

Towards a consequentialist understanding of cognitive penetration

Dustin Stokes
University of Utah

If cognitive penetration of perception occurs, it is one interesting way that concepts and perception can relate. There are at least two ways this might be characterized. First, the content of cognitive states like belief are, on one plausible account, structured by concepts. And on one account of cognitive penetration, the content of one's cognitive states affects, in some intelligible way, the content of perceptual experience. Thus, in a naturally described way, one's concepts—understood as mentally represented Fregean senses—partly determine one's perceptual experiences. Second, one can think of concepts as recognitional capacities. And some have recently argued that one's experiences change by virtue of the acquisition of new recognitional capacities. This involves a kind of diachronic cognitive penetration, where one's concepts—understood as abilities—affect how one perceives the world.

Of course things are not so simple. What is clear is that cognitive penetration, if it occurs, would be of significant philosophical and scientific importance (and in addition to the possibilities outlined just above). But whether the phenomenon occurs is debated. Much of this debate is unfortunately rooted in lack of consensus on just what the phenomenon is or would be. There is little to no agreement, on either side of the debate, regarding a definition or analysis of the target phenomenon. For example, the first characterization mentioned above invokes a debatable semantic criterion for cognitive penetration. And the second characterization suggests long-term cognitive effects on perception, and some theorists claim, sometimes only implicitly, that cognitive penetration must be a direct, synchronic relation.

This lack of conceptual consensus has important consequences. Perhaps most importantly, if cognitive penetration is to be empirically testable, then there must be some

agreement on just what one is testing for (and thus how experiments should be designed and controlled). As it stands, empirical data are interpreted differently by different theorists, and by appeal to different criteria for what “counts” as a case of cognitive penetration. A variety of sceptical (that is, non-cognitive penetration) interpretations are invoked, and without an uncontroversial definition of cognitive penetration in hand, adjudication is difficult if not impossible. Indeed, how could one claim that the sceptical interpretations are less plausible, and that an observed mental phenomenon is *best* explained as cognitive penetration, if the latter notion is not clearly defined or, better, not at least defined in a way that is agreeable to both parties of the debate?

This paper outlines a methodological strategy for resolution and one that will, hopefully, encourage progress on an important (possible) aspect of the human mind. The simple prescriptive thesis is this: cognitive penetration should be understood (if not defined) in terms of its consequences. This consequentialism about cognitive penetration exploits the clearest point of agreement in the debate, namely, the importance of a phenomenon like cognitive penetration.

The discussion proceeds as follows. §I briefly outlines two extant definitions of cognitive penetration. §II clarifies how these definitions yield divergent verdicts on a pair of empirical case types. §III turns to the alleged consequences of cognitive penetration. And §IV characterizes and argues for the consequentialist strategy.

I. Two definitions

The term ‘cognitively penetrable’ was first coined by Zenon Pylyshyn (1980), and his most recent definition is the one most often cited in the literature. Here is its clearest statement: “[I]f a system is cognitively penetrable then the function it computes is sensitive, in a semantically coherent way, to the organism’s goals and beliefs, that is, it can be altered in

a way that bears some logical relation to what the person knows” (Pylyshyn 1999). This partial definition identifies a necessary condition for cognitive penetration that we might call the *semantic criterion* (See Macpherson 2012; Stokes 2013). The criterion as presented by Pylyshyn--call it (SC)--is ambiguous. In some places, Pylyshyn seems to have in mind a logical, inference-supporting relation. So a cognitive state like a belief penetrates a perceptual experience only if the content of the belief could support an inference to the content of the resulting experience. This makes the semantic criterion a rationality criterion. One worry here is that, at least as discussed in the literature, it is unclear that worries about cognitive penetration are worries about a rational relation. Indeed, most theorists are concerned with cognitive penetration qua *non-rational* relation--going back to the concerns about theory-ladenness of the 1950s and 60s, and up through the epistemic circularity worries current in literature today (Siegel 2011; Lyons 2011; Raftopoulos 2006).¹ Elsewhere, Pylyshyn ostensibly takes the criterion to require only representational coherence: “This is the essence of what we mean by cognitive penetration: it is an influence that is coherent...when the meaning of the representation is taken into account” (Pylyshyn 1999: 365, fn3). This weakened semantic criterion requires that one could, perhaps under idealized conditions, identify how the content of the penetrating state, say a belief, affected the content of the resultant perceptual experience. Here again, however, one may worry about the motivation for the criterion thus interpreted, since it imposes a kind of operationalist condition on cognitive penetration.

The motivation for Pylyshyn’s semantic criterion, for either of the above interpretations, is ultimately clear. The problem is that this motivation is insufficiently theory-neutral. To see this, one can trace Pylyshyn’s thinking back to his earlier articulations of the concept of cognitive penetrability. The broader goal for Pylyshyn’s earlier relevant

¹ These are discussed in detail in Section III.

work was to introduce a criterion that marked off the proper subject matter for cognitive science. Nearly two decades earlier, Pylyshyn wrote:

Much of the paper elaborates various conditions that need to be met if a *literal view of mental activity as computation* is to serve as the basis for explanatory theories. The coherence of such a view depends on there being a principled distinction between functions whose explanation requires that we posit internal representations and those that we can appropriately describe as merely instantiating causal physical or biological laws. In this paper the distinction is empirically grounded in a methodological criterion called the ‘cognitive impenetrability condition’ (Pylyshyn 1980: 111; emphasis added).

Thus initially, Pylyshyn took *cognitive penetrability* to distinguish cognitive phenomena (understood computationally) from *cognitively impenetrable* functional architecture (understood biologically). This picture was then amended a few years later, where cognitively penetrable systems are influenced by systems the explanation of which requires terms of rules and representation, and explanation of that influence will also require terms of rules and representations (Pylyshyn 1984). It was here that something like Pylyshyn’s current semantic criterion emerged; and its application placed perception under the category of biological, functional architecture. One odd consequence, one might think, is that this puts perception outside the purview of cognitive science. And this is for the reason that cognitive science, understood in Pylyshyn’s way, involves a commitment to the computationalist doctrine.

It is this doctrine that ultimately motivates a semantic criterion, no matter its particular form across the nearly twenty years that Pylyshyn has written on the topic. The central worry here is that the computationalist theory is controversial, and most certainly not one to which all parties to the cognitive penetration debate commit. Indeed, a number of philosophers as well as theorists who count themselves “cognitive scientists” resist or reject commitment to this form of computationalism, and included among these are theorists who otherwise have interest in and have written on whether cognitive penetration occurs. So if Pylyshyn’s notion of cognitive penetration centres around the semantic criterion, and the

only motivation for this criterion is the computationalist one, then this definition fails, ironically, for reasons of theory-ladenness.

A recent alternative definition follows Pylyshyn by maintaining that cognitive penetration is not merely a causal relation between cognition and perception, while excluding his semantic criterion.

(CP) A perceptual experience E is cognitively penetrated if and only if (1) E is causally dependent upon some cognitive state C and (2) the causal link between E and C is internal and mental.

This definition accomplishes a few things. First, it makes explicit that the phenomenon of interest involves (at least partly) an effect on perceptual *experience*. Thus on today's orthodox understanding, the phenomenal character of a sensory experience is affected by a cognitive state like belief. Clause (1) thus ensures that instances where states like belief or memory are affected by antecedent cognitive states do not count as cognitive penetration. Clause (2) maintains that this relation must be an internal one, and the causal chain must involve mental states or processes (but with no restriction on how long that chain is). This clause insures that instances where one's beliefs or other cognitive states cause an action of some kind which then causes a (change in) perceptual experience do not count as cognitive penetration.

Much could be said about (CP). Suffice it to say that this definition is not, like most definitions, without its problems.² (Indeed, problems of application of the definition will emerge in §II below.) So instead of any extended analysis on those scores, it will prove more instructive, for the purposes of this paper, to instead identify the ways in which (SC) and (CP) deliver divergent verdicts on different sets of cases.

² (CP) is taken from Stokes 2012; forthcoming. See Stokes 2013 for one possible counterexample to (CP).

II. Divergent verdicts

Defining cognitive penetration, as a theoretical task, is of little final importance to theorists involved in this debate. A definition is instead subservient to answering empirical questions about the mind, namely, whether cognition and perception relate in some specially important way/s. So getting the definition right is only important to the degree that it assists in performing and interpreting the results of experimentation and reflection. And it is for this reason that an uncontroversial--that is, generally agreed upon--definition is needed. Without this agreement, theorists apply distinct criteria to the same data, and come away with diverging verdicts on whether that data evidences cognitive penetration. Two kinds of case make this clear.

The first set of cases involves apparent effects of evaluative attitudes on visual size perception. In a now famous experiment, Bruner and Goodman (1947) asked child subjects to report the size of American coins. In both control and experimental groups, subjects were placed before a wooden box with a glass face. Subjects were to turn a knob, located on the bottom right corner of the box face, in order to adjust the circumference of a patch of light to match the presented targets. Targets were placed in the subject's left hand, and held six inches to the left of, and on the same horizontal plane as, the adjustable light patch. Target stimuli for experimental subjects were ordinary coins of values ranging from 1 cent to 50 cents. Target stimuli for controls were cardboard cutout analogues of the varying coins. After an initial training period with the apparatus, subjects then took as much time as they wished to match the light patch to the respective target stimuli. Control subjects reported the size of the cardboard cutouts with near perfect accuracy. Experimental subjects consistently overestimated the size of the coins, and by differences (by comparison with controls) as high as 30%. In a second variation of the study, subjects were separated into groups of "poor" and "rich" children. