action. So critical realist vs interface could be a false contrast: we might need both.

Relatedly, there may be stimulus features that are useful to (monotonically) represent as an input to a broad variety of these action-oriented computations, across different contexts. Surely it is plausible to speculate that spatial structure is like this. For example, it is useful both for computing the energy cost of movement and for object identification. Moreover, such flexible-use representations are plausibly evolutionarily accessible (they don't require a designer with foresight), because the broad use could be an exaptation from a more simple use. For example, spatial distance perception might evolve first because it is useful for simple computations of cost of motion; but then it is exapted for a host of other useful purposes.

A natural speculation would then be that if we study toy models that are subject to more realistic computational constraints and where a more realistic range of physical parameters is relevant to the organism's fitness, we would get results that instead support the hybrid realist view. I would note in this regard that (ironically) the perceptual strategies that dominate in HSP's models are actually, contrary to advertising, only interface with respect to resource quantity, and are *overall* critical realist, because the organisms "veridically" (i.e. structurally veridically) perceive layout properties of their environment such as where environmental boundaries are. Future work that investigates the properties of these models might be illuminating[18].

So the argument for the fitness perception thesis is questionable. Let's turn to the payoff distribution thesis: non-monotonicity of payoff with respect to physical parameters.

3.2 The Payoff Distribution Thesis

Again, this is the thesis that fitness payoff structure is non-monotonically related to underlying physical structure (however that is understood). HSP have two arguments for this thesis, an empirical argument, and a philosophical argument which I will call the "indifference argument".

[18] In this direction, some preliminary joint work with Amalie Trewartha suggested that if we include these layout properties as features whose perceptual strategy can vary between organisms, a veridical *layout* perception tends to dominate, even if an interface strategy dominates for resource quantity.

The empirical argument is that organisms are homeostatic systems that are trying to maintain parameters in their internal state (e.g. temperature) within a livable range. So fitness functions vis a vis physical parameters will often be normal distributions (and therefore non-monotonic) that represent this "not too little, not too much" mode of interaction with the physical world (HSP 2015 p. 1486).

One problem here is that there seem to be obvious counterexamples. Again, spatial distance is a good case. The energy cost of motion along a path through physical space is a monotonic function of distance along the path. Cost of motion is obviously very useful to compute, and so it's not surprising that organisms have this capacity (admittedly overall computation of optimal path in an environment with hills and obstacles etc. is rather more complicated, but cost comparisons along a single path could be a component of that computation, and could be evolutionarily more ancient). Similar points could be made about duration.

Now, rather than further developing this objection, I want to immediately note that there is an attractive escape route here for the interface theorist, which would also help with the earlier objections. They could jettison the fitness perception thesis, and instead focus on the consciously perceived features of the environment that are at least the immediate *inputs* to computations of fitness payoffs (or at least, to action choices). Call these *intermediate features*. Unlike strongly pragmatic features (like, say, tastiness), these features might lead to quite different behaviors combined with different utilities and background information – i.e. they might have the kind of pragmatic flexibility I just suggested is characteristic of spatial distance. Still, we might ask: given that these intermediate features are adapted to computation of optimal action choice, and optimal action choice is often only very indirectly and non-monotonically related to physical stimulus features, why think that what is in conscious perception hasn't already been transformed beyond recognition? Why think that anything like structural transparency obtains for intermediate features?

Call the view that perception of intermediate features is structurally opaque *weak perceptual pragmatism*. To my mind, it is a more compelling challenge to perceptual transparency, because it doesn't require the strong, and rather implausible commitments of the strong pragmatist discussed in the previous section. For example, although it is surprising as a view of duration or spatial distance, weak pragmatism at least avoids implausibly treating them as directly tied to particular kinds of action payoff. This makes it at least somewhat more viable to maintain that these are intermediate features for which structural transparency does not obtain, and which therefore are *not physical input parameters in the relevant sense*. Thus, even if it is true that cost of motion is a monotonic function of spatial distance, that wouldn't show that the payoff distribution thesis is false, because spatial distance itself could be non-monotonically related to the true physical structure of the world (more on this momentarily).

Still, there are two important objections to weak perceptual pragmatism that I want to consider. These will also helpfully motivate HSP's other argument for Payoff Non-monotonicity, the indifference argument.

First, one might object that the appeals to Darwinian evolution, homeostasis, etc. in the argument assume that we know that we are organisms of a certain kind with a certain history. But if perception is globally non-veridical, doesn't that call this into question, and therefore call into question some of the premises of the argument? Is the argument self-undermining?

Interestingly, this is a point where, in one way, HSP's position is actually rather stronger than they allow for. They think that we can know we are products of Darwinian selection even conditional on doubting the existence biological entities like animals and plants that perception presents us with (because perception is "non-veridical"). They also speculate that even if perception isn't reliable, other aspects of our cognition (e.g. mathematical reasoning and other kinds of abstract reasoning) could be reliable, allowing us to know that Darwinism is true (2015a p.1500), 2019 ch.4). This suggests a kind of transcendental argument for Darwinism: a priori, it is the best explanation for the existence of complex structure in *any* situation prior to investigation. So we don't need to make any substantial empirical claims about the world to believe that we are the products of Darwinian natural selection. But actually I don't think they need this kind of speculative defense. We noted earlier that perception being non-veridical in the *structural* sense is perfectly compatible with it being veridical in the *content-involving* sense. And it's the latter sense that it is relevant to whether our beliefs based on experience are *true*. So there's no reason why HSP can't hold that there really exist such things as populations of organisms, their prey and predators, environmental resources etc. etc., and that the best explanation of their features and distribution is that they were created by Darwinian natural selection. True enough, our entire scheme for describing the biological world might be only non-transparently related to physics (or whatever exists fundamentally), but that's no reason for saying it isn't a roughly *correct* description of a real mind-independent system.

