

IGNORANCE BY ACQUAINTANCE

*It seems better to persevere in the attempt to analyze and clarify
the somewhat muddy ideas commonly called up by the word 'experience',
since it is not improbable that in this process
we may come upon something of fundamental importance
to the theory of knowledge*
—Russell, 'The nature of acquaintance'

Something very unusual is going on here
—Chalmers 2003

Aims in this handout:

- Get a sense for Chalmers's apparatus in 'The content and epistemology of phenomenal belief'
- Application to action: (1) agentive and causal concepts (2) care and feeding of indiscriminability

I. On CEPB

Consider Black-and-White Mary, who is just seeing a red thing for the first time: she goes R.

We can distinguish the following sorts of concept which Mary might deploy in referring to R:

1. 'That phenomenal property most people have when they see a red thing'
2. 'That phenomenal property I have when I see a red thing'
3. 'That phenomenal property which I am having now'
4. '*This very property*'

Notes about these concepts:

- Mary could have had any of the first three without having seen a red thing.

- There is a sense in which each of the first three leaves open just which property it is referring to. This is easy to see in the case of the first two: Mary would have been able to deploy (1) or (2) in her BW room. She could deploy either of those without knowing what it is like to see a red thing.
- Something similar is true of the third concept, tho this can be slightly harder to see. Since it is available prior to leaving the BW room, it can be used to refer either to R (after she leaves) or to B or W (before she leaves). Accordingly *that concept* is not specifically tied to any particular phenomenal property. It's rather like 'here' or 'me' or 'now', which are available even if I have no idea who where or when I am. If this is obscure, it is perhaps because we tend to think of such "demonstrations" as involving some sort of fixation on the referent; however, the third concept can be deployed "blindly" while thinking about mathematics.
- However, none of this is true of the fourth concept. It is not available until BWM has left, it cannot be deployed blindly, and if one is deploying it one plausibly knows what it is like to see a red thing.

This last—the sort of concept that goes with *knowing what it's like to see a red thing*—is the sort of concept that Chalmers is presenting a story about, his concept *R* (the third is *E* and the first and second are *red_C* and *red_I* respectively).

Key doctrines:

- A. A metaphysical doctrine: One is "acquainted with" the phenomenal properties of one's experience.
- B. A psychological doctrine: (1) There is such a thing as a *quasi-direct phenomenal concept*: a concept the deployment of which divides into (a) an act of considering a predicate in "cognitive background" (b) an act of "acquaintance-based" attention to one's experience; (2) we are often in a position to deploy an extensive range of quasi-direct phenomenal concepts.
- C. A metasemantical doctrine: The (rigid) referent of such a concept is the phenomenal property which (a) satisfies the predicate and (b) is maximally similar to some phenomenal property instantiated in one's experience.

D. A doctrine about content: A quasi-direct phenomenal concept K presents its referent under a nature-revealing mode, in the following sense. Suppose that K is a QDPC: then, for any concept F that is *explicitly more fundamental* than K, it is a priori whether K = F.

- Contrast: 'water' is not nature-revealing. Why? It is equivalent to something like the existentially quantified concept 'the local watery stuff', thus explicitly not especially fundamental; concepts like 'H₂O' and 'XYZ' are more fundamental; it is not a priori whether water = H₂O or whether water = XYZ.
- QDPCs are nature-revealing for the simple reason that they are conceptually absolutely fundamental, hence it is vacuously true that there is no explicitly more fundamental concept such that blah blah. (We go into Chalmers's actual discussion of this below at significantly greater length.)

Terminology: a quasi-direct phenomenal concept for which the referent is identical to a phenomenal property instantiated in one's experience (for which similarity goes to identity) AND for which there is a reasonably wide range of nearby possibilities for which this is true is a *direct* phenomenal concept; otherwise a *pseudo-direct* phenomenal concept. *Direct (etc) phenomenal judgment*: judgement predicating a direct phenomenal concept of one's own experience (where the experience is demonstrated in the act of attention that is a constituent fo the DPC).

E. An epistemological doctrine: A direct phenomenal judgement about one's own experience is prima facie justified by acquaintance.

A. On acquaintance

Two questions: (1) What is acquaintance? (2) Why believe in it?

1. The nature of acquaintance

- "a special relation to the phenomenal properties instantiated in our experience: a relation that we do not bear to the other instantiated properties in question, and a relation that is required in order to form a direct concept of a property in the manner described. This relation would seem to be a peculiarly intimate one, made possible by the fact that experiences lie at the heart of the mind rather than

standing at a distance from it; and it seems to be a relation that carries the potential for conceptual and epistemic consequences. We might call this relation *acquaintance*."

- "Acquaintance can be regarded as a basic sort of epistemic relation between a subject and a property."
- "Acquaintance is not itself a conceptual relation: rather, it makes certain sorts of concepts possible. And it is not itself a justificatory relation: rather, it makes certain sorts of justification possible."
- "One need not regard the acquaintance relation that a subject bears to a phenomenal property as something ontologically over and above the subject's instantiation of the property, requiring a subject-relation-quality ontology at the fundamental level. It is arguable that it is a conceptual truth that to have a phenomenal quality is to be acquainted with it (at least in so far as we have a concept of acquaintance that is not wholly theoretical). Certainly it is hard to conceive of a scenario in which a phenomenal quality is instantiated but no one is acquainted with it. If so, then the picture I have sketched is combined with a simple subject-quality ontology, combined with this conceptual truth. The ontological ground of all this might lie in the nature of phenomenal qualities, rather than in some ontologically further relation."

