

## Anti-individualism and phenomenal content

Byrne, Darragh

DOI:

[10.1007/s10670-019-00179-2](https://doi.org/10.1007/s10670-019-00179-2)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Byrne, D 2020, 'Anti-individualism and phenomenal content', *Erkenntnis*. <https://doi.org/10.1007/s10670-019-00179-2>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.



# Anti-individualism and Phenomenal Content

Darragh Byrne<sup>1</sup> 

Received: 10 May 2018 / Accepted: 3 October 2019  
© The Author(s) 2020

## Abstract

The paper addresses a *prima facie* tension between two popular views about concepts. The first is the doctrine that some concepts are constitutively perceptual/experiential, so that they can be possessed only by suitably experienced subjects. This is a classic empiricist theme, but its most conspicuous recent appearance is in literature on *phenomenal concepts*. The second view is anti-individualism: here, the view that concept possession depends not only on a thinker's internal states and relations to the concepts' referents, but also on certain of her relations to sociolinguistic peers. In recent works, Derek Ball and Michael Tye have in effect argued that the doctrines are incompatible, and their conclusion is that no concepts depend on experience. In reply, Bénédicte Veillet endorses those authors' incompatibilism, but argues that it is anti-individualism (about the concepts at issue) that we should reject. I develop an approach to reconciliation that is more promising than any considered by these theorists. Against Veillet, I defend a version of anti-individualism about phenomenal concepts, but against Ball and Tye, I argue that they can be possessed only by suitably experienced thinkers.

My project is to resolve a *prima facie* tension between two popular views about concepts. The first is that some concepts are constitutively perceptual or experiential, in a sense that implies that they can be possessed only by suitably experienced subjects. This is a classic empiricist doctrine, but its most conspicuous recent appearance is in literature on *phenomenal concepts*. The second view is *anti-individualism*, which says (roughly) that the concepts a thinker possesses can depend not only on her internal states and her relations to the concepts' referents, but also on certain of her relations to sociolinguistic peers. Ball (2009) and Tye (2009) argue, in effect, that the doctrines are incompatible and they conclude that no concepts depend on experience. In reply, Veillet (2012) endorses Ball and Tye's incompatibilism but argues that it is anti-individualism (about the concepts at issue) that we should

---

✉ Darragh Byrne  
d.byrne@bham.ac.uk

<sup>1</sup> University of Birmingham, Birmingham, UK

reject. In this paper I defend an approach to reconciliation that is more promising than those considered by these authors.

In Sects. 1 and 2 I elaborate the two doctrines in turn, and in Sect. 3 I examine the debate between Ball/Tye and Veillet. In Sect. 4 I develop my approach to reconciliation, and finally in Sect. 5 I work through the consequences of this. Against Veillet I defend anti-individualism about phenomenal concepts, but against Ball and Tye I argue that they are only possessed by suitably experienced thinkers.

## 1 The Experience-Dependency of Phenomenal Concepts

Locke and Hume held that certain ideas of colours and sounds cannot be entertained by the congenitally blind and deaf.<sup>1</sup> But in contemporary philosophy of mind, the most conspicuous advocates of the doctrine that certain concepts are *experience-dependent* are found in the literature on *phenomenal concepts* (PCs). According to ‘PC-theorists’—these are concepts which pick out (types of) phenomenal qualities from the first-person, introspective perspective, and so seem to afford a distinctively *direct* way of thinking about those qualities. Some authors construe this in epistemic terms, holding that possession of PCs involves *acquaintance* with phenomenal qualities, and that this may be difficult to reconcile with physicalism.<sup>2</sup> On the other hand, many PC-theorists maintain that the directness at issue is not epistemic but merely semantic, and indeed, that reflection on PCs’ distinctively direct semantics provides resources sufficient to rebut influential anti-physicalist arguments.<sup>3</sup>

To understand the arguments to come below, we need a sense of how this conception of direct reference relates to theories of concepts in general. For illustrative purposes let’s assume that these theories fall into two kinds. According to the first kind of theory—*covariational* theory, concepts are states/properties (of persons, or their brains) whose *tokenings* or *activations* co-vary with the perceptual or linguistic salience to their possessors of the concepts’ referents (or instantiations of their referents/elements of their extensions). Of course, not all correspondences are taken to be constitutive of concept possession: advocates of simple covariational theories identify the relevant ones as those which occur in *normal* or *optimal* conditions, while more sophisticated theories incorporate alternative ways to designate the canonical ones.<sup>4</sup> Meanwhile according to the second—*inferential/conceptual role* theory—concepts are individuated by their positions in the networks of inferences which their possessors are disposed to make—again, in conditions that are *normal*

<sup>1</sup> See e.g. Locke (1975, §II.2, ‘Of simple ideas’: pp. 99–100); Hume (2008, §II, ‘Of the origin of ideas’: p. 98).

<sup>2</sup> This claim is made explicitly by Levene (2007, p. 162–165), Nida-Rümelin (2007, p. 307) and Goff (2011), but it is, arguably, implicit in much influential anti-physicalist work, e.g. Jackson (1986) and Chalmers (1996).

<sup>3</sup> Prominent advocates of this physicalist ‘PC strategy’ include Loar (1990/97, 2003), Sturgeon (1994), Papineau (1998, 2002, 2007), Perry (2001), Lycan (1996), Balog (2012) and, in work that predates the antithetical line considered below, Tye (1999, 2003).

<sup>4</sup> See e.g. Dretske (1981) and Fodor (1980/90, 1998). Versions of this approach are sometimes labelled *indicator* or *informational* semantics.

or *optimal* or designated as canonical in some other way.<sup>5</sup> Although there are obvious differences between these two kinds of theories of concepts, many theorists appropriate elements of both.

PC-theorists characterize the distinctive semantics they claim PCs to exhibit by contrasting them with concepts whose reference is mediated by (contingently exhibited) properties of the concepts' referents/extensions. E.g. we think of water as *wet*, *transparent*, *thirst-quenching* etc.—and even if these properties don't constitute the meaning of the term 'water' or the content of the associated concept, WATER, they serve to fix the referent/extension<sup>6</sup> of those representations, so they play an important semantic role. In contrast, it's claimed, by exploiting the introspective perspective, PCs of what it's like to see colours, suffer pains, etc. refer to those phenomenal qualities in a way that's not mediated by reference-fixing descriptive information. A consequence of this emphasised by the physicalist PC-theorists is that propositions incorporating these concepts are not a priori connected to propositions couched in physical/scientific terms (or at any rate, are not so connected in virtue of the PCs they contain). And according to these philosophers, this suffices to explain away the epistemic data exploited by anti-physicalist arguments (e.g. that it's conceivable that there are zombies, and that science will never explain why experiences exhibit their phenomenal qualities).<sup>7</sup>

The contentions that PCs refer directly and are conceptually isolated make clear that PC-theorists endorse a version of covariational theory of such concepts, rather than an inferential role one. On straightforward versions, the direct relation between a PC and its referent is assumed to be causal. But on other accounts, it is held to be one of incorporation/constitution.<sup>8</sup>

So far in this section I've not mentioned the contention that a subject's possession of a PC depends on her having had a relevant experience. Indeed, I'm not sure whether this is an *essential* component of PC-theory, or, if it is, precisely how strong we should take the modal notion of dependence it incorporates to be. But in some form at least, it's an extremely natural and attractive assumption for a PC-theorist to make, and some of the more detailed accounts of PCs offered by PC-theorists rely on it.<sup>9</sup> PC-theorists propose that typical thinkers acquire their PCs through acts of introspective attention and that these introspective acts fix the PCs' references. Many stop short of an insistence that possession of a PC without relevant experience is *impossible*—e.g. Tye (1999, p. 712) concedes that a subject might acquire a PC through brain surgery or a miracle, and Papineau (2002, §2.8) suggests

<sup>5</sup> See e.g. Harman (1973, 1987), Peacocke (1992), Brandom (1994) and Greenberg and Harman (2006).

<sup>6</sup> Words in small capitals denote concepts. For brevity I'll generally omit the word 'extension' henceforth, and stick to 'referent'. I don't mean this to imply any substantial assumptions about the semantic values of predicate expressions and the concepts they express.

<sup>7</sup> For a defence of the view that PCs are in this sense 'topic-neutral' against a recent attack, see my (2016).

<sup>8</sup> Tye (1999, 2003) is a good example of the causal approach. Papineau (1998, 2002, 2007) and Balog (2012) develop constitutional ones. Elements of both are anticipated in Loar's pioneering (1990/97).

<sup>9</sup> E.g. Loar (1990/97, 2003) construes them as 'recognition', where recognition is understood in terms of perception/introspection of relevant instances of the concepts' referents.

that experience-dependency is ‘a quite contingent matter’—but these concessions involve exotic cases: caveats to the doctrine that all *normal* thinkers with PCs have had relevant experiences. Moreover, although the claim that PCs refer directly is the one that plays the most important role in physicalist PC-theorists’ responses to anti-physicalist arguments, the experience-dependency claim features prominently in their response to one of them—Jackson’s knowledge argument. And the scenario envisaged in the course of that argument is often invoked to illustrate claims about PCs. The physicalists concede to Jackson that while confined to her monochrome room, Mary cannot know what it’s like to see red, but they contend that this is not because what it’s like is non-physical. They propose, rather, that it’s because—not having had the relevant experience—she lacks a requisite experience-dependent PC.<sup>10</sup>

Although PC-theorists suppose that PCs are acquired through acts of introspective attention to token experiences, they assume that the concepts’ referents are phenomenal types. They assume, moreover, that the PCs survive the experiences’ cessation: coffee drinkers can think in the relevant way about what it’s like to taste coffee long after the cup is empty. Furthermore, although they assume that PCs are acquired through episodes of demonstrative attention, most theorists agree that in general, referents of PCs do not vary with context of use, and so are not demonstrative/indexical concepts.

It deserves emphasis that the claim here is that because PCs refer directly, they are not equivalent to *any* descriptive concepts. E.g., according to PC-theorists, the PC of phenomenal red that Mary acquires on her release is importantly distinct even from the co-referring descriptive concepts she might express with phrases like, ‘what it’s like to see red’ and ‘the property that others call “phenomenal red”’. Before her first visual colour experience, Mary knew plenty about colours, and used the word ‘red’ to express a colour concept. She also knew what it was like to see shades of grey, and what myriad other experiences were like, so she had the general concept PHENOMENAL EXPERIENCE. Hence, before her first visual colour experience she might well have used a phrase like ‘what it’s like to see red’, to express a descriptive content which PC-theorists insist is distinct from the co-referring PC she acquires later. I’m not sure whether PC-theorists would say that the concept pre-release Mary expressed by the word ‘red’ is connected to descriptive information in the way they suggest concepts like WATER are; or indeed, whether it’s the same as the concept that normally-experienced thinkers express with ‘red’, but irrespective of these issues, PC-theorists certainly maintain that this colour concept is distinct from the PC of phenomenal red which she acquires on her release.

A final preliminary matter which is relevant to my argument below concerns whether, if PCs are non-indexical/demonstrative and non-descriptive, we should presume them to be expressed in natural language, and if so, by what words? Mary is often represented as reporting her transformative experience with an exclamation such as, ‘So *that*’s what it’s like to see red!’, but should PC-theorists suppose that

<sup>10</sup> Ball (2009) argues that the experience-dependency claim plays a crucial role in the knowledge argument as well as in the physicalist rejoinder.

any of the terms in that sentence express her newly-acquired PC? We have already seen that they ought not to assume that her phrase, ‘what it’s like to see red’ does, as it is descriptive, and expresses a content Mary might well have entertained before her release from the room. Next, the PC-theorist ought not to think that her, ‘red’ expresses the PC either, since red is a property of external objects, not of experiences, and moreover, seems also to express a concept she possessed before release from the room. Perhaps the most plausible suggestion is that her ‘*that*’ expresses the PC: but if this is correct, it involves a non-standard use of the term, since ‘that’ is more usually a demonstrative/indexical expression whose referent varies with context of use.<sup>11</sup> Similar reflections seem true of words and phrases like ‘pain’, ‘what it’s like to suffer pain’, etc.

At this point, the PC-theorist might suggest that even if the concepts expressed by words like ‘red’ and ‘pain’ are not exclusively phenomenal, they may have phenomenal components, at least when used by ordinarily-experienced subjects (as opposed to Mary and her ilk). Perhaps the concepts RED and PAIN are complex, with functional or descriptive components (so that e.g. (to illustrate) they might be related, respectively, to COLOUR and UNPLEASANT) but also phenomenal components. This possibility suggests a refinement of our focus: if concepts like RED and PAIN are complexes with phenomenal components, then what we’re really interested in are those components, and the PC-theorist can be expected to hive these off for separate discussion. Following Chalmers (2003, p. 225), let’s use the label ‘pure PC’s for these putative non-descriptive, non-indexical conceptual units which pick out phenomenal qualities ‘directly, in terms of [their] intrinsic phenomenal nature’ and whose references are fixed through acts of introspective attention to phenomenal qualities.

Should PC-theorists suppose that there are words that exclusively express pure PCs, as opposed to complexes with (perhaps) phenomenal components? A negative answer has some independent appeal. When trying to explain to friends, students etc. what we mean by ‘phenomenal qualities’, many of us struggle to find the right words. One finds oneself resorting to emphasized (and, I’ve suggested, non-standard uses of) demonstratives (‘I mean pain’s feeling like *this*’ or ‘what Mary expresses with “*That* is what it’s like to see red”’).

Although I think this is an important suggestion<sup>12</sup> it’s hard to resist an obvious rejoinder. Even if we don’t already have words for pure PCs, subjects with such concepts could simply introduce neologisms for them by stipulation. It is difficult to deny that this is possible, since we have the terms we need to articulate such stipulations. Thus, a thinker could fix the referent of a neologism, ‘red\*’, with the

<sup>11</sup> Note that if one can express a PC with a non-standard use of ‘that’, one might also do it with a *referential use* of the description, ‘what it’s like to see red’: but this use would also be non-standard. The point remains that the expressed *concept* is not descriptive.

<sup>12</sup> The idea that phenomenal contents are not linguistically expressed is not new: Alex Byrne (2002) and Benj Hellie (2004) do not use the terminology of ‘PCs’ etc., but in effect, both respond to the knowledge argument by arguing that what Mary learns is *ineffable*. Block (1996) is also sympathetic. (Byrne (2002) provides useful references to earlier advocates.) Notice that even if PCs are not linguistically expressed, they can still be held to play many of the roles in thinking that PC-theorists assign them. The problems about consciousness which preoccupy PC-theorists concern thought, not language.

resolution expressed by ‘henceforth I’ll use “red\*” to express *this* (phenomenal type)’—while deploying his pure PC. And his subsequent uses of ‘red\*’ might then express that PC, rather than any descriptive or indexical concept.<sup>13</sup>

## 2 Anti-individualism

One feature that many covariational theories of concepts share with many inferential role theories is a commitment to *individualism*: a characterization of a subject’s concepts in terms of features of *her*—albeit features that may include her relations to the concepts’ referents. In contrast, *anti-individualism* (a.k.a. *social externalism*) proposes that thinkers who fail to satisfy some of the conditions in terms of which the covariation and inferential role theories identify concepts can qualify nonetheless as possessing them in virtue of certain relations to others in their sociolinguistic communities.<sup>14</sup>

The most influential arguments for anti-individualism issue from Tyler Burge’s (1979) famous story of Al, the mostly-competent user of the term ‘arthritis’ who sincerely asserts to his doctor, ‘I have arthritis in my thigh’. When Al’s doctor corrects him by explaining that arthritis is a condition one can only have in joints, he accepts the correction—accepts on the basis of the doctor’s authority that he misused the term ‘arthritis’. There are two ways in which we might interpret Al’s thinking. First, we could suggest that before being corrected by the doctor, he used the term ‘arthritis’ to express an idiosyncratic concept—THARTHRTIS—which picks out a (possible but non-actual) medical condition which is like arthritis except that one can have it in the thigh as well as in joints. On this interpretation, Al expressed a true belief when he first spoke to the doctor, but then modified/replaced his idiosyncratic concept on the basis of the doctor’s correction. The second interpretation is that the belief Al expressed before the doctor corrected him was the false one that he had arthritis in his thigh, and the doctor’s intervention led him to revise this belief, but not to modify a concept. This second interpretation commits us to an attribution to pre-correction Al of the ordinary concept, ARTHRITIS, notwithstanding his misconception of its referent.

Individualist theories (of both the covariational and inferential role kinds) comport with the first interpretation of the Al story, and not with the second. But Burge and others argue that the second interpretation fits our ordinary intuitions

<sup>13</sup> The suggestion that subjects might introduce terms for PCs by stipulation is made by Papineau (2002, §4.11) and also by Ball (2009, p. 950) and Tye (2009, p. 69). (It resonates mischievously with Wittgenstein’s (1953, §258) sceptical discussion of the man who undertakes to record instances of a sensation type by writing ‘S’ in his diary whenever tokens of it occur. In the literature on Wittgenstein’s infamous Private Language Argument, Edward Craig (1982) defends a version of the stipulation suggestion.).