So far so good : however, there's still trouble here. The problem is that by the interface theorist's own lights, empirical evidence can only ever tell us about the form of the mapping function from properties *at different levels of the interface*. For example, on their view, a property like resource quantity or spatial distance is an *output* of the scrambling function. So even if there's an empirical argument that the mapping from resource quantity to perceptual space is non-monotonic, that doesn't show that the map from the underlying physical space to the perceptual space *is* non-monotonic. It leaves it completely open what that is. I think this is a serious objection : it suggests that any empirical argument for the interface view really will be self-undermining! This motivates their more a priori argument : the indifference argument (see below).

Second, there is a point made by Bertrand Russell in "the Problems of Philosophy" (and noted by HSP) which might seem to support perceptual transparency.

> "If a regiment of men are marching along a road, the shape of the regiment will look different from different points of view, but the men will appear arranged in the same order from all points of view. Hence, we regard the order as true also in physical space, whereas the shape is only supposed to correspond to the physical space so far as is required for the preservation of the order." (2001 p.51)

In other words: perceptual invariance (across changes in the position of the perceiving subject) is evidence for structural transparency (e.g. perception is structurally transparent with respect to *order* of perceived objects).

Now, HSP have an interesting response to this objection, which exploits what they call the "invention of symmetry" theorem (2015 p.1498). What the theorem shows is that invariance of the psychometric function across symmetry translations like rotation and change

of position is *consistent* with the function being non-transparent. So there is no necessary implication from invariance to structural transparency. For example, we might suppose that the physical world just consists in a completely unstructured set of points!! One can consistently have a psychometric function on this unstructured set that gives a phenomenal world where, e.g. from various different viewing points, it looks like there is a group of ten squirrels standing in a line.

That's an important point, but it is limited in following way. It may be objected that even if non-transparency is *consistent* with perceptual invariance, in many cases (but not all), the *best explanation* for perceptual invariance is that we are perceiving structure that is objective physical structure. Presumably that is what we would say for Russell's regiment, for example. And presumably that is the reason that the hypothesis of a completely unstructured fundamental world is highly unattractive.

The interface theorist may respond as follows. Such an explanation is a *downwards inference* – an inference from manifest perceptual structure to the underlying physical structure. But the form of this relationship is exactly what is at stake in the debate between the realist and the interface theorist. And indeed it is surely true that if we are not allowed to assume any prior constraints on the form of the connection between manifest properties and underlying physical properties (call this the *grounding function*) then we can't legitimately rule out, e.g. an unstructured fundamental world. Now, this is precisely HSP's second argument against structural transparency (on my reading of them). They say : if we consider the space of all possible grounding functions, the functions that that give us perceptual transparency occupy a vanishingly small region of the space of all such functions. So by a principle of indifference, we should not assume that the kinds of principles which (presumably) are implicitly operative in ordinary downwards inferences (e.g. those performed by scientists) are revealing of objective structure. Rather, they are just ways of nicely systematizing invariances and connections in the psychometric function - we are forever stuck in the headset![19]

They will also wheel out the indifference argument in response to another obvious objection. We might say: haven't we learned that spatio-temporal structure and causal structure is transparently related to fundamental physical structure? True enough, these days we take seriously fundamental theories on which space and time are non-fundamental and emergent (e.g. Wuthrich (2019)). But still, the connecting function is not a non-monotonic scrambler; otherwise we could not have made the downwards inference to the physical deeper structure. In response, HSP can say that the assumption that such downwards inferences reveal objective fundamental structure is question-begging against the interface theorist. If we are initially indifferent among possible grounding functions, the chance of this is effectively zero.

Now of course, whether this is convincing depends on how much force the indifference argument has. The problems with it are both technical and philosophical. On the more technical (but still philosophical!) end, we should ask : what exactly is the space of grounding functions? What kind of mathematical objects are we talking about here, and how are they measured? What is the argument that structural maps are a measure zero subset? The way I will proceed is to grant (quite charitably) that we are operating with a reasonable answer to this question, on

[19] see HSP (2015) and Prakash et al (2020). I should also note here the resemblance to epistemic arguments against realist views of fundamentality/naturalness (see e.g. Cohen and Callender (2009)).

which by a natural mathematical measure, structural functions really are a vanishingly small subset[20]. My objection will be more philosophical: why should the realist concede from this that they should have low prior probability? This will be the subject of the final section of this paper, where I also develop in more detail the realist's view.

4. Realism vs the Interface Theory (aka Why Reality is REAL after all)

I now consider in more detail what non-structural grounding *is* and why rejecting it really is an interesting form of idealism.

Now, I suspect some will object that the interface theory is *not* actually saying anything very radical or interesting. Recall that the theory claims perception has evolved to track the *pay-off structure* of the environment, and that this is non-monotonically related to the physical structure of the environment. The objector claims that this simply amounts to the banal claim that we perceive features of the environment that are useful to know about given our adaptive needs, and that these are distinct from, and only indirectly related to whatever physical features the Interface theorists is including in W (= the input to the psychometric function). What are these? On one important reading, states of W are supposed to be *fundamental physical states of affairs* (this is certainly suggested by HSPs' discussion). So the interface theory can start sounding like the combination of the following two positions:

> **Non-fundamental Perception** : The environmental features that we perceive are not fundamental physical features
>
> **Non-structural grounding** : the mapping from fundamental physics to the high-level features that perceptual experience presents to us is not a structure-preserving mapping.