2. The case for acquaintance

A theoretical argument:

- "It is equally possible to regard acquaintance as a theoretical notion, inferred to give a unified account of the distinctive conceptual and epistemic character that we have reason to believe is present in the phenomenal domain. "
 - a. *Can* form nature-revealing concepts of phenomenal properties; *cannot* form nature-revealing concepts of anything else; this divergence is best explained by positing a distinctive relation between subjects and phenomenal properties.
 - b. Need an explanation of why direct phenomenal judgements are justified, and statistical stuff doesn't do the trick.
 - c. Even nicer that there is a package which dusts off both these explanations at once.

Buttressing considerations:

- “Some philosophers (e.g. Russell 1910; Fumerton 1995) have held that we are “acquainted with acquaintance”, and have made the case of its existence that way. I think there is something to the idea that our special epistemic relation to experience is revealed in our experience, but I note that the proponent of acquaintance is not forced to rely on such a thesis.”
- “Even if acquaintance is a theoretical notion, it clearly gains some pre-theoretical support from the intuitive view that beliefs can be epistemically grounded in experiences, where experiences are not themselves beliefs but nevertheless have an epistemic status that can help justify a belief. One might view acquaintance as capturing that epistemic status.”

B. Quasi-direct phenomenal concepts

Once again, two questions: (1) What are these beasts? (2) Why believe in them? And one further question: (3) How many of them are there?

1. The nature of quasi-direct phenomenal concepts

- “a quasi-direct phenomenal concept arises from an act of (intended) demonstration, along with a characteristic sort of cognitive act.”
- “The clearest cases of [quasi-]direct phenomenal concepts arise when a subject attends to the quality of an experience, and forms a concept wholly based on the attention to the quality, “taking up” the quality into the concept.”
- “As with all acts of demonstration and attention, phenomenal demonstration and attention involves a cognitive element. Reference to a phenomenal quality is determined in part by cognitive elements of a demonstration. These cognitive elements will also enter into determining the content of a corresponding [quasi-]direct phenomenal concept. Consider two individuals with identical visual experiences. These individuals might engage in different acts of demonstration — e.g. one might demonstrate a red quality experienced in the right half of the visual field, and the other a green quality experienced in the left half of the visual field — and thus form distinct [quasi-]direct phenomenal concepts. Or they might attend to the same location in the visual field, but

demonstrate distinct qualities associated with that location: e.g. one might demonstrate a highly specific shade of phenomenal redness, and the other a less specific shade, again resulting in distinct [quasi-]direct phenomenal concepts. These differences will be due to differences in the cognitive backgrounds of the demonstrations in the two individuals.”

Couple of noteworthy points:

- We zeroed in on BW Mary but there is no requirement of “novelty” behind the quasi-direct pcp: can deploy it for the hundredth time (“yep, this is still what it’s like to see red”).
- Other sorts of “robustly phenomenal” concepts are ones involving imagination, memory, and projection, but quasi-direct phenomenal concepts are those which involve genuine attention to phn qualities.

2. The case for quasi-direct phenomenal concepts

Argument here seems to run as follows:

- a. BW Mary has new knowledge of phenomenal properties when first seeing a red thing;
- b. But she was already in a position to know that seeing a red thing is *red_I* and that seeing a red thing is *red_C*;
- c. So that does not suffice for knowing what it is like to see a red thing, so those are not our only concepts of R;
- d. And she was already in a position to entertain the thought that seeing a red thing is *E*, although she was not in a position to entertain the thought with the truth-condition that seeing a red thing is R in that manner; moreover, she could have *inattentively* entertained the thought that seeing a red thing is *E* after leaving the BW room with that truth-condition;
- e. So we need some further concept to explain the nature of knowing what it is like to see a red thing.
- f. Apparently what is missing here is *attentiveness* to R, so that should be built in to the basis of the relevant concept;
- g. And cognitive background comes along to provide such necessities as focusing on the right versus the left, focusing at high or low level of determinacy, and the like.

Good argument.

3. How widely available are these?

These concepts are supposed to explain “knowledge of what it is like”, so whenever that is available, they are.

However, they might be utterly unavailable in some cases: “for all I have said, it may be that some phenomenal states, such as fleeting or background phenomenal states, *cannot* be taken up into a direct phenomenal concept, perhaps because they cannot be subject to the right sort of attention”.

C. The determination of reference

Here the questions are: (1) Why tie the reference to the demonstrated quality at all? (2) Why involve the cognitive background at all? (3) What if the two get out of whack?

1. The tie to the demonstrated quality

Consider Inverted Mary, who is a physical-functional-environmental duplicate of Mary but for whom what it is like to see a red thing is the same as what it is like for Mary to see a green thing, and so forth: Inverted Mary goes G when she looks at a stop sign, and so forth.

When Inverted Mary is freed from her BW cell, what does she learn? The same thing that Mary learns? No, Mary comes to learn something with the truth-condition that she goes R when she sees a red thing, while Inverted Mary comes to learn something with the truth-condition that she goes G when she sees a red thing.

But if so:

This is already enough to draw a strong conclusion about the irreducibility of content. Recall that Mary and Inverted Mary are physical/functional and environmental twins, even after they see red things for the first time. Nevertheless, they have beliefs with different contents. It follows that belief content does not supervene conceptually on physical/functional properties. And it follows from

this that intentional properties are not conceptually supervenient on physical/functional properties, in the general case. [...]

To start with, it is natural to hold that the content of phenomenal concepts and beliefs supervenes conceptually on the combination of physical and phenomenal properties. Mary and Inverted Mary are physical twins, but they are phenomenally distinct, and this phenomenal distinctness (Mary experiences phenomenal redness, Inverted Mary experiences phenomenal greenness) precisely mirrors their intentional distinctness (Mary believes that tomatoes cause R experiences, Inverted Mary believes that tomatoes cause G experiences). It is very plausible to suppose that their intentional distinctness holds in virtue of their phenomenal distinctness.

The alternative is that the intentional content of the phenomenal concept is conceptually independent of both physical and phenomenal properties. If that is so, it should be conceivable that two subjects have the same physical and phenomenal properties, while having phenomenal beliefs that differ in content. Such a case might involve Mary and Mary' as physical and phenomenal twins, who are both experiencing phenomenal redness for the first time (while being phenomenally identical in all other respects), with Mary acquiring the belief that tomatoes cause R experiences while Mary' acquires the belief that tomatoes cause G experiences. It is not at all clear that such a case is conceivable.

2. The cognitive background

This is fairly clear: am I demonstrating over on the left or on the right? So as to demonstrate maximally specific phenomenal shade or something more general? The act of attention alone won't get the answer, need something else.

Inclined to think one can treat the contribution of the cognitive background as a *predicate*, which whatever the ultimate referent of the concept is must satisfy.

3. Defective quasi-direct phenomenal concepts

The case of Nancy:

Along with this cognitive element comes the possibility of failed demonstration, if the cognitive element and the targeted experiential elements mismatch sufficiently. Take Nancy, who attends to a patch of phenomenal color, acting cognitively as if to demonstrate a highly specific phenomenal shade. Nancy has not attended sufficiently closely to notice that the patch has a non-uniform phenomenal color: let us say it is a veridical experience of a square colored with different shades of red on its left and right side. In such a case, the demonstrative phenomenal concept will presumably refer to no quality at all: given its cognitive structure, it could refer only to a specific quality, but it would break symmetry for it to refer to either instantiated quality, and presumably uninstantiated qualities cannot be demonstrated.

What of any associated direct phenomenal concept? It is not out of the question that the subject forms *some* substantive concept where a direct phenomenal concept would normally be formed; perhaps a concept of an intermediate uninstantiated shade of phenomenal red, at least if the instantiated shades are not too different. Like a direct phenomenal concept, this concept will have a content that depends constitutively on associated qualities of experience (Inverted Nancy might form a concept of an intermediate phenomenal green), but it will not truly be a direct phenomenal concept, since its content will not directly mirror an underlying quality.

I believe that what is being said here is that something like the following lexical ordering of rules determines the referent of a quasi-direct phenomenal concept:

1. Look at the predicate that is contributed by the cognitive background: whatever the referent is, it must satisfy the predicate (“given its cognitive structure, it could refer only to a specific quality”);
2. Then look at the attended phenomenal quality: the referent is the quality which is as similar to it as possible, given the first rule (“It is not out of the question that the subject forms *some* substantive concept where a direct phenomenal concept would normally be formed; perhaps a concept of an intermediate uninstantiated shade

of phenomenal red, at least if the instantiated shades are not too different. [...] this concept will have a content that depends constitutively on associated qualities of experience (Inverted Nancy might form a concept of an intermediate phenomenal green”).

Is this a good way to think about things? Somewhat inclined to think so: want to say that there is some distinctive “knowing what it’s like” concept made available even when there is this sort of mismatch; plausible that Inverted Mary considerations bring some sort of dependence on the underlying quality; plausible that in general overconstraint brings with it some sort of necessity of compromise if referent is to be secured. Not entirely confident that the determination of reference has to go exactly like this: one reason to think so might be that the cognitive background plays an overriding role in reasoning, and that (for reasons we will see shortly) it therefore has to play an overriding role in the determination of referent as well.

We can treat these concepts as “rigid”, if we like: as referring to the same thing relative to every metaphysically possible world.

D. Fregean considerations

1. Something very unusual

We have seen the facts about the referent of QDPCs, now let us turn to facts about their sense.

To begin with, recall that there is in some sense a dependence of the referent of QDPCs on one’s phenomenal properties.