In my (MS) I explore a position more sceptical of linguistic expression. To offset the appearance that we express phenomenal contents with terms like ‘*this*’ in the examples above, I consider the possibility that such contents might be ‘conveyed’ rather than expressed, in a manner in some ways akin to Gricean implicature.

<sup>14</sup> Thus, as I’ll emphasize below, the variety of semantic externalism in focus here is the sociolinguistic one associated especially with Burge, and not the distinct variety promoted by Putnam, Kripke, etc.



and attributive practice better than the first. They argue, moreover, that the way in which the first interpretation involves two concepts (ARTHRITIS and THARTHRTIS) instead of one generates implausible consequences about mundane beliefs, such as are expressed by sentences which AI does not misuse, e.g. 'Arthritis is painful'. Individualism implies that either (1) before the doctor's intervention, AI did not have the belief which the doctor expresses with that sentence (because he lacked a requisite concept); or (2) the doctor uses the sentence (at different times) to express two distinct beliefs—the one he shared with pre-correction AI, and the one he shares with post-correction AI (and many others). Anti-individualists urge that neither (1) nor (2) is plausible, so we should favour the one-concept interpretation.

We can formulate the anti-individualist position in the terms used above by saying that before the conversation with his doctor, AI possessed the concept ARTHRITIS even though he did not fully satisfy the covariational or inferentialist conditions in whose terms individualist theories characterize concept possession. Burge's presentation suggests that the relevant conditions AI fails to satisfy are inferentialist ones: it's assumed that his assent to the sentence, 'I have arthritis in my thigh' is evidence that some of the inferences AI is disposed to make between (propositions containing) ARTHRITIS and THIGH are at odds with those in terms of which inferentialist individualist theory would characterize possession of ARTHRITIS. As I'll illustrate in Sects. 4 and 5 below, we can also formulate versions of Burge's thought experiment which engage more directly with covariational individualist theories by involving relations between concepts/terms and their referents rather than inferential relations between concepts/terms.

The above is a rather schematic illustration of anti-individualism, but it's enough to indicate the way in which its advocates characterize concept possession in *interpersonal* terms. The key point is that AI is disposed to *defer* to his doctor over use of the term 'arthritis', because he acknowledges that the doctor's conception of arthritis is more accurate or complete than his. It's a good (and under-explored) question how exactly this mechanism of interpersonal deference is supposed to work.<sup>15</sup> For us, an important point to emphasize is that the proposal is *not* that what AI expresses with the term 'arthritis' is a metalinguistic content such as he might express by 'the disease my doctor calls "arthritis"'. Rather, the claim is that, in some implicit/automatic way, deference affords AI a way to piggyback on his doctor's superior conception.

On the other hand, anti-individualists don't generally think that a disposition to defer over the use of a term is *sufficient* for concept possession. Burge emphasises that AI is a generally competent user of 'arthritis'—i.e. most of his uses align with others—and I assume that the significance of this, in the terms introduced above, is as evidence that he satisfies some (perhaps most) of the conditions in terms of which individualist theories characterize concept possession.<sup>16</sup> Many philosophers

<sup>15</sup> See Greenberg (2014) for a plausibly pessimistic but fertile discussion.

<sup>16</sup> Burge (1979, p. 91) notes that we should not attribute the concept ORANG-UTAN to a subject who believed that 'orang-utan' picks out a kind of drink. Moreover, if the only occurrences of 'arthritis' that ever occurred in sentences asserted by AI were opaque occurrences in sentences attributing attitudes about the disease to the doctor, our inclination might well be to attribute to AI the metalinguistic content suggested above rather than ARTHRITIS. (For related discussion, see Sect. 5 below.).



who are broadly sympathetic to covariational or inferential role theories accept that deference can also play a role. Thus, just as many theorists endorse elements of both covariational and inferential role theories of concepts, many more seem to endorse accounts that combine elements of both covariational/inferential role theories and anti-individualism. It's a good (and, again, under-explored<sup>17</sup>) question just how this compromise is supposed to work. I assume the rough idea (and one that will suffice for my purposes here) is that to possess a concept, a thinker has to meet (or come close to meeting) *some* of the conditions associated with individualist approaches, but that shortcomings in their satisfaction of these can be compensated for by suitable inclinations to defer. E.g. it seems plausible that to qualify as possessing ARTHRITIS, Al needs to be disposed to endorse *some* pertinent inferences, (e.g. from ARTHRITIS to DISEASE) and/or to enjoy some relevant causal relations to instances of arthritis, even if the misconception evidenced by his assent to, 'I have arthritis in my thigh' is off-set by his deference to a medical expert.

Two further qualifications are in order before we return to the phenomenal case. First, it's worth pointing out that even if our intuitions accord with Burge's over familiar examples like the one just rehearsed, the case they support is at best an inference-to-the-best-explanation of the data codified in those intuitions: it was never going to be a conclusive knock-down argument for anti-individualism about relevant concepts.<sup>18</sup> (I return to the question of how to evaluate Burgean arguments in Sect. 4.)

Second, notice that Burgean anti-individualism is quite different from the variety of semantic externalism about the meanings of terms like 'water' defended by Putnam (1975) and Kripke (1980)—and the differences matter here. On the Putnam/Kripke view, the environmental features on which linguistic meanings (and, let's assume, conceptual contents) depend are natural (usually chemical/structural) properties of the terms' referents; whereas on the anti-individualist picture, the relevant environmental features are sociolinguistic. The conception of direct reference promoted by PC-theorists certainly suggests a version of Putnam/Kripke externalism about those concepts,<sup>19</sup> but whether PC-theorists should also endorse Burgean anti-individualism about them is moot.

### 3 PCs Without Experience?

In the previous section, I explained the anti-individualist proposal that thinkers whose conceptions of concepts' referents are inaccurate or incomplete can qualify as possessing the concepts in virtue of their inclinations to *defer* to peers whom they recognize to be more authoritative. Ball (2009) and Tye (2009) argue that the same

<sup>17</sup> Though again, see Greenberg (2014).

<sup>18</sup> Ball (2009, p. 954) appears to agree: 'Burge's arguments do not show that his opponent's view is *incoherent*, but that it is *implausible*.'

<sup>19</sup> That is, it suggests that the contents at issue depend on their referents, though perhaps in this case not that they 'ain't in the head'!

is true of introspective concepts of phenomenal qualities. Moreover, they maintain, if a thinker can possess those concepts in the deferential way, then she can possess them without having had relevant experiences, so such concepts are not experience-dependent. They defend their contention dramatically in terms of Jackson's Mary. After Mary is released from the room, she can converse with her peers about phenomenal red. All parties agree that whatever words and concepts the others use to do this, post-release Mary can use too. But now, surely she could have learned the *word* that expresses the relevant concept—and so come deferentially to possess that concept—*before her release*. The PC-theorist must deny this, holding that before her release she possessed no more than a non-phenomenal, impoverished version of the concept (or a different concept). But Ball and Tye argue that this is as implausible as the individualist's contention that before being corrected by his doctor, Al possessed the idiosyncratic concept, THARTHRTIS.

Veillet (2012) has responded to the Ball/Tye argument, defending the individualist, two-concept interpretation of Mary's case against their Burgean critique.<sup>20</sup> In this section I examine this debate. As we'll see, Veillet's response to Ball and Tye is effective: their argument against the two-concept interpretation of Mary is unpersuasive. However, Veillet provides no reason to favour an individualist interpretation of the concepts at issue: so the question whether thinkers can possess these concepts in the deferential way remains open. In Sect. 4 I develop a more plausible version of anti-individualism about these concepts than the one considered by Ball and Tye, but in Sect. 5 I argue that even on this conception, such concepts are experience-dependent. (So the position I defend is distinct from both of those championed by these authors.)

A complication that makes this a little difficult to excavate fairly is that to avoid begging the question against Ball and Tye, we had better not call the controversial concepts 'PCs'. As I explained in Sect. 1, the thesis that PCs are experience-dependent may not be essential to PC-theory, but it's widely endorsed by PC-theorists, and by attacking it, Ball and Tye take themselves to be arguing that (to quote the title of Ball (2009)) 'There are no Phenomenal Concepts'. On the other hand Ball and Tye don't deny that thinkers have introspective concepts of phenomenal properties<sup>21</sup>—their contention is simply that such concepts are not experience-dependent. So we need a neutral term for concepts which pick out phenomenal qualities from

<sup>20</sup> Alter (2013) has responded to Ball and Tye's argument in a different way. He argues that while it may be conceded that pre-release Mary possesses the relevant PC, she does not at that time *master* it, and moreover, that it is mastery and not possession that matters to the arguments pertaining to the metaphysics of consciousness. Ball (2013) has replied, arguing, *inter alia*, that even if mere possession is in some sense thinner than full mastery, pre-release Mary's mere possession of the relevant concept suffices for her to know what it's like to see red.