The problem with non-fundamental perception (considered alone) is it's not at all a surprising claim. Granted, it has been historically tempting to think that at least space and time as perceived by us might be fundamental features of the universe. And indeed HSP emphasize in support of their view that space-time may well turn out to be non-fundamental (note that this supports interpreting them as limiting W to fundamental states of affairs). But still, it's not a radical or surprising claim that we only perceive high-level features of the world, and not fundamental features. Pretty much everyone thinks of the project of fundamental physics as getting behind the appearances that constitute the "manifest world", and inferring the detailed physical structure that is behind these appearances. And certainly, when philosophers or cognitive scientists have claimed that evolution has given us perceptual systems that typically present the world in an "accurate" or "veridical" way, they did not mean to claim that it provides a direct window onto the fundamental physical world.

[20] Prakash et al (2020) considers four classes of structures which could be used in psycho-metric modelling and argues that only a vanishingly small subclass are homomorphisms. Of course, other kinds of structures could be relevant (as they are well aware), and one could question the underlying indifference principle.

Furthermore, if we allow non-fundamental states of affairs into W, this raises the question "which ones"? Among the high-level states of affairs will be precisely the states of affairs that we have evolved to perceive (whatever these are!)! If we include these in W, then perception *will* be "veridical" by HSP's own lights! Furthermore, if W is deliberately limited to some set of non-fundamental states of affairs that are *not* those we have evolved to perceive, but for which a psychometric function F still exists, why is it interesting to be told that the mapping from W states to the states we *are* evolved to perceive, is a somewhat indirect and non-structural one? (more on this momentarily).

Relatedly, it is reasonable to wonder why supplementing non-fundamental perception with non-structural grounding is such a bold move. It is commonly believed that the story one would have to tell to get from a fundamental physical description of the world to the kind of description that mentions that manifest phenomena we know and love, could involve a very complex series of abstractions and inferences. Why think that the mapping implicit in this story would be a nice, simple, structure-preserving one? And who exactly has claimed otherwise?

Although I think HSP are at fault for not addressing this kind of objection, I also think that there is a compelling answer here. Consider again the problem of downwards inference: what are the correct epistemic principles to use to infer the lower-level grounds of the observable manifest world? A good answer to this question needs to explain how we can rule out what I call *micro-sceptical scenarios*. These are scenarios where the fundamental world is nothing like what we have come to believe, but nonetheless grounds our manifest world through a counterintuitively complex and gruesome grounding function. For example, consider *Game of Thrones world*. This is a world which, given knowledge of its fundamental layout, we would intuitively describe as a world where something like the world of the Game of Thrones novels is playing out. Now, consider the hypothesis that this is in fact the world that we (locally) live in; furthermore the reason why we don't see any of the game of thrones participants is that we and all the objects we observe are *metaphysical junk*: our manifest world is related to the fundamental level in a complex and gruesome way. If informed of our existence, philosophers in the game of thrones world would think of us as fanciful and purely notional constructions, rather than solid concrete beings. Or for another example, consider *the dust world hypothesis* : the fundamental world is random swirling dust; nonetheless the manifest world exists in a way that is derived in complex gruesome way from the dust[21].

What is "metaphysical junk"? I am assuming here a plenitudinous view of objects and properties. On this view, there is an object for every function from possible worlds to sets of space-time points (it's actual and possible space-time trajectories). And there is a property or relation for every function from worlds to sets of objects (or sets of n-tuples of objects for n-place relations; for quantities we can consider functions from objects to real numbers). I assume that a subset of these objects and properties are privileged as fundamental – e.g. these could be individual space-time points, or particles, and the features ascribed to them by fundamental physics. The rest includes the objects we know and love like mountains and trees, but also so much else! Although we typically ignore most of the metaphysical junk, it at least exists. A well-known challenge is – what exactly is distinguishes the junk from the entities we know and love? Here the issue is : how do we know that we aren't junk?

[21] Chalmers also (2022) discusses this sceptical scenario, which he attributes to sci-fi writer Greg Egan.

In an important discussion which I consider in detail elsewhere (Lee (manuscript)), Shoemaker (1988) gives a method of constructing systems of junk objects ("ghosts") in such a way that the correlations across modal space between events in these junk systems mimic the correlations we find in real functional systems, and the objects are spatiotemporally like real objects[22]. This raises the question – why, if at all, do these ghosts not count as genuine functional systems? Admitting that they are would seem to be catastrophic – for example, we would have to admit that the world is densely populated with "ghost brains" that functionally replicate all manner of different consciously perceiving brains. But as Shoemaker notes, it's not actually clear what the relevant difference is between the thin correlational causation that ties together ghost systems and the causation that we observe and theorize in the world. This is the *ghost world puzzle*. It asks what rules out ghosts existing that are functionally like us. The micro-sceptical puzzler on the other hand asks : how do we know that *we* aren't ghosts? (Obviously, there is a close relationship here with triviality arguments against functionalism[23], although I find that Shoemaker's set up gets us into the issues from a different angle in an illuminating way).