Contextual dependence of the referent of some concept is nothing super-exciting. We have seen examples such as ‘I’, ‘here’, and ‘now’. In the realm of phenomenal concepts, we have seen that *E* and *red_I* and *red_C* are also contextual in this way: the referent of *E* depends on whatever phenomenal property I have now, that of *red_I* depends on whatever phenomenal property red things usually cause me to instantiate, and so forth. Chalmers also thinks that ‘water’ is a contextually dependent concept: its uses pick out the watery stuff that is typically most abundant in my location (H₂O on Earth, XYZ on Twin Earth).

However, Chalmers thinks that this is not the sort of dependence that is involved in QDPCs. The argument seems to run as follows:

Mary is in a position to know, merely on the basis of knowing that she is *R*, that she is not in Inverted Mary's situation. Similarly, Inverted Mary is in a position to know, merely on the basis of knowing that she is *G*, that she is not in Mary's situation.

Knowledge that one is *R* and knowledge that one is *G*, therefore, put one in a position to "rule out" different "epistemic possibilities": once Mary knows that she is *R*, she can rule out the possibility that she is in Inverted Mary's situation (but not that she is in her own situation!), while Inverted Mary's knowledge that she is *G* puts her in a position to rule out the possibility that she is in Mary's situation.

More generally, for any phenomenal property *F* that is in fact incompatible with *R*, Mary's knowledge that she is *R* puts her in a position to rule out that she is *F*; mutatis mutandis for Inverted Mary. At every epistemic possibility for Mary, she is *R*.

By contrast, when Oscar (on Earth) or Twin Oscar (on Twin Earth) comes to know that there is water in his glass, he is not thereby in a position to know that he is not in the situation of the other. Neither is thereby in a position to know that XYZ or H₂O is not in his glass; neither is thereby in a position to rule out those epistemic possibilities.

Let's think of things in terms of "rigidity", or dependence of referent on possible worlds. We have already seen that *R* and *G* rigidly refer to *R* and *G*, respectively. In this respect, they are no different from Oscar and Twin Oscar's 'water' concepts. Where they differ is in that *R* and *G* seem to be rigid in another sense in which the 'water' concepts are not: *R* refers to *R* (*G* refers to *G*) in every *epistemic* possibility; by contrast, 'water' refers to H₂O in some epistemic possibilities, XYZ in others.

Now hold on, you say: this begs the question against the analytic functionalist who thinks that *R* is a functional concept like 'that neural property which red things cause me to instance'. From the perspective of such a theorist, Mary's *R* is no better off than Oscar's

'water': it refers to R-fibers in her actual case, but G-fibers in others ... ; just as Oscar's 'water' refers to H₂O in his actual case, XYZ in others.

Well yes: but recall the adherence to "phenomenal realism" at the top of the paper: this is in effect the view that there is some concept like *R* which is not analyzable. (Inverted Mary is already unacceptable to the analytic functionalist.)

Chalmers endorses a view of sense according to which two concepts have the same sense just if they contribute in the same way to providing one a basis for ruling out epistemic possibilities (so-called "two-dimensionalism"). Accordingly, *R* has a distinct sense from *G*, and one which is also distinct from that of any functional or physical concept. We could think of this sort of sense, because it leaves no options for which property its referent might be, as being "nature-revealing". If so, quasi-direct phenomenal concepts are, as desired, nature-revealing.

2. Acquaintance makes the scene

This epistemic rigidity is the fact that acquaintance is initially wheeled out to explain.

Terminological excursus

A token QDPC gets to be *direct*, recall, just if (a) it involves meshing cognitive background and underlying quality and (b) throughout a reasonable range of counterfactual circumstances, "it" remains so; otherwise *pseudo-direct*. How do we track "it" around among counterfactual circumstances?

Consider the QDPC that Mary tokens. This has a certain sense, involves a certain cognitive background, involves a certain ostended quality, is deployed by a certain person at a certain bit of their life-history. Any of these could be used to set up types to which it belongs.

Which type is desired here? Don't want it to be the sense, or the ostended quality: the idea is supposed to be that if it were a bit phenomenal-redder I would deploy a concept to reflect this. Rather, maybe something like a combination of the cognitive background and life-history criteria.

(To see why (b) adds something, suppose that in my cognitive background I apply some extremely stringent condition on my phenomenal property: it has to be *utterly* constant. I am in a circumstance in which, although this is true, there is a very large chance of a small amount of noise interfering with perfect constancy. In such a case we would say that I got lucky in avoiding cognitive mismatch.)

A direct phenomenal *judgement* is one which applies a direct phenomenal concept to an experience picked out in the same act of attention.

F. Epistemological considerations

Direct phenomenal judgements are “incorrigible”: first, they are true (this can be seen by unpacking the definition); second, there is a modal dimension to this, due to the tracking in nearby neighbourhoods.

Don't want to get too much into epistemology here, but a big idea in the last several decades has been that this sort of statistical condition does not suffice for normativity: we want something suitably “internalistic”, reflecting an “awareness of evidence” or some such. Example: I guess about the weather in Brampton regularly and completely without any evidential basis; a mischievous spirit unbeknownst to me foresees my guesses and fixes the weather to match my guesses; plausibly, despite my perfect record of accuracy I lack justification for these guesses.