<sup>21</sup> E.g. Tye (2009, p. 56) characterises a kind of concept with whose existence he has no quibble as 'those concepts we use in thoughts or judgements formed on the basis of introspection of phenomenal character of our phenomenal states'. And in the course of Ball's defence of his view that Mary could possess a concept of the relevant kind by deference, he considers what she expresses, upon her release, with the sentence, 'That is what it's like to see red', and explains that in that sentence, "That" refers to an experience of red, to which Mary is attending in introspection, or to some feature of such an experience' (2009, p. 947).

the introspective perspective but are not *defined* as experience-dependent. From here on I'll use the term 'introspective concepts' (IC), leaving it implicit that the referents in the cases of interest here are phenomenal. I'll also assume that the ICs at issue are 'pure' in the sense introduced in Sect. 1—i.e. that their referents are *exclusively* phenomenal. (My usage here implies a concession to Tye and Ball that a concept could be introspective without being experience-dependent; and one might worry that this begs the question not just against the doctrine that ICs are experience-dependent, but also against any conception that regards such concepts as essentially tied to perception. But as long as anti-individualism remains in the running this can be avoided. For if a thinker can possess an IC in the deferential way, then even if its possession does not require *her* to have had introspective contact with its referent (so that it is not experience-dependent in the sense characterized in Sect. 1) the concept may yet be experience-dependent in the weaker sense that it requires (either her or) some of the thinkers to whom she defers (or the ones to whom they defer, etc.) to have had such contact (so that the concept might still be regarded as essentially perceptual.)

Another preliminary question concerns the *terms* that are used to express these concepts. Ball and Tye run their arguments using terms like 'red', and 'what it's like to experience red'; but for the reasons I suggested in Sect. 1, this is misleading: neither of those terms, at least when used in standard ways, is a plausible candidate for expression of a (pure) PC or an IC. An advocate of the view that such concepts are not expressed in natural languages at all would have an obvious rejoinder to the Ball/Tye argument, since the medium through which anti-individualists propose that concepts can be deferentially possessed is linguistic; but in response to this (and as I suggested in Sect. 1) Ball (2009, pp. 949–951) and Tye (2009, pp. 69–70) argue that a speaker might simply introduce a term for such a concept by stipulation. And they insist (plausibly) that their arguments would be equally effective if transposed into neologistic terms. In light of this I'll follow Veillet (2012, p. 102) and use a neologism: 'red\*'.

In these terms, then, we envisage Mary using the term 'red\*' before and after her release from the monochrome room in discussions with better-experienced peers. Orthodox PC-theorists have to hold that Mary expresses *two* concepts with this—an austere, non-phenomenal one which she utilizes before her release from the room, and the full-blooded PC that she comes to deploy after it. Let's ask which of these concepts Mary's more experienced interlocutors expressed in conversations with her before her release. Perhaps they didn't possess the austere non-phenomenal concept which Mary does, in which case their earlier discussions with Mary involved equivocation. Or perhaps they did possess the austere non-phenomenal concept as well as the richer PC, and they used 'red\*' to express each of them on different occasions. Tye and Ball assume that the PC-theorist will opt for the second possibility (and Veillet agrees) but they argue that the two-concept picture it delivers is implausible.

First, they maintain, ordinarily experienced subjects are not introspectively aware that they possess pairs of concepts of this sort, and that the associated words are ambiguous. Moreover, the picture has further implausible consequences. E.g., if we suppose that before release, Mary thought, 'I have not experienced red\*', and after, she thinks 'I have experienced red\*', advocates of the two-concept interpretation

must deny that these two thoughts contradict one another (since according to them, ‘red\*’ expresses distinct concepts in the two sentences)—and this is implausible. According to Ball and Tye, it’s better to suppose that only one concept is involved here. And since it seems that Mary possessed this concept before her first colour experience, it cannot be experience-dependent.

In response to this, Veillet (2012) defends the two-concept interpretation. To illustrate, she compares Mary to *Annie*, a novice astronomer who uses ‘Hesperus’ to express the concept *HESPERUS*, but who does not possess the concept, *PHOSPHORUS*. One day, Annie is introduced to the name, ‘Phosphorus’ by being told (only) that it refers to Hesperus, and she proceeds to use the name ‘Phosphorus’ in conversations with other astronomers. According to Veillet, the fact that Annie has become familiar with the term, ‘Phosphorus’ is not a reason to suppose that she now possesses the concept it is used by others to express—it’s more plausible, she suggests, to suppose that Annie uses it to express a concept she already has, e.g. *HESPERUS*.<sup>22</sup> And so, Veillet argues, the fact that Mary can use the term ‘red\*’ while in the room is not a definitive reason to suppose that she uses it at that time to express the same concept as her more experienced peers and her later self. A plausible alternative is that she uses it to express a co-referring non-phenomenal concept. Veillet’s label for this is ‘RED PHENOMENAL CHARACTER’: it is (I assume) what I have been describing as a *descriptive* concept of a phenomenal quality—something like the one that, I suggested in Sect. 1, is expressed by the phrase, ‘what it’s like to see red’.

This allows Veillet to respond to the considerations raised by Ball and Tye against the two-concept interpretation. Since we don’t need to interpret pre-release Mary as using ‘red\*’ to express the PC it’s normally used to express, it is not implausible to interpret her pre-release utterance of ‘I have not experienced red\*’ as consistent with her post-release declaration, ‘I have experienced red\*’. Veillet also replies to Ball and Tye’s argument that ordinarily-experienced thinkers are not introspectively aware of using terms like ‘red\*’ to express more than one concept. She argues that this is because, generally-speaking, they do not do so. Even if experienced subjects *possess* the austere concept that pre-release Mary expresses with ‘red\*’, they might never have reason to express it (with that term or any other).

## 4 A More Prosaic Case

Veillet’s careful analysis establishes that there’s nothing incoherent about the two-concept interpretation that orthodox PC-theorists need to give of the Mary scenario: so she provides a coherent way for individualists about ICs to rebut the Ball/Tye argument against PCs. On the other hand, the one-concept interpretation

<sup>22</sup> As will emerge, I agree with Veillet that we have little reason to suppose that Annie uses ‘Phosphorus’ to express *PHOSPHORUS*; but for what it’s worth, I don’t think it’s especially plausible that what she uses it to express is *HESPERUS*. Given the role that ‘Phosphorus’ is likely to have in her thinking, I think it’s much more plausible that it expresses something like ‘the planet that astronomers call “Hesperus”’. (For more on metalinguistic examples like this, see Sect. 5 below.).

championed by Ball and Tye is also coherent: Veillet offers no positive reasons to prefer the two-concept one. So the position we're left in is more of a stalemate than a defence of individualism about ICs.

Let's look a little closer at the dialectical structure of Burgean arguments for anti-individualism about concepts of a given kind. As I observed in Sect. 2, these are never 'knock down': they're a variety of inference-to-the-best-explanation. We're told a story about a thinker, and invited to agree that the most compelling interpretation of it involves a single concept she shares with others, as opposed to a pair of concepts (one shared and one idiosyncratic)—even though the thinker fails to satisfy some of the conditions in terms of which individualist theories characterize possession of the (shared) concept. Hence, evaluation of a Burgean argument boils down to a weighing of two things against one another. On one side we have the interpretative intuitions elicited by the story, and on the other is our commitment to the mooted individualist possession conditions.

As we saw in Sect. 2, we can think of Burge's famous argument as involving putative individualist possession conditions (for ARTHRITIS) that are inferential, and which seem not quite to be met by a thinker who assents to the sentence, 'I have arthritis in my thigh'. When we turn our attention to PCs (and to related concepts like RED) we shift our attention to *covariational* candidates for the role of relevant mooted individualist possession condition, for as we saw in Sect. 1, PC-theorists typically endorse covariation theories of these concepts rather than inferentialist ones. Thus, the relevant idiosyncrasy exhibited by Mary is that she is not related to instantiations of red\* in the (introspective) way that more typical thinkers are; and this is relevant because a typical individualist theory of RED\* would characterize its possession in terms of a thinker's relation to instances of red\*. More specifically, as we saw in Sect. 1, PC-theorists characterize possession PCs as essentially involving distinctively introspective causal—or perhaps even constitutive—relations to experiences exhibiting them, and the crucial point here is that Mary's lack of experience provides very strong evidence that she fails to meet one of these individualist conditions. The upshot here is that if it is to persuade us to repudiate orthodox PC-theory, the Burgean argument has to offer a story of which the one-concept interpretation elicits a stronger pull than the assumption that RED\* has this (experience-involving) individualist possession condition.