The connection here with the interface theory is that it is precisely in the business of challenging whether our downwards inferences are reliable in the sense that they might reveal the objective physical structure of the world. Arguably, ordinary scientific practice assumes that there is an epistemically tractable, structure-preserving map from the fundamental physical world to the manifest world (it might be complex and multi-stage, but it is not totally scrambling). The interface theorist, both with the empirical argument and the indifference argument, aims to show that this is almost certainly *not* the case. To my mind, this amounts to saying that we are almost certainly *are* in a micro-sceptical scenario. (If anything then, the interface challenge is stronger than a mere sceptical challenge that asks : how do you know that you're not in the sceptical situation?[24]).

One way in which this is (arguably) a clearer framing of the interface theory that HSP's own is this. It's very natural to read the interface theory as treating us as non-junk ordinary material objects, with non-junk ordinary brains, and then picturing the non-monotonic scrambling as occurring in our brains – the perceptual system acts as a kind of distorting lens. But of course, as HSP themselves make clear, objects like human bodies and brains are all part of the manifest phenomenal interface, and so are the *results* of the scrambling function. So in fact the "scrambling brain" seems to fall out of the picture when we think things through. As mentioned, this is a place where it's reasonable to suspect that there's a kind of inconsistency in the empirical argument from evolution (which very much seems to be directed at *establishing* a scrambling brain) – but I won't press this further here. I will note that the indifference argument does not depend on the scrambling brain idea, nor on considerations of evolution, and in this way is very much akin to the kind of philosophical challenge we get from micro-scepticism and the possibility of ghost-systems.

---

[22] Shoemaker implicitly assumes a flat space-time, and then takes ghost objects to be regions of space-time whose trajectories have the same shape as regions occupied by ordinary objects in a source world (e.g. a world where a completely different distribution of "ordinary" macro-objects exists).

[23] See Sprevak (2018) for a helpful review.

[24] In this way the interface theorist's challenge is similar to arguments for the Boltzmann brain hypothesis (e.g. Dogramaci (2019)) and the simulation hypothesis (Bostrom (2003)).

This is why I think that the interface theory is not at all banal idea, but rather is interesting and challenging. Let's now consider how a realist (understood as someone who thinks that downwards inference to the objective structural of the world is tractable) might respond. The first thing is to briefly consider what our downwards inference principles typically actually are.

I won't be able to get into much detail on this, but I think it's helpful here to distinguish *formal principles* and *causal principles*. Formal principles might include HSPs favored monotonicity principle, but might also include stronger kinds of structure-preservation, such as linearity[25]. Another kind of formal principle (not unrelated to these others), is *derivational complexity minimization*. There are different ways of measuring complexity. A salient measure here might *Kolmogorov complexity* – the minimum length of program needed to generate a derivation. I think it would not be surprising if it turns out we implicitly strive to minimize this kind of computational complexity in postulating grounds for higher-level states of affairs, although I offer this only as an empirical speculation.

Causal principles tell us how to relate causal structure at higher and lower levels. Two related ideas seem very important here. The first is spatio-temporal locality. We start with the assumption (which has very much turned out to be defeasible), that causal processes operate in a spatio-temporally local way. The second is a principle of causal-mechanism. A causal process occurring in a certain region of space-time is undergirded by a lower-level causal process occurring in the same region. One way to make vivid these principles is precisely to consider how they might break down if we are Shoemakerian ghosts. Intuitively, that would mean that when we trace the causal processes going on, say, locally inside the ghost, we would find a mismatch at the fundamental level whereby "ghost processes" are not mechanistically underpinned in the way our causal principles lead us to believe (for example, the ghost might exist in a completely empty region of space-time!).

Now, there are many interesting questions (unfortunately not addressed here) about how exactly to formulate these principles and how they are related. One thing to flag immediately is whatever view we end up taking of them will translate into a certain understanding of the "structural preservation" at stake in this discussion, which might therefore not *just* include monotonicity as a condition (as already suggested above). For example, theories of the world that violate our causal principles might be said to *not* be structure-preserving in an important sense.[26] And as I argued above, the monotonicity condition itself is in need of refinement, because some non-monotonic maps *are* epistemically reversible.

Once we have completed the *descriptive* task of figuring out how downwards inference works in practice, how might we justify or unify these principles? Although I don't have space to pursue it detail here, I do want to briefly discuss what I see as an intriguing and potentially powerful strategy for addressing this issue. The idea is that we should *start* with the

---

[25] For example, chemical properties like shell-numbers in atoms might be linearly related to features of the quantum-mechanical wave function describing the atom.

[26] If we are trying to triangulate multiple kinds of "structure-preservation" this also makes it more plausible that we could leverage a subset of principles to argue that preservation of one kind or other is violated.

fundamental world, and then consider the question "*what are the real patterns here*?", with the goal of building *upwards* from physical *to* natural high-level objects and properties. One might hope that this would reveal our downwards inferences to involve exactly those principles that would construct a fundamental world from which our manifest world can be recovered as "natural structure".

The way for the realist to approach this, I think, is to ask "what patterns in the fundamental evolution would be of interest to a being who is *purely* interested in the fundamental evolution? (e.g. they do not from the outset have any of our parochial interests in medium sized objects like food and mates). Let us further suppose that such a being is interested in finding convenient compact ways of *summarizing* the fundamental evolution. Famously, Humeans about dynamical laws believe that the laws can be recovered from the fundamental distribution as compact summary of how the world evolves[27]. To my mind, an attractive speculation (ambitious but not totally implausible) is that this approach (broadly understood) can be successfully generalized beyond laws to many other worldly patterns[28]. A nice example is the notion of a material object and the notion of center of mass, as applied in a well-known component of Newton's theory. A set of particles rigidly stuck together will behave like a particle of the same mass located at the object's center of mass. One way of thinking of this is that because the particles are stuck together, there will be great redundancy in a complete description of their trajectories, because they are highly correlated. So if we are concerned with knowing approximately for each particle where it will end up, treating the system as an object is a highly efficient way to summarize these particle trajectories[29].