Acquaintance has been wheeled out to explain the availability of nature-revealing concepts, and it returns to provide the internalistic element.

G. Application

Chalmers's key examples:

- Black-and-White Mary upon leaving her cell, seeing a red thing, going R, focusing on R = what her visual experience is like, and thinking—‘the phenomenal property of (my typical) experiences of seeing a red thing = *this very property*’;
- Inverted Mary going G, focusing on G, thinking ‘the phenomenal property of (my typical) experiences of seeing a red thing = *this very property*’

- Nancy seeing a thing which causes her to go half R16 and half R18, thinking ‘the homogeneous phenomenal property of my current visual experience = *this very property*’

Running through the cases in application:

- Regular Mary:
 - a) Is acquainted with R/phenomenal red;
 - b) Focuses “acquaintance-based attention” on her total experience; presupposes that the referent will be a property of visual aspects of her total experience caused by the thing she is looking at (and, perhaps, that it is caused in the typical way by a red thing); thinks ‘the current phenomenal property of (aspect A) of my experience = *this very property*’;
 - c) What that token thereby refers to is that property which is of a visual aspect of total experience caused by the thing she is looking at (and, perhaps, that is caused in the typical way by a red thing), which is also closest to some phenomenal property instanced in her current experience: namely, phenomenal red;
 - d) Phenomenal red's nature is thereby revealed by that token for the trivial reason discussed above;
 - e) The judgement is prima facie justified because Mary's experience *is* phenomenal red, and this is (let's assume) safely the case.
- Inverted Mary:
 - a) Is acquainted with G/phenomenal green;
 - b) Focuses “acquaintance-based attention” on her total experience; presupposes that the referent will be a property of visual aspects of her total experience caused by the thing she is looking at (and, perhaps, that it is caused in the typical way by a red thing); thinks ‘the current phenomenal property of (aspect A) of my experience = *this very property*’;
 - c) What that token thereby refers to is that property which is of a visual aspect of total experience caused by the thing she is looking at (and, perhaps, that is caused in the typical way by a red thing), which is also closest to some

phenomenal property instanced in her current experience: namely, phenomenal *green*;

- d) Phenomenal green's nature is thereby revealed by that token;
 - e) The judgement is *prima facie* justified because Inverted Mary's experience *is* phenomenal green, and this is (let's assume) safely the case.
- Nancy:
 - a) Is acquainted with a mix of R16 and R18;
 - b) Focuses "acquaintance-based attention" on her total experience; presupposes that the referent will be a homogeneous phenomenal property of visual aspects of her total experience; thinks 'the current phenomenal property of (aspect A) of my experience = *this very property*';
 - c) What that token thereby refers to is that property which is a homogeneous phenomenal property of visual aspects of her total experience; which is also closest to some phenomenal property instanced in her current experience: namely, R17;
 - d) R17's nature is thereby revealed by that token;
 - e) The judgement is *not* *prima facie* justified because the QDPC does not refer to a property Nancy's experience in fact instantiates.

We can now generalize this model: let us say that an *acquaintance-based concept* is a concept that fits this model, without prejudicing what sort of property the object of acquaintance is. Chalmers thinks that the referents of acquaintance-based concepts are phenomenal properties; I think they are kinds of action.

We can say that a *direct* acquaintance-based concept is one which affords knowledge of its referent, in the sense that if one deploys it in self-ascriptive judgement, the resulting *direct acquaintance-based judgement* is *thereby* knowledgeable. (Perhaps the conditions Chalmers imposes for directness in his sense suffice for directness in my sense.)

II. Causal concepts of action

Why suppose the model applies to first-person concepts of action? I think the heart of the case for the model is this fairly quotidian observation:

- * There is a certain class of concepts of types of experiences such that each is
 - i. Learnable only "from the first-person" (by those who have undergone (or imagined) something like its referent);
 - ii. (Hence?) Indefinable/fundamental;
 - iii. Associated with a certain limited degree of privileged first-person access.

(A)–(E) can be seen as providing a nice clean "explanation" of this against the background of classical semantics. (A), (B), and (D) explain (i); adding (C) explains (ii); adding (E) explains (iii). None of this takes a stand on what exactly the kinds of experiences are, so one who agrees with me on this question should accept that the model applies to action.

Of course, one could perform a *modus tollens*: the model doesn't apply to concepts of action, hence actions aren't kinds of experience.

The main reason for thinking this would be that concepts of action are *causal* concepts, hence learnable from the third-person. A cluster of related data surrounding this idea:

1. In the same circumstances understood causally, Mary goes R and Inverted Mary goes G. This is conceivable: it shows that DACs of phenomenal properties are independent of causal concepts. Not so for concepts of action. The analogue would be Joan and Inverted Joan, in the same causal circumstances, but such that Joan is gardening and Inverted Joan *is*—not just seems to be—chopping garlic. That is inconceivable.
2. We can use action-terms to describe the behaviour of even *weak* zombies. This weak zombie is pressing a button; that weak zombie is combing its hair (thanks to Elena Koren for pushing me on this point).
3. When we think of a weak zombie, what we think of is a being with qualities but without *perspective*. Perspective is simple and unified in the strongest sense possible; actions are complex and sometimes in disharmony.
4. Easy to imagine someone being strong-zombified piecemeal: first RHS of vision turns off, then LHS, then high tones, then low tones, then itch By contrast, actions are interdependent, hard to

imagine someone being weak zombified by continuing to write a thesis but stopping writing any of its chapters, etc.