This suggests a way in which to resolve the stalemate in the PC-theorist's favour, in light of Veillet's discussion. The theory of PCs (some key elements of which I outlined in Sect. 1) is intuitively and theoretically appealing. If the pull it elicits is stronger than the inclination to adopt Ball and Tye's one-concept interpretation of the Mary story, we're justified in continuing to believe it. And by demonstrating the availability of a coherent two-concept interpretation, Veillet's discussion deflates the appeal of the one-concept one, to some degree at least.

One reservation about this is that the view defended by Veillet—combining as it does individualism about a small rarefied class of concepts and anti-individualism about all the rest—is a rather disunified, inelegant theory of intentionality. More importantly, I think we can do better than the qualified resolution just considered. Below I'll defend a version of the Burgean argument for anti-individualism about ICs that we have not yet considered. If this did not suggest itself in the previous

section, this is because our focus there was on such an *exotic* case. In this section I'll consider a subject whose predicament is much more prosaic than Mary's. However, the version of anti-individualism about ICs I'll defend is consistent with the contention that ICs are experience-dependent, and in Sect. 5 I'll defend that contention. So while the position I'll develop opposes Veillet's, it also opposes Ball and Tye's.

In the previous section I noted that Ball and Tye illustrate their main argument by envisaging Mary and her peers using terms like 'red', and 'what it's like to see red' instead of philosophers' neologisms like 'red\*'. I complained that this is misleading—since they're poor candidates for terms we might take to express ICs—and I followed Veillet in sanitizing the discussion by replacing those terms with the neologism. However, an important part of the standard argument for anti-individualism about any concept is the lemma that fairly competent speakers can misapply it, and then accept correction from others whom they recognize as more authoritative. (This is the part of the argument that suggests the thinker's failure to satisfy a mooted individualist possession condition.) Ball and Tye defend instances of this lemma in various ways: in this section I want to focus on one (on which they touch only briefly) which involves a more prosaic scenario than Mary's, but in which the use of words like 'red' is neither misleading nor dispensable.

Ball (2009, p. 951) and Tye (2009, pp. 63–65) point out that some speakers systematically misapply 'red' to objects which are not red and 'pain' to experiences that are not painful, and they reflect that many such speakers are happy to accept correction when these errors are pointed out to them, much as Al accepts correction from his doctor.<sup>23</sup> As I've emphasized, these terms don't express ICs (or at least they do not exclusively) but Ball and Tye argue that thinkers who are disposed to misapply these terms can be expected also to misapply the related concepts of phenomenal qualities: e.g., (in the terms introduced above:) thinkers who misapply RED can be expected also to misapply RED\*. And thinkers who accept correction and revise earlier applications of RED can be expected to do the same in respect of RED\*. Tye pushes this line forcibly:

Anyone who is willing to accept correction as to whether a given shade should really be counted as a shade of red should be willing to accept correction as to whether a given token experience she is undergoing while viewing that shade should properly be counted as having the phenomenal character of experiences as of red, assuming that she takes herself to be a normal perceiver in normal viewing conditions. (2009, p. 64)

He goes on to argue that a thinker who accepted the first correction but not the second would be 'deeply irrational'. Veillet (2012: 112) takes issue with this line of argument. She points out that it's one thing to concede that a thinker can mistakenly believe an experience to correspond to a certain colour, and quite another to concede that she can mistakenly believe it to correspond to a given type of phenomenal property. PC-theorists generally agree that a thinker could possess an IC like RED\*

<sup>23</sup> Burge (1979) also uses colour examples.

without knowing which colour it is a phenomenal instance of<sup>24</sup>; so they might well hold that a thinker could be prepared to defer over application of a colour concept without being prepared to defer over a concept like RED\*.

In a moment I'll argue that Veillet's interpretation of Ball and Tye's argument is uncharitable—but it's encouraged by carelessness in passages like the one from Tye I quoted above. (And as I said, Ball's and Tye's presentations of this argument are brief). There, Tye considers whether a thinker would 'accept correction' over whether an experience has '*the phenomenal character of experiences as of red*' (2009, p. 64, italics added) and it's easy to read that as a question about a descriptive concept which incorporates the colour concept, RED—e.g. as about the concept Veillet call 'RED PHENOMENAL CHARACTER' (which I've suggested is expressed by the phrase, 'what it's like to see red'). And Veillet is right to point out that PC-theorists can concede that thinkers might defer over that concept without conceding that they might defer over the likes of RED\*.

But two questions remain. First, might a thinker misapply RED\*? The Cartesian view that first-person phenomenal judgements are infallible has a venerable pedigree, but few contemporary advocates. There's a constitutive difference between undergoing a phenomenal experience and making a phenomenal judgement about it, so it should be possible for the two to come apart.<sup>25</sup> This invites my second question: could circumstances that led a thinker to believe that she had misapplied an IC and so to revise an earlier application count as circumstances involving *deference* of the sort relevant to anti-individualism?

To get clear on this, let's begin with RED, and take a closer look at the kind of case suggested by Ball's and Tye's remarks. Suppose that Alice is a normally-experienced perceiver who mistakenly describes a certain area of a painting as 'red'. Shortly afterwards, her description is challenged by Bob, who knows that the area in question was orange. Alice recognizes that Bob is a better colour taxonomizer than her, and so she accepts his correction. For familiar Burgean reasons, it seems natural to attribute a single concept, RED, to Alice both before and after correction—and so to accept that her possession of that concept is (at least in part) deferential. To get extra clear on this, let's put it in the explanatory terms I used in Sect. 2 and earlier in this section. It seems that Alice fails at least to some degree to satisfy a condition in terms of which an individualist would characterize possession of RED. As with the case of PCs this condition is a covariational (rather than inferential) one, and her idiosyncrasy is evidenced by her disposition to apply 'red' to an area of the painting which is not red. As I mentioned in Sect. 1, covariational theorists don't assume that all connections between a concept and the world are constitutive of possession of the concept—the relevant ones are those that arise in conditions that are *normal*

<sup>24</sup> This is illustrated by Nida-Rümelin's (1998) example of Marianna, whose predicament is just like Mary's except that when she is released from the room what she confronts is not an object whose colour she can anticipate, but an array of different-coloured splodges of paint. PC-theorists assume that by looking at a red splodge, she can acquire RED\* without knowing what colour it corresponds to.

<sup>25</sup> Chalmers (2003, pp. 234–254) reflects that judgements invoking 'direct PCs'—i.e. PCs formed and deployed in the immediate introspective presence of their referents—may be infallible. But my interest is in judgements invoking 'standing' ICs which are not in this way guaranteed to be correct.



or *optimal* or whatever—but let’s assume that such conditions obtain in this case so that Alice’s use is certainly at odds with the mooted possession condition. The anti-individualist claim, then, is that this shortcoming is off-set by her preparedness to defer. An individualist would insist that before being corrected, Alice used ‘red’ to express an idiosyncratic colour concept (one which includes some orange in its extension). But the anti-individualist emphasizes her preparedness to defer and the familiar considerations to do with mundane beliefs expressed using ‘red’ to argue that we should prefer an anti-individualist attribution to her of RED.

Suppose next that when Alice made her judgement about the painting’s colour, she also applied an IC (presumably, RED\*) in a judgement about what it was like to see the colour. I suggest that what Ball and Tye are driving at is the suggestion that if Alice’s application of RED to the painting was an error, then her application of the IC to the experience was probably an error too. Here, an individualist would characterize possession of RED\* in terms (inter alia, perhaps) of a certain introspective relation to red\* (again, modulo the usual qualification about *normal/optimal* conditions) and if Alice misapplied the IC, this is evidence that she failed to meet that condition. (Let’s assume that conditions were normal/optimal for application of RED\* just as we assumed they were for application of RED.)

Should we agree that Alice probably misapplied RED\*? I suggest that Alice has reason to suspect that she did. It’s likely that she believes that red\* experiences are typically caused by (or, indeed, *are*) perceptions of red objects, and so in general she’ll expect her applications of RED and RED\* to keep in step. Indeed, it’s likely that she’ll *endeavour* to keep them in step. If Bob convinces her to revise the earlier application of RED, a revision of her earlier application of RED\* would be licenced—indeed, prescribed—by this policy. Against this theoretical setting, it would seem (as Tye suggests) *irrational* for her to revise the earlier application of RED without also revising her earlier application of RED\*.