Let's say that a notion that features in such an efficient summary has "humean objective significance" or just is "humean" or "objectively significant" for short. If the notion of a material object is humean, then notice that, plausibly, any property that helps efficiently predict/explain its trajectory will *also* be humean. So for example, folk psychology could be of interest to a being only concerned with efficiently describing physics, because it is useful for predicting how the correlated particles that make up human bodies will move around. I also think it's plausible that thermodynamic properties like temperature can be given the humean treatment : for example if I want to know roughly where a particle will end up, knowing the thermodynamic properties of the parts of the system that it is embedded in will often be very useful[30]. In this way, the kind of time-directed causal structure that depends on the thermodynamic arrow of time could emerge naturally through humean reasoning from the fundamental level.

These ideas obviously stand in need of much further development, but let's suppose we have recovered a world of high-level patterns as objectively significant in this way[31]. Then I

---

[27] Lewis argued that the laws are axioms in a theory of the world that optimally trades off strength and simplicity (see e.g. Lewis (1994)). For further discussion see Loewer (1996), Cohen and Callender (2009), Ismael (2015), Hall (2015), Callendar (2023), Jaag and Loew (2020).

[28] The idea here is *not* that the ordinary macro-world falls out from the details of a best-system account like Lewis's. It's that Humean reasoning in the spirit of best-systems accounts can recover the macro-world.

[29] It's no coincidence that Dennett (1991) also uses this example in his discussion of 'real patterns'.

[30] For example, consider a particle that is part of a gas that cools or dissipates, or a particle that is part of a moving piece of machinery in an engine whose functioning can be modelled in thermodynamic terms. Thanks to Amalie Trewartha for helpful discussion on this point.

[31] One pressing issue here is this : famously Humean accounts of laws risk making laws anthropocentric because "best system" could mean *best for us* given our cognitive limitations and place in the world (Lewis called this

would make two further, closely connected proposals in a realist spirit. One is that much of our manifest ontology (e.g. the idea of a material object) is objectively significant[32]. A humean super-being would care about mountains and trees and animals and planets, and the features of them that we use to explain and predict their behavior. The second proposal is that when we think about our downwards inference principles, they are likely to give us a fundamental account that makes our manifest ontology objectively significant. That is, objective significance for manifest categories is no accident from the point of view of how we reason about the world – perhaps with thinking of it this way, we construct our physics *so as* to make our manifest world come out as a humean real pattern.

Now, this latter idea, if correct, does invite a line of objection very much in the spirit of the interface theory. Given that our perceptual systems were defined with survival and mating in mind, why think that they would deliver categories with objective significance? Why would it be useful for us Darwinian beings to perceive and think about categories that also would be of interest to a disinterested Humean super-being? Wouldn't that be to illicitly assume that our parochial human perspective offers a direct window onto how the world really is (where now that is understood as the vision of a humean super-being)? Call this the *modified Darwinian debunking argument*.

There is also a version of the indifference argument here. The interface theorist is likely to accuse the realist indulging in these humean speculations as massively begging the question. Of course if we make downwards inferences that reverse engineer the manifest world as objectively significant, realism will seem to have been vindicated. But if we start off uncertain about the form of the grounding function – in particular if we are completely indifferent over the space of possible grounding functions (however ever that be modelled and measured), then the chance that this kind of reasoning is successful is effectively zero.

At this point, the issue of burden of proof in the end-game very much rears its head. In addressing these challenges, I think it's helpful to distinguish two different types of debunking arguments : what I'll call *by-your-lights* and *by-my-lights* debunking arguments. The difference is this. The by-your-lights debunker presents a challenge that is easier to meet, because they are willing to provisionally grant the world-view of their opponent. They challenge a particular assumption of the opponent by asking them to explain, *given the resources of their world-view*, how it is that the assumption is reliably held (or has whatever other epistemically important feature that is at issue). Take for example, our belief in the presence of medium-sized perceptible objects like humans, trees, chairs etc. A by-your-lights debunking challenge against these beliefs can be successfully met by pointing out that given that these items exist, and given that knowing about them would be important for the survival of an organism like us (appealing to our overall scientific world-view), there is every reason why we would have a perceptual system that reliably informed us about them. To meet this challenge then, there is nothing wrong with simply appealing to the assumption in question, as part of a world-view that explains why our making that assumption is reliable (the challenge is therefore in the spirit

"ratbag idealism"). I assume my super-being has an interest in efficiently summarizing the fundamental facts – is that really a parochial human interest that undermines the claim that the revealed patterns are "objective" in an interesting sense?

[32] There is an intriguing resemblance here with "natural scene statistics" approaches to explaining perceptual and cognitive categories in cognitive science (e.g. Geisler (2008)).

of Quine's *epistemological naturalism* (1969)). Notice that this doesn't mean it is trivial to meet the challenge. Consider for example, the intuition that a dualist view of conscious experience is true. One can imagine a theorist, who by their own lights, has this intuition because of the way their brain works, not because non-physical experiential properties exist and have caused the intuition; so the epistemic challenge is effective even if we grant provisionally that dualism is true.