My response is two-fold:

1. Inclined to think that concepts of action *entail* but are not *entailed by* causal concepts. This is compatible with conceivability of weak zombies (causal concepts apply, action concepts don't) but not with conceivability of action inverts (if the action concepts apply, this fixes the causal concepts which apply). Inclined for a long time to think standard inversion in the same boat: unhappy with a conception of Q-teal and Q-canary yellow being causally treated as more similar than either is to Q-turquoise (classic paper on this by Hilbert and Kalderon).
Consequence of this is that once one has come to know what it is like to A, one can extract a causal concept "A" and apply it to zombies. Probably also in many cases having a causal concept of a sort of action is available as an aid to the imagination: watching someone who is Aing, I form a causal concept "Aing", and experiment around in my imagination until I imagine an action type, my DAC of which which entails "Aing". (I note that Goldman, who thinks that a great many mental concepts are based in acquaintance, also accepts an extensive role for ones that are more "theoretically" based. My approach therefore inherits his data to some degree.)
2. Inclined to think "turning on the lights" on a weak zombie is a simple (by which I mean "uncomplex", not "easy"!) matter: *conscious life* is the most fundamental kind of experience/action, thereby bestowing the status of experience/action on all of its parts "at once".

III. Acquaintance-based concepts and indiscriminability

- Recall that a is indiscriminable from b/an F (in manner W) just if it can't be known (in manner W) that a ≠ b/that a isn't F.
- We say that a and b are *trivially* indiscriminable if this is so because a = b; *substantially* indiscriminable otherwise.

A. Applications of indiscriminability

1. Analyzing intention

Context: a few weeks ago we hypothesized that intending to A is doing something which is introspectively indiscriminable from a case of *A-ing in good circumstances*, where good circumstances are defined by reference to the context of attribution.

A bit of support for thinking they're not defined primarily by reference to the attributee (though in certain cases we might wish to do this):

- Thinking that the notion of intention is primarily social, useful for prediction and assignment of guilt. These are highly commonplace practices and thus should not rest too much on access to information that might be hidden, such as what exactly someone does or does not know or mistakenly believe (what someone *should* know by contrast is not hidden).
- Shua Knobe has interesting experimental evidence bearing on this: in minimal pairs differing solely in terms of the goodness or evil of the outcome, people are more inclined to ascribe intention in the evil case than in the good case; significantly, the famous cases are ones in which the acting subject is ignorant of some fact that the attributor knows.
Knobe's minimal pair: Greg (Edgar) dumps Q in Alan's well, ignorant of the fact that Q will thereby flow into the city water supply. Did Greg (Edgar) Q-ify the city water supply intentionally? For Greg Q is fluoride; for Edgar it is lye.
***this sucks, change it

How to build this hypothesis into the acquaintance-based concept model?

Basic idea:

- For Bill to intend to A (from my perspective) is for Bill to be performing some action a such that I am unable to produce a direct acquaintance-based judgement (entailing) that a is not a case of *A-ing in good circumstances*.

A-ing in good circumstances? How does that get into the act? Stay tuned ...

Example: Bill is running away from a bear. He still counts as in good circumstances (goodness typically has to do with lack of significant relevant ignorance, recall). Hence it is true that he is running away from a bear in good circumstances; hence trivially indiscriminable from doing this; hence counts as intending to run away from a bear.

Example: Bill is running away from a bear in a dream. Taking up Bill's perspective, I am in a position to produce the DAJ that this experience is a case of running away from a bear in good circumstances. The perspective I take up isn't inconsistent, so I'm in no position to produce a DAJ that entails otherwise. Hence Bill's exp is substantively indiscriminable from running away from a bear in good circumstances; hence counts as intending to run away from a bear.

Example: Setiya's Fred, who was going home by the shortest route possible but has unknowingly diverged from that route. We count him as *intending* to go home by the shortest route possible though he is no longer *going* home by the shortest route possible because, taking up his perspective, we can't tell that the experience isn't a case of doing so.

Example: Davidson's ten carbon copies: (a) *making* ten cc; (b) *intending* to make them; hence plausibly (c) making them *intentionally* despite (d) no position to know whether (a). Problem, if think Aing intentionally = in a position to make DAB that Aing, (c) and (d) are incompatible.

2. Analyzing narrowness

More context: a phenomenal property is supposed to be *narrow*. What does this mean? F is narrow just if ...