To be sure, we have not yet described a case of Alice *deferring to Bob over her application of RED\**—at least not directly. For in the scenario just considered, Alice’s reason for revising her application of RED\* is not a recognition that Bob has a better conception of RED\*’s referent (i.e. of red\*). Rather, her reason is the conjunctive one, that, first, she expects her applications of RED and RED\* to align, and second, her recognition of Bob’s authority in regard to the referent of RED.

It’s a good question whether we should take a thinker’s inclination to revise an earlier application of a concept for this kind of complex, indirect reason to support anti-individualism about the concept at issue. I think we probably should; but we can pass over that for now, for there is a more direct kind of scenario we can consider instead. Suppose that at the time Alice stipulatively introduces ‘red\*’ to express her IC of (the type) what it’s like (for her) to see red—let’s call this time ‘ $t_1$ ’—she tells Bob about this, and he undertakes to follow suit. I.e., (as the PC-theorist will have it) at  $t_1$  he undertakes henceforth to use ‘red\*’ to express his IC of (the type:) what it’s like (for him) to see red. Subsequently, they use the term ‘red\*’ in conversations with one another. The analysis of this I’ve offered suggests that there’s a *precariousness* to this. The PC-theorist will say that Alice’s post- $t_1$  uses of ‘red\*’ refer to the phenomenal quality she was introspectively attending to when she acquired her IC, and

that Bob's uses refer to the quality to which he was introspectively attending when he acquired his. For both of them this will have been a time sometime before  $t_1$ —let's assume it was the same time, and call it ' $t_0$ '. The PC-theorist can be expected to concede that for all that's been said, these phenomenal qualities might differ: different people might experience colours in different ways. Moreover, even if what it was like for Alice at  $t_0$  is the same as what it was like for Bob, their ICs might exhibit a different degree of specificity or one of them might not locate the border between red\* and orange\* exactly where the other does. On the other hand, one might reason optimistically that at the time they acquired their IC(s), what it was like for Alice may well have been much the same as what it was like for Bob, and that their concept-forming intentions, partitioning etc. were much the same. And in that case, one might suppose, their meanings would *align*: both would use 'red\*' to express the same concept. Presumably, since they bother to enter into their arrangement with 'red\*', Alice and Bob side with the optimists on this: they assume that their meanings align.

Now we can extend the art gallery story. Suppose Alice made her judgements at  $t_2$ , and that right after correcting her application of 'red', at  $t_3$ , Bob goes further, and mischievously suggests, 'I bet you described your colour experience back there as "red\*", and that you got that wrong too!' And Alice might be persuaded by this. She might reason to herself that since (as has already been granted) Bob is better than her at perceptually distinguishing red from orange, he's probably also better at introspectively distinguishing red\* from orange\*. Now, making this judgement would commit Alice to a certain *modesty* in regard to her application of an IC, but not, I'll argue, *irrationality*. Applications of ICs involve *memory*—in applying RED\* to a current experience, Alice in effect judges that her current experience is phenomenally like the one to which she was attending when she acquired RED\*—and thinkers recognize that their memories are fallible. In the scenario envisaged above, Alice comes at  $t_3$  in effect to believe that at  $t_2$ , Bob was better than her at judging whether the phenomenal character that he attended to at  $t_2$  was the same as the phenomenal character that she attended to at  $t_0$ . Now, since the earlier of these experiences was Alice's and the later was Bob's this may seem like an *interpersonal* phenomenal comparison of a kind that, it seems, no ordinary person could make. But remember: both Alice and Bob assume that their meanings align—that at  $t_0$ , what it was like for Alice was *the same* as what it was like for Bob. Assuming this alignment, Alice can reason that Bob's comparison of what it was like for him at  $t_2$  with what it was like for her at  $t_0$  is just a comparison of what it was like for him at  $t_2$  with what it was like for *him* at  $t_0$ . Thus if she accedes to him over the use of 'red\*', this is modest, insofar as it commits Alice to thinking of Bob as better than her at certain (cross-temporal, memory-involving) phenomenal comparisons; but it does not commit her to crediting him with a supernatural-looking talent for *interpersonal* phenomenal comparisons.<sup>26</sup>

<sup>26</sup> The doctrine that we cannot make sense even in principle of interpersonal phenomenal comparisons of a kind touched on here is the notorious 'Frege-Schlick view'. For discussion, see Shoemaker (1982, 2006, Stalnaker 2006), Stalnaker (1999, 2006) and Block (2007). I lack the space to go into this here, but obviously I'm committed to rejection of this strong thesis, even if in the latter part of this paragraph I suggest in effect that the idea of interpersonal phenomenal comparisons taken as a matter of course (as opposed to *in principle*) seems more problematic than that of cross-temporal intra-personal phenomenal comparisons.

Here we can also see that Alice's inclination to defer to Bob over 'red' and RED does not *compel* her to accept Bob's correction in regard to 'red\*'. The other way she could rationally react to his intervention is to repudiate her optimism about meaning alignment. If she's pretty sure that her  $t_2$  use of 'red\*' and application of RED\* were correct, she can reason that perhaps she and Bob don't use 'red\*' to pick out the same phenomenal quality—perhaps, e.g. at  $t_0$ , what it was like for him was different to what it was like for her, so they acquired different ICs—and on that basis she can decline to defer. She might have other reasons for pessimism over alignment. E.g. she may be impressed by familiar Cartesian intuitions about the *idiosyncrasy* of phenomenal experience. Even if she does not think it likely that e.g. Bob's experiences are inverted relative to hers, or that either of them is a zombie, she might, even so, think it unlikely that their phenomenal qualities at  $t_0$  were *the same*. Or she might have what she takes to be evidence relevant to this particular case. E.g. past experience with Bob might suggest to her that while he's good at discriminating colours, he's less good at consistent application of the policy to keep applications of colour concepts and phenomenal colour concepts in step.

So, would thinkers with words for ICs grasp them in the deferential way? The answer suggested by the reflections above is that they *could* do, but they also could *not*. It depends (a) on the degree to which they regard themselves as better (or worse) than others at (cross-temporal, memory-involving) phenomenal comparisons, and (b) the degree of confidence they have in the assumption that their relevant phenomenal experiences are/were like those of their peers. If the quantity (a) is higher than (b)—if thinkers are more confident that their introspective applications of ICs are more reliable than others' than they are confident that others have experiences of the same phenomenal kinds—then they won't defer. And in that case, the correct interpretation is the individualist one: they possess idiosyncratic ICs in a way that does not involve deference.<sup>27</sup>

On the other hand, if the quantity (a) is lower than (b)—if thinkers with words for ICs are less confident that their introspective IC-involving judgements are better than others' than they are confident that others have experiences of the same phenomenal kinds—then they'll be prepared, on occasion, to defer. And in that case, the familiar Burgean considerations are apt. Perhaps when Alice introduced the term 'red\*' at  $t_1$ , its reference was fixed to the phenomenal quality she was attending to when she formed the concept RED\*, and perhaps we should favour an individualist construal of the concept she expressed by her first uses thereafter of the term 'red\*'. But if she is disposed thereafter to defer to Bob over application of the term in something like the manner described above, the one-concept interpretation seems much more attractive, for all of the familiar reasons. An individualist would characterize her possession of RED\* in terms of a relation (in normal/optimal conditions) to instantiations of red\* and she now has reason to believe—so we now have reason

<sup>27</sup> On this view, the concepts are idiosyncratic because their identity depends only on how things are with the thinkers themselves, and because distinct thinkers *might* have distinct ICs. This version of individualism does not rule out the possibility that as a matter of (as it were, coincidental) fact, different thinkers' ICs are the same.

to believe—that at  $t_2$  she failed to satisfy that condition. As usual, an individualist interpretation, according to which the concept she possessed at  $t_2$  was idiosyncratic, is available. Indeed, the remark above about how her initial uses of ‘red\*’ were probably individualistic confirms the availability of a candidate for this role. So if we have reasons to insist that the concepts she and Bob express continue to be distinct, the individualist interpretation remains coherent, just as it did in the Mary case. But if Alice reasons in anything like the way envisaged above, it’s hard to see what those reasons could be.