The by-my-lights challenger, by contrast, is an individual who is *not* willing to provisionally grant the assumption in question, and wants an *independent* argument for thinking that it is correct. Famously, this kind of challenge is very hard to meet – to establish anything we need to make assumptions, but those assumptions themselves can be attacked with a by-my-lights challenge. So it's not particularly disturbing if in the end our belief-system can't be given this very demanding kind of vindication. This, famously, is the moral many draw from considering traditional sceptical attacks on human knowledge. On the other hand, if we can't provide the holistic kind of vindication demanded by the by-your-lights challenger, that is more epistemically disturbing.

Now, the way I see it, if the interface theorist's debunking challenges are understood in the weaker by-your-lights sense, they are still non-trivial challenges, but ones that can potentially be met. As the interface theorist's own empirical argument from homeostasis etc. illustrates, one could envisage investigating how our perceptual systems actually evolutionarily developed, and finding that we have the rug pulled out from under our feet, because our best theory tells us that perception is likely to be structurally opaque. However, I do not believe that this is what we will actually find. In particular, if we start by assuming that ourselves, our food, mates, predators and other behaviorally significant objects and features are objectively significant (i.e. that these objects are all humean objects), then it would be adaptive to perceive humean objects, and it would be adaptive to perceive the humean features that explain their behaviors (consider what we said earlier about spatial distance). Furthermore, in so far as it's useful to be able to mentally model the causal mechanisms underpinning manifest causal interactions, it would be adaptive to be disposed to engage in downwards inferences that uncover the humean causal structure underpinning these interactions; and so our reasoning would be revealed (by our lights) to be reliable.

It's true that there can turn out to be local exceptions to this. For example, color and smell could turn out to be quite gruesome and therefore non-humean. In this way, a kind of evolutionary debunking argument against the humean significance of these properties can succeed. It can also turn out that our epistemic principles are unreliable in various ways and need to be locally revised. But presumably we could not bootstrap our way to extremely radical revisions. More generally, if all we face is a by-your-lights challenge, a broadly realist view seems to be structurally built in as an upshot[33].

Of course there is something unsatisfying about this. It would be great to be able to assume nothing and build our worldview from a completely a priori starting point. But of course

[33] It must be acknowledged here that the history of physics shows that we are capable of empirically establishing a surprising degree of "mismatch" between manifest structure and the fundamental structure (quantum mechanics being one famous example). I still think there are surely limits built in here given the priors reflected in our downwards inference principles, but it's an interesting question what the extent is to which they limit the space of viable physical theories (relatedly, are there limits to what we could *observe* as data for our theories?).

this is overambitious. Arguably, the proponent of a by-my-lights version of the interface challenge is guilty of this kind of unreasonable demand. To say we should be indifferent between different downwards inference principles is to say that we should start with no assumptions whatsoever about how the manifest world is generated. But of course if we start there, there is no hope of knowing anything about that generative base. Similarly, if I start with no assumptions about how my perceptual experience is causally generated, there is no hope of recovering the manifest perceptual world. Philosophers have long learned to live with this kind of result, and learned to only expect the more holistic 'internal' kind of vindication, not a full-blooded refutation of the radical sceptic. If the interface theorist is only an old-fashioned radical sceptic, then theory loses some of its interest, or at least fails to be a compelling threat to the realist.

Now, this doesn't mean that progress can't be made within the realist holistic project. We can figure out the relationship between our principles, including trying to recover as much as we can from as few principles as possible, or modifying or clarifying principles when they clash with each other or otherwise result in unattractive consequences. The Humean program that I sketched above is supposed to be an example of this kind of theorizing. It doesn't start from nowhere, because the humean superbeing *must themselves be operating with some principles* – e.g. they want an "efficient summary" of the fundamental world in some substantive sense. What if they had a different interest?

Is this the end of the story? Actually no. At this point the interface theorist regains composure and launches into the following defense. They can ask why, even if their view is structurally akin to, say, radical scepticism about the existence of the external world, it should be considered objectionable in the same way. In particular, they may point to the fact that structural non-veridicality is perfectly consistent with content-based veridicality. So the "scepticism" they are serving up is completely consistent with our having a largely correct view of the world!! They might say: the idea that we should expect or strive for a model of the world that mirrors or structurally corresponds to the objective fundamental structure of nature is an overambitious or hubristic metaphysical gloss on our theorizing that is a philosopher's fantasy and not part of our ordinary theoretical understanding; moreover, it is completely dispensable, pragmatically speaking. In this way, it is quite unlike the idea that our beliefs are more or less true, or that our belief forming methods are at least somewhat reliable – that's an assumption without which enquiry into the world, and in fact just human life in general, cannot go on. Why not settle for a more "internal" Kantian kind of realism, rather than the "external" realism that structural correspondence demands?

Along these lines, one can make an interesting comparison here to the debate about moral realism (i.e. that mind-independent moral norms are part of the fundamental furniture of the world). It is common to offer evolutionary debunking arguments against moral realism[34], and it has also been common for moral realists to respond by saying that they can at least meet a by-your-lights debunking challenge[35], and that by-my-lights challenge really amounts to an

[34] e.g. Street (2006) is a classic example.

[35] If we are allowed to take for granted what the moral facts *are*, then it's at least *less* challenging to explain how our physical evolution lead to us having reliable moral beliefs. For example, if murdering people in your community is objectively wrong, and evolution predictably lead us to think that it is wrong (because e.g. it was adaptive for us to live in peaceful communities), then there is a sense in which this moral belief is reliably formed.

unworrying form of moral scepticism. In this case though, I would be inclined to side with the debunkers! Moreover, it is quite plausible to make the same kind of appeal to modesty and pragmatism – moral realism just doesn't seem to achieve much theoretically or pragmatically in the way that a belief in the reliability of our epistemic norms probably does.