- ... intrinsic duplicates must share F? —but in my view, cases of *running from a bear* and *running from a bear in a dream* are intrinsically distinct.
- ... for all possible creatures, if those creatures are the same inside the skin, one has an experience which is F just if the other does? —but zombies are the same inside the skin as we are, hence no property of an experience is narrow.
- ... for all *nomologically* possible creatures, blah blah blah? —(i) assumes zombies are nomologically impossible why?; (ii) what about creatures that don't have skin?; (iii) even supposing we have some sharp conception of the boundary of the body of a creature, why couldn't I offload some of the processing of my experiences to my computer? (iv) assumes creatures are the only things that can have experience.

- ... it is *this* kind of property, where foils are obvious and paradigms are given by cases of pairs of creatures seeing/acting the same way on twin objects, inversion, illusion, error, dream, envatment, etc? —but how do we get beyond the initial sample, to determine an indefinite extension and antiextension? how do we build the “rails to infinity”?
 - In a metaphysical way: appeal to natural kind structure in the world? —But I think that the natural kinds of experience are “broad” properties like *running from a bear*.
 - In an epistemological way: by appeal to idealization of the epistemic capacities used in establishing the paradigms and foils? —OK, I think what was going on there was that the foils are introspectively discriminable from one another and the paradigms not.

So the idea is that a property is narrow just if it is shared among experiences which are introspectively indiscriminable.

Another question: how to make sense of this in the acquaintance-based concept model? Basic idea:

- For e and e* to be paradigm sharers of a narrow property (from my perspective) is for me to be unable to produce direct acquaintance-based judgements about both e and e* which are a priori inconsistent.

The forms of these judgements, recall, are *this* experience is like *this*, while *that* experience is like *that*. The subject terms are indexical (the experience I am undergoing), while the predicates are substantive. The judgements are therefore a priori inconsistent just if it is a priori that no one has an experience which is both like *this* and like *that*. Since the concepts are nature-revealing, this will be so just if it is *impossible* for an experience to be both like *this* and like *that*. Example: phenomenal-red-all-over v phenomenal-green-all-over; counterexample: phenomenal-green-on-the-left v phenomenal-red-on-the-right.

Example: Recollecting my experience yesterday at the time of writing this, I was attentively wrapping up an undergrad lecture on

Frege. I issue a DAJ of that experience which entails that the subject is doing so; issuing a DAJ of my current experience, it has me attentively typing notes for a grad lecture on indiscriminability. No one could be doing both things at once, so I am in a position to discriminate these experiences in the relevant sense; hence the experiences differ phenomenally.

Example: Renee is looking out the window at people's hats and coats; world w' is exactly like Renee's world, so Renee's counterpart at w' is having the same kind of experience. As a result, since a DAJ is knowledgeable I am trivially in no position to issue in DAJs that entail that no one could be having experiences of both types. Hence they are phenomenally the same.

Example: Vat-Renee is not a zombie and has a brain the same as Renee's. For whatever reason, I am not in a position to issue DAJs that reveal the true nature of Vat-Renee's experience; hence they are phenomenally the same.

3. Philosophical questions about these analyses

The proposals raise the following questions:

1. Both applications require "introspective" knowledge of the other: how can we make sense of that?
2. How can there be the requisite substantive gaps in my ability to issue direct acquaintance-based judgements: cases in which, though the requisite DAJ is true, I can't produce it?
3. What to do about the inclusion of the circumstances in the analysis of intention? More generally, are there distinctive requirements we need to beware of in transferring over a model built for a focused range of "narrow" properties into a highly extensive range of "broad" properties?

B. Introspecting you

Easy question first. Note that on the acquaintance-based concepts model, if we *imagine* being F, we thereby gain acquaintance with F; and are thereby in a position to deploy acquaintance-based concepts of F. (Also in a position to deploy acquaintance-based concepts of *imagining* being F—ignore this.)

Now, if I am in position to know that Joan has the property I have imagined, then I am in position to know that Joan is F.

Our characterization of Joan's experience might be imperfect either by being incomplete, or by being inaccurate. But *indiscriminability* is a modal notion, allowing for extensive idealization. We can assume that our imagination is not flawed in either of these ways. Our imagined experiences are as good as the real thing, for our purposes.

What if we wish to impose limits on this assumption? Well why would we? One answer—maybe we think there are *aliens*, creatures with experiences that we are not in a position to imagine: Nagel's bat? Wittgenstein's lion? Hyperintelligent, crazed, and depraved shapeshifters from planet Grogon? If so, this would mean that I can not form any direct acquaintance-based judgement about those experiences; a fortiori not be in a position to form the relevant sorts of DAJs about them. Accordingly the alien would intend everything and share every narrow property with everyone. Weird: response is to kick them out of the domain. Is that bad? Well it would be bad if I were trying to provide a *friendly* analysis of the notions of intention and phenomenality, on which they end up as robust features of reality. Fortunately, I am not. If there are aliens, this strengthens my view that phenomenality and experience (intention and action) should be segregated: these notions are supposed by internalists to have at least similar ranges of application, but they turn out not to.

C. Ignorance and error

God's doxastic position mirrors the world. She is omniscient, and infallible:

- *Omniscience:* $P \supset BP$;
- *Infallibility:* $BP \supset P$.