## 5 Resolutions

In this final section I clarify and evaluate the ramifications of the version of anti-individualism about ICs developed above, and revisit the pivotal case of Mary. The position we’ve reached is that ICs can be possessed in the deferential way, and they can be possessed in the non-deferential way. This view is not disunified and inelegant, in the way I suggested at the beginning of the previous section that Veillet’s position seems to be: rather it’s a version of anti-individualism. Anti-individualists should (and I think do) assume that *any* concept can be possessed in the non-deferential way, if the possessor is resolute enough. Suppose that when Al is confronted by his doctor he digs in his heels and insists that what *he* meant by ‘arthritis’ was most definitely a condition one can have in thighs as well as joints. In that case, I take it, we’d opt for the individualist presumption that before the doctor’s intervention, the concept he expressed was THARTHRTIS and not ARTHRITIS. Burge’s contention is not that concept-possession cannot be individualistic. Rather, it’s that the anti-individualistic construal is the more plausible, given that thinkers don’t generally exhibit this kind of obstinacy—given that they’re generally prepared to defer.<sup>28</sup>

Perhaps thinkers defer less over ICs than over other concepts. This is what we might expect if—as I argued above—they are rarely expressed by ordinary terms of natural languages, since the mechanism of deference described by standard Burgean anti-individualism is linguistic. But I’ve argued that even if they were only expressed by specially-introduced neologisms, thinkers like Alice might well be prepared to defer—and that if there are other cases in which they are not, this will be because of their assessments of the quantities (a) and (b) rather than anything to do with the special nature of ICs. Since anti-individualism is the view that concepts can

<sup>28</sup> Chalmers (2012, Ch. 6, §9) argues that any term can be used in both deferential and non-deferential ways. An anonymous referee for *Erkenntnis* suggests that advocates of Chalmers’ view have a straightforward response to Ball and Tye: even if the relevant terms can be used deferentially, they can also be used non-deferentially. We can see here why this would be ineffective as a reply to Ball and Tye. To begin with, the thesis they oppose, in the first instance, concerns the conditions necessary for *possession* of certain concepts rather than their linguistic expression. Of course the thesis that certain concepts are experience-dependent has a consequence involving linguistic expression—i.e. that relevant concepts can (assuming they can be expressed linguistically) be expressed non-deferentially. But as we’ve just seen, anti-individualism about these concepts does not dispute this. It’s not about whether it’s *possible* for someone to use ‘red\*’ non-deferentially, but rather whether, on balance, we should interpret their ordinary uses as non-deferential.

be possessed in the deferential way, not that they can *only* be so possessed, I submit that this constitutes a compelling argument for anti-individualism about ICs.

In Sect. 4 my attention was directed mainly at the rarefied case of the stipulated neologism, ‘red\*’, but there may be reasons to suspect that deferential possession of phenomenal contents is more widespread. Above I conceded that even if the concepts expressed by terms like ‘red’ are not pure ICs, they might be complexes with phenomenal components. And if they are, we might well suppose that when a thinker possesses such a concept in the deferential way, this gives him a deferential purchase on its phenomenal component. If there were good reasons to deny this—to suppose that only the non-phenomenal parts of such a complex content could be deferentially possessed—those reasons (or analogues of them) ought to have emerged in the course of our reflections on the neologistic case.

To conclude, I return to Mary’s case, for we now have the theoretical resources we need to explain why—even if ICs can be (and/or often are) possessed in the deferential way as anti-individualists maintain—it remains plausible to regard them as experience-dependent.

As emerged in Sect. 2, familiar versions of anti-individualism don’t licence the contention that an inclination to defer over the application of a term is *sufficient* for possession of the concept it is normally used to express. Burge emphasises that Al is a *mostly* competent user of ‘arthritis’—and I assume his reason for this is that if his use of the term was radically deviant—e.g. if he failed to assent to the sentence, ‘Arthritis is a disease’—we’d be reluctant to attribute ARTHRITIS to him. In the terms I introduced in Sect. 2: plausible theories characterize concept-possession in terms of (individualistic) referential and/or inferential relations, even if they make the anti-individualist concession that shortcomings in regard to those are compensable by suitable deferential dispositions.

In the previous section I suggested that in general, evaluation of Burgean arguments is a matter of weighing the interpretative intuitions elicited by the relevant stories against our commitment to the claim that a given individualist condition is (or conditions are) necessary for concept possession. Once again: in the argument involving Al, these conditions (mooted to be necessary for possession of ARTHRITIS) are inferential ones, but in the arguments involving Mary and Alice considered in Sects. 3 and 4, the mooted condition (on possession of RED\*) is a covariational one involving the concept’s referent.

And here’s the key point. The argument for anti-individualism about Alice developed in Sect. 4 is much more persuasive than the one about Mary discussed in Sect. 3, because Alice comes much closer to meeting the relevant mooted individualist possession condition than (pre-release) Mary. Recall: the mooted possession condition involves an introspective relation (which might be causal or constitutional) between the thinker and red\* experiences. Alice’s misapplication of ‘red\*’ provides evidence that her relation to red\* experiences is not quite the one in terms of which an individualist covariational theory would characterize possession of RED\*, *but it’s close*. She introspectively knows/remembers red\* with *some* degree of accuracy, even if not the same accuracy as Bob, and no doubt many of her introspective applications are correct. In contrast, before her release, Mary comes *nowhere near* to meeting the mooted covariational possession condition: she stands in *no*

introspective relation to red\*, since she's never had a red\* experience. To be sure, we can assume that Mary knows a lot about red\*, and some of the ways in which she manifests this knowledge may be evidence that she satisfies various *inferential* conditions which we might take to be partially constitutive of possession of RED\* (as well, no doubt, as of co-referring concepts her possession of which is not in question). But neither these nor her inclination to defer seems sufficient to compensate for the way she comes nowhere near to meeting the covariational condition. Thus, while Alice resembles Al—the thinker who is generally proficient with ‘arthritis’ but fails to some degree to meet one relevant mooted individualist possession condition, Mary is much more like the thinker who appears to fail much more comprehensively to meet the possession conditions an individualist might offer for ARTHRITIS, e.g. by failing to assent to ‘Arthritis is a disease’.

Another way to bring this out is to compare the concepts enjoyed by Alice and Mary to the metalinguistic content that either of them might express with a phrase like, ‘the phenomenal quality that people call “red\*”’. I mentioned in Sect. 1 that it's crucial to PC-theory that thinkers' PCs are not descriptive, metalinguistic contents like this (since even by PC-theorists' lights, an inexperienced subject like Mary could have such a concept). And there's no reason to suppose that the concept Alice expresses with ‘red\*’ is of this sort. As just observed, she has an introspective memory, however imperfect, of what it's like to see red, and even if she and Bob had never coined the term, ‘red\*’, she'd have the IC, RED\*. So she could possess RED\* without possessing the metalinguistic concept, and (like anyone else) she could have the metalinguistic concept without the IC. In contrast, it's much harder to be convinced that the concept Mary expresses with ‘red\*’ is not simply the metalinguistic one, since her only route to the referent of RED\* is via linguistic deference. In contrast to Alice, there's no way that Mary could possess the concept she uses ‘red\*’ to express without also possessing the metalinguistic concept, and vice versa. Alice could without contradiction think the thought she would express with, ‘perhaps red\* is not what other people call “red\*”’; but it's hard to see how Mary could think this without contradiction.

Metalinguistic contents like these are discussed in the familiar debate over Al and ARTHRITIS. Anti-individualists urge that an unpalatable consequence of individualism is that it entails that when pre-correction Al uttered mundane sentences like ‘Arthritis is painful’, he did not express the same first-order beliefs as others do with those sentences; but some individualists propose to off-set this with the reflection that individualism accommodates the suggestion that Al held closely related and widely shared metalinguistic beliefs, e.g. that the condition that others call ‘arthritis’ is painful.<sup>29</sup> Anti-individualists reply to this in various ways; but the one I want to note here is that an implausible feature of the metalinguistic proposal is how it ties the relevant beliefs to particular languages. The point of the proposal was to accommodate the attractive assumption that some of Al's relevant beliefs are shared with other thinkers, but the accommodation disappoints by allowing only English

<sup>29</sup> See e.g., Donnellan (1993), Chalmers (2002).

speakers to share these beliefs, since only English speakers have beliefs about others' uses of the term 'arthritis'.<sup>30</sup>

When we spell out the analogue of this debate for the arguments over RED\*, we once again find important differences between the case of Mary and all the rest: in particular, the proposal to off-set the implausibility of individualism by reflection on available metalinguistic beliefs is more compelling in the Mary case than in that of Alice. Here, the *prima facie* unwelcome consequence of individualism is that when pre-release Mary and  $t_2$  Alice say something ('mundane') like, 'Ripe tomatoes cause red\* experiences', they don't express a belief that they share with others; and the individualist rejoinder is that we can off-set the *prima facie* implausibility of this with the reflection that they might instead have the belief that ripe tomatoes cause experiences with the phenomenal character that others call 'red\*'. In Alice's case the appeal of this rejoinder is compromised just as it was in Al's by the way in which the relevant beliefs can be shared only with speakers of a certain language. Just as we want to allow that Al can share beliefs about arthritis with thinkers who don't use the word 'arthritis', we should want to allow that Alice can share beliefs about red\* with thinkers who don't use the neologism. In contrast, since pre-release Mary's only access to the referent of 'red\*' is linguistic/deferential, the proposal that she shares beliefs with experienced peers who don't use that term is much less plausible. To be sure, we've been assuming that Mary is as prepared as Alice to defer; but in Mary's case the deferential inclination is so closely tied to the word 'red\*' that it's hard to see what more there can be to it than what she'd get by associating 'red\*' with the description, 'the phenomenal quality that others call "red\*"'. Thus, while the reflection on metalinguistic beliefs does no more to rehabilitate individualism in Alice's case than it does in Al's, it suggests a very plausible analysis of Mary's deferential disposition.