To my mind, this is the most interesting line of defense that an interface theorist can make. But note that it is spiritually far-removed from HSP's interface theory. They see themselves as iconoclasts smashing up common sense with their radical lessons from evolution. This interface theorist is offering their view in the spirit of theoretical restraint and pragmatism.

Now, although I think this is a better way to think about the theory, I'm not ultimately convinced that it is an attractive resting place. There's a more minor and a more major defense the structural realist has up their sleeve. The minor defense is this: they could potentially make the case that the comparison between structural transparency and moral realism is unconvincing, because the moral realist can't even in the end meet the by-your-lights challenge. Even if we take for granted what the moral facts are (construed realistically), do they really explain our moral intuitions in a way that makes it clear that these intuitions are a reliable form of moral perception into a mind-independent moral realm, in the way that visual perception can be argued (in a by-my-lights spirit) to be reliable perception of medium sized objects? That actually does not seem plausible (to me at least!), although of course much more needs to be said (maybe the comparison with perception is unfair). On the other hand, the by-my-lights defense of structural transparency looks to be quite solid (modulo the inevitable sense of question-beggingness).

This defense is more minor, because it doesn't tackle head on the accusation that structural transparency is theoretically otiose and hubristic; it just shows there's a kind of internal coherence to it. The more major response does tackle this head-on though. Here's the thing: one way to put the "modest" view we are contemplating is as saying that there's nothing problematic about the idea that the manifest world, and the rest of the known universe (i.e. the world we think we know through theoretical inference), is really a system of junk objects and properties/relations, only remotely and gruesomely related to the true noumenal structure of the world. But that leads to an acute version of Shoemaker's ghost world puzzle. For if the functional structure that we are familiar with (e.g. the structure involved in our brain's functioning), can exist consistently with our being junk objects, then why doesn't that lead to a massive explosion of equally robust functional structure densely populating the universe? Specifically, if assume a plenitudinous ontology[36], then although we may then have to admit systems of ghost objects and properties whose interactions *mimic* real causal interaction in a modal-correlational sense, we might have hoped that ultimately these interactions are not underpinned mechanistically at the fundamental level in such a way to count as *robustly causal* (e.g. imagine, again, reading a system of ghosts into a completely empty space-time). But if we ourselves are junk, this way of distinguishing ourselves from junk is (arguably) doomed, and we risk having to think of other systems of junk objects and properties as not just existing, but also *not differing in a metaphysically significant way from the world we know and love* (e.g. being

[36] An important lacuna in this paper is (admittedly) the consideration of sparse ontologies that simply deny that junk objects and property instantiations really exist. In other work I consider the viability of these kinds of views and their potential to solve the ghost world puzzle in a different way (Lee (2023), Lee (manuscript)).

robustly functional in just the same way). But that's a wild view : for example, it would seem to entail that every possible functional structure that could characterize every possible brain of the same complexity as the human brain, actually is instantiated in our universe[37]. So the problem is this : metaphysical modesty about who we are, and how our manifest world relates to the fundamental world, might lead to extreme *liberality* about what other beings with genuine functional structure inhabit the universe. And that's not a modest view at all – in fact it's extremely theoretically crazy!!!!

In my current thinking, this is why, in the end, I lean in the direction of structural-transparency realism. To believe that we are alone in the world (qua robustly functional beings), we have to think of the mind as a mirror of nature. I must admit though that a principle that says "avoid functional explosion" is an odd kind of foundation stone. So even if we like this realist picture, there is surely interesting work to be done taxonomizing and theorizing its foundational assumptions; once we have done this, perhaps there is a way of justifying the resistance to functional explosion in a way shows how it follows from, or at least coheres nicely with, other basic tenets. It's also natural to wonder whether the interface theorist can avoid the excesses of explosion without embracing structural transparency. Denying plenitude and holding that unlike our manifest world, the alleged junk-systems do not really exist, is one tempting path along these lines – a path that I sceptically discuss in detail elsewhere (Lee (2023), Lee (manuscript)). One can also consider views that deny the existence of any objective fundamental structure that would constitute the unknowable noumenon, thereby sidestepping the issue of structural transparency. One interesting (but rather obscure) view along these lines also metaphysically privileges the manifest world *above* the underlying structure (it is not a mere "construction", even if the so-called "fundamental" world is) so that there is no possibility of experience failing to present "the world as it really is" (the manifest world is the world at its most real!!). Whether such views can be developed in a compelling way is a question I will have to leave for another time.

References

Beck, J. (2018). Analog mental representation. *Wiley Interdisciplinary Reviews: Cognitive Science*, *9*(6), e1479.

Berkeley, G. (1713/1979). *Three Dialogues between Hylas and Philonous*. Hackett Publishing Company.

Bostrom, N. (2003). Are we living in a computer simulation?. *The philosophical quarterly*, *53*(211), 243-255.

[37] And of course it's not a big step from this to the view that there exist an infinite variety of consciously experiencing subjects! In my view, even if we reject functionalism about consciousness, the problem is serious because we will still have an infinite variety of "quasi-conscious" subjects (see Lee (2019) for elaboration).

Callendar, C. (2023). Humean Laws of Nature : The End of the Good Old Days. In M. Hicks, S. Jaag, C. Loews eds. *Humean Laws for Humean Agents*. Oxford University Press.

Chalmers, D. (2006). Perception and the Fall from Eden. In T. Gendler & J. Hawthorne, eds. *Perceptual Experience*. Oxford University Press.

Chalmers, D. (2015). Panpsychism and panprotopsychism. *Consciousness in the physical world: Perspectives on Russellian monism*, 246-276.

Chalmers, D. (2017). Idealism and the Mind-Body Problem.

Chalmers, D. (2022). *Reality + : Virtual Worlds and the Problems of Philosophy.* Norton.

Cohen, J. (2015). Perceptual representation, veridicality, and the interface theory of perception, 1–10.

Cohen, J., and Callender, C. (2009). A better best system account of lawhood. *Philosophical Studies*, 145(1), 1–34.

Dennett, D. C. (1993). *Consciousness explained*. Penguin uk.

Dennett, D. C. (1991). Real patterns. *The journal of Philosophy*, *88*(1), 27-51.

Dogramaci, S. (2019). Does my total evidence support that I'm a Boltzmann Brain? *Philosophical Studies* 177 (12):3717-3723.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192.

Goff, P. (2017). *Consciousness and fundamental reality*. Oxford University Press.

Hall, N. (2015). Humean reductionism about laws of nature. In B.Loewer and J. Schaffer (Eds.), The Blackwell companion to David Lewis (pp. 262–277). Oxford: Blackwell.

Hoffman, D. D., Singh, M., & Prakash, C. (2015a). The Interface Theory of Perception. *Psychonomic Bulletin & Review*, *22*(6), 1480–506. http://doi.org/10.3758/s13423-015-0890-8

Hoffman, D. D., Singh, M., & Prakash, C. (2015b). Probing the interface theory of perception : Reply to commentaries, *Psychonomic Bulletin & Review*, *22*(6), 1551–1576.

Hoffman, D. (2019). *The case against reality: Why evolution hid the truth from our eyes*. WW Norton & Company.

Ismael, J. (2015). How to be Humean. In B. Loewer & J. Schaffer (Eds.), *The Blackwell Companion to David Lewis*, 188–205. Oxford: Blackwell.

Jaag, S., & Loew, C. (2020). Making best systems best for us. *Synthese*, *197*, 2525-2550.

Langton, R. (1998). *Kantian humility: Our ignorance of things in themselves*. Oxford University Press.

Lee, G. (2016). Worlds, Voyages and Experiences: Commentary on Pelczar's Sensorama. *Analysis*, *76*(4), 453-461.

Lee, G. (2019). Alien Subjectivity and the Importance of Consciousness. In A.Pautz and D.Stoljar eds. *Blockheads!: Essays on Ned Block's Philosophy of Mind and Consciousness*. MIT Press.

Lee, G. (2023). Against Magnitude Realism. *Crítica. Revista Hispanoamericana De Filosofía*, *55*(163), 13-44

Lee, G. (manuscript). Getting out of Ghost World : Grounding and High-level Structure.

Lewis, D. (1983). New work for a theory of universals. *Australasian journal of philosophy*, *61*(4), 343-377.

Lewis, D. (1994). Humean supervenience debugged. *Mind*, 103, 473–90.

Lewis, D. (2001) Ramseyan Humility. In David Braddon-Mitchell & Robert Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*. MIT Press. pp. 203-222

Loewer, B. (1996). Humean supervenience. *Philosophical Topics*, *24*(1), 101-127.

Mclaughlin, B. P., & Green, E. J. (2015). Are icons sense data ?, *Psychonomic Bulletin & Review*, *22*(6), 1541–1545.

Mørch, H. H. (2014). Panpsychism and causation: A new argument and a solution to the combination problem. PhD Thesis, University of Oslo.

Neander, K. (2017). *A mark of the mental: In defense of informational teleosemantics*. MIT Press.

Pautz, A. (2020). *Perception*. Routledge.

Pelczar, M. (2015). *Sensorama: A phenomenalist analysis of spacetime and its contents*. OUP Oxford.

Prakash, C., Fields, C., Hoffman, D. D., Prentner, R., & Singh, M. (2020). Fact, fiction, and fitness. *Entropy*, *22*(5), 514.

Prakash, C., Stephens, K. D., Hoffman, D. D., Singh, M., & Fields, C. (2021). Fitness beats truth in the evolution of perception. *Acta Biotheoretica*, *69*, 319-341.

Putnam, H. (1981). *Reason, truth and history*. Cambridge University Press.

Putnam, H. (1983). *Realism and Reason: Philosophical Papers, Volume 3*. Cambridge: Cambridge University Press

Quine, W. (1969). Epistemology Naturalized. in *Ontological Relativity and Other Essays*, New York: Columbia University Press.

Roelofs, L. (2019). *Combining minds: How to think about composite subjectivity*. Oxford University Press.

Russell, B. (2001). *The problems of philosophy*. OUP Oxford.

Shea, N. (2018). *Representation in cognitive science*. Oxford University Press.

Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive psychology*, *1*(1), 1-17.

Shoemaker, S. (1988). On what there are. *Philosophical Topics*, *16*(1), 201-223.

Sider, T. (2013). *Writing the Book of the World*. OUP Oxford.

Sprevak, M. (2018). "Triviality Arguments about Computational Implementation". In Sprevak and Colombo eds. *The Routledge Companion to the Computational Mind*. Routledge.

Strawson, G. (2009). Realistic monism: why physicalism entails panpsychism. *Journal of Consciousness Studies*, *13*(10-11).

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical studies*, 109-166.

Wüthrich, C. (2019). The emergence of space and time. In *The Routledge handbook of emergence* (pp. 315-326). Routledge.