These don't need to go together. One might be omniscient but fallible if one has some contradictions in one's beliefs. Or one might be infallible but not omniscient by having no opinion at all (at one extreme). (The notion of omniscience suggests *all-knowingness*; I am thinking of a weaker notion of *all-believingness*. We could imagine a stronger omniscience principle concerning knowledge as well. The strengthening of infallibility is trivial, because knowledge entails truth.)

We can imagine restrictions on these principles, by taking B to concern certain subjects at times by epistemic methods, or to have modal force; or by restricting the scope of P to certain subjects-matter.

Happily, we differ from God, being subject (at least considering things unrestrictedly) to:

- *Ignorance*: $P \wedge \neg BP$;
- *Error*: $\neg BP \wedge P$.

(Strictly speaking, “ignorance” is lack of *knowledge*, and is thus compatible with the absence of ignorance in my technical sense.)

Substantive inability to produce the requisite DAJ could be due to either inability to overcome ignorance or to avoid error (ignoring cases where knowledge is unavailable despite production of the relevant true belief).

The “presupposition” behind an ABC is the clear place to put pressure. When the presupposition is expansive, it can leave the content of the concept exceedingly indeterminate. When it is sharp, there is a risk of error. Which one shall we choose, to build into our theory? And what gives the content of the presupposition? It would be nice if we could tie the presupposition to the good case, somehow: that way, our views of two narrow things—intention and phenomenal character—would flow in parallel. We would find in each context that intention and phenomenal character co-supervene. If we detected freedom, this would undermine the doctrine that $E = A$.

Working hypothesis: the characterization of the good case should entail the presupposition.

1. Ignorance

Actual ignorance of experience is commonplace: experience is too complex to conceptualize in most ordinary cases.

What we need here is something like *necessary* ignorance, or at least “medically” necessary ignorance (in the sense of ignorance that is practically irremediable in the context of ascribing intentions or phenomenal sameness). Chalmers’s story acknowledges the possibility of such necessary ignorance: a change in a phenomenal property that is too “fleeting” for us

to be in a position to ascribe a predicate which singles it out determinately; a “background” phenomenal property which we lack a relevant body of contrasting cases to have formed abilities to do so.

So consider the following hypothesis: the presupposition is always weaker than the characterization of the good case. So weak, in fact, as to permit extensive indeterminacy in the sense of an ABC. Accordingly, the referent is always something disjunctive: taking a short walk up G street or taking a long walk up B street or “walking in a dream” or ...

The thought then is that we figure out what someone intends to do by inference, by reducing the indeterminacy in the direct judgement by adding in the remainder of the content of the description of the good case.

The difficulty here is that, if all that is immediately presented to the first-person perspective is the highly disjunctive thing, it is hard to sustain the idea that *experience* is broad.

Worse, two guys could be phenomenally alike but differ in their intentions (once the remainder of the content of the good case returns). That would snap the connection between experience and action.

We should keep looking.

2. Error

We can resolve this difficulty by making the presupposition identical to the characterization of the good case. Suppose that we build in the full description of the good case into the presupposition: then the direct judgement will not be indeterminate about whether Fred is on G street, but will in fact entail that. Then the presupposition will be false when Fred is on B street, and the ABC will be misfire: it will refer to the property of taking a short walk on G street.

But how do we go about making this error *medically necessary*?

Back when I used to be a direct realist, I thought that among the kinds of experiences were properties like *seeing something red*, which can be mimicked but not duplicated by dreams. The name of the game was to make it so that Mary and Vat-Mary both refer with their acquaintance-based concepts to *seeing something red*. Strategy: it is medically necessary

for us to assume that perception is veridical when introspecting perceptual states; we don't know how to do it without making this assumption. (Hume: "men are carried, by a natural instinct or prepossession, to repose faith in their senses; and that, without any reasoning, or even almost before the use of reason, we always suppose an external universe, which depends not on our perception [...]. Even the animal creation are governed by a like opinion, and preserve this belief of external objects, in all their thoughts, designs, and actions.")

But this doesn't carry over well into the current case. Suppose that the short walk was up B street. We certainly shouldn't inflate the making of some assumption about G street into a medical necessity!

Of course, this is not a problem if we just identify the presupposition with the description of the good case.

But we don't want to require making *some particular level of detail* about the good case medically necessary either. I should be able to form ABCs that are more abstract. Why can't I form an ABC of the disjunctive property?

3. A sliding scale approach

Solution is obvious: assure that the presupposition equals the characterization of the good case, but let the stringency of both of these float. If we float the stringency down, we reduce the risk of error at the cost of less specific answers about the nature of the experience; if we float it up, we get more specific answers at the cost of risk of error.

Although we are under no medically necessary compulsion to fall into error when reflecting on Fred's case (neither is he: if he had different beliefs about what street he was on, he would not), the alternative is indeterminacy. But determinacy about some connection with the external world is also an option, so we can *sometimes* find the external world present to the first-person perspective.

Well we *can*, but maybe this always requires some epistemic screwup. Maybe the vaguer presupposition is somehow better. But I don't see why that would be. Risk-aversion is a pragmatic matter. Just so long as the external world *can legitimately* be present to the first-person perspective, we save externalism.