All in all, then, the weighing exercises on which I have argued assessments of Burgean arguments depend deliver different verdicts in these two cases: they suggest that the argument may be effective in Alice's case, but that in Mary's it is much less so.

Neither Ball, Tye nor Veillet consider prosaic cases like Alice's in detail and so I suspect that if readers of Ball and Tye find their argument against experience-dependency compelling it is because they assume that if a concept can be possessed in the deferential way, then its possession can be *entirely* deferential—as Mary's would have to be. But I've argued that that was never a feature of plausible versions of Burgean anti-individualism; and Alice's story demonstrates as well as any that this would be a false dilemma. Concepts can be possessed deferentially while being experience-dependent. Anti-individualism about ICs does not entail that there are no PCs.

**Acknowledgements** For helpful comments and discussion I'm grateful to members of audiences in Belfast, Birmingham, Exeter, Helsinki, Keele, and Oslo. I'm especially grateful to Joe Morrison, Sarah

<sup>30</sup> See e.g., Fodor (1982), Tye (2009, pp. 65–66).



Sawyer, Scott Sturgeon, Pär Sundström, James Tartaglia, Henry Taylor, Naomi Thompson and two anonymous referees for *Erkenntnis*.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Alter, T. (2013). Social externalism and the knowledge argument. *Mind*, 122, 481–496.
- Ball, D. (2009). There are no phenomenal concepts. *Mind*, 118(472), 935–962.
- Ball, D. (2013). Consciousness and conceptual mastery. *Mind*, 122(486), 497–508.
- Balog, K. (2012). Acquaintance and the mind-body problem. In S. Gozzano & C. Hill (Eds.), *New perspectives on type identity: The mental and the physical* (pp. 16–43). Cambridge: Cambridge University Press.
- Block, N. (1996). Mental paint and mental latex. In E. Villanueva (Ed.), *Philosophical issues* 7 (pp. 19–49). Atascadero, CA: Ridgeview.
- Block, N. (2007). Wittgenstein and qualia. *Philosophical Perspectives*, 21, 73–115.
- Brandom, R. (1994). *Making it explicit*. Cambridge, MA: Harvard University Press.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4(1), 73–122.
- Byrne, A. (2002). Something about Mary. *Grazer Philosophische Studien*, 63, 123–140.
- Byrne, D. (2016). Do phenomenal concepts misrepresent? *Philosophical Psychology*, 29(5), 669–678.
- Byrne, D. (MS). Implicature and phenomenal content.
- Chalmers, D. (1996). *The conscious mind*. Oxford: Oxford University Press.
- Chalmers, D. (2002). The components of content. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 608–633). Oxford: Oxford University Press.
- Chalmers, D. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 220–272). Oxford: Oxford University Press.
- Chalmers, D. (2012). *Constructing the world*. Oxford: Oxford University Press.
- Craig, E. (1982). Meaning, use and privacy. *Mind*, 91, 541–564.
- Donnellan, K. (1993). There is a word for that kind of thing: An investigation of two thought experiments. *Philosophical Perspectives*, 7, 155–171.
- Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT.
- Fodor, J. (1980/90). Psychosemantics or where do truth conditions come from? In W. Lycan (Ed.), *Mind and cognition* (pp. 312–337). Oxford: Blackwell.
- Fodor, J. (1982). Cognitive science and the twin-Earth problem. *Notre Dame Journal of Formal Logic*, 23, 98–118.
- Fodor, J. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Oxford University Press.
- Goff, P. (2011). A posteriori physicalists get our phenomenal concepts wrong. *Australasian Journal of Philosophy*, 89(2), 191–209.
- Greenberg, M. (2014). Troubles for content I. In A. Burgess & B. Sherman (Eds.), *Metasemantics: New essays on the foundations of meaning* (pp. 147–168). Oxford: Oxford University Press.
- Greenberg, M., & Harman, G. (2006). Conceptual role semantics. In E. Lepore & B. Smith (Eds.), *Oxford handbook of philosophy of language* (pp. 295–322). Oxford: Oxford University Press.
- Harman, G. (1973). *Thought*. Princeton, NJ: Princeton University Press.
- Harman, G. (1987). (Nonsolipsistic) conceptual role semantics. In E. Lepore (Ed.), *New directions in semantics* (pp. 55–81). London: Academic Press.
- Hellie, B. (2004). Inexpressible truths and the allure of the knowledge argument. In Y. Nagasawa, P. Ludlow, & D. Stoljar (Eds.), *There's something about Mary* (pp. 333–364). Cambridge, MA: MIT Press.
- Hume, D. (2008) *An enquiry concerning human understanding*. In P. Millican (Ed.). Oxford: Oxford University Press.
- Jackson, F. (1986). 'What Mary didn't know. *Journal of Philosophy*, 83, 291–295.
- Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: MIT Press.

- Levene, J. (2007). Phenomenal concepts and the materialist constraint. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism* (pp. 143–166). Oxford: Oxford University Press.
- Loar, B. (1990/97). Phenomenal states (revised version). In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 597–616). Cambridge, MA: MIT Press.
- Loar, B. (2003). Qualia, properties, modality. *Philosophical Issues*, 1(1), 113–129.
- Locke, J. (1975) *An essay concerning human understanding*. In P. Nidditch (Ed.). Oxford: Oxford University Press.
- Lycan, W. (1996). *Consciousness and experience*. Cambridge, MA: MIT Press.
- Nida-Rümelin, M. (1998). On belief about experiences: An epistemological distinction applied to the knowledge argument. *Philosophy and Phenomenological Research*, 58(1), 51–73.
- Nida-Rümelin, M. (2007). Grasping phenomenal properties. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism* (pp. 307–338). Oxford: Oxford University Press.
- Papineau, D. (1998). Mind the gap. In J. Tomberlin (Ed.), *Philosophical perspectives* (Vol. 12, pp. 373–389).
- Papineau, D. (2002). *Thinking about consciousness*. Oxford: Oxford University Press.
- Papineau, D. (2007). Phenomenal and perceptual concepts. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism* (pp. 111–144). Oxford: Oxford University Press.
- Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.
- Perry, J. (2001). *Knowledge, possibility, consciousness*. Cambridge, MA: MIT Press.
- Putnam, H. (1975). The meaning of ‘meaning’. In H. Putnam (Ed.), *Mind, language, and reality: Philosophical papers* (Vol. II, pp. 215–271). Cambridge: Cambridge University Press.
- Shoemaker, S. (1982). The inverted spectrum. *Journal of Philosophy*, 74(7), 357–381.
- Shoemaker, S. (2006). The Frege-Schlick view. In J. Thomson & A. Byrne (Eds.), *Content and modality* (pp. 18–33). Oxford: Oxford University Press.
- Stalnaker, R. (1999). Comparing qualia across persons. *Philosophical Topics*, 26, 385–405.
- Stalnaker, R. (2006). Reply to Shoemaker. In J. Thomson & A. Byrne (Eds.), *Content and modality* (pp. 251–295). Oxford: Oxford University Press.
- Sturgeon, S. (1994). The epistemic view of subjectivity. *Journal of Philosophy*, 91, 221–236.
- Tye, M. (1999). Phenomenal consciousness: The explanatory gap as a cognitive illusion. *Mind*, 108, 705–725.
- Tye, M. (2003). A theory of phenomenal concepts. In A. O’Hear (Ed.), *Minds and persons* (Vol. 53, pp. 91–105). Cambridge: Royal Institute of Philosophy Supplement.
- Tye, M. (2009). *Consciousness revisited: Materialism without phenomenal concepts*. Cambridge, MA: MIT Press.
- Veillet, B. (2012). In defense of phenomenal concepts. *Philosophical Papers*, 41(1), 97–127.
- Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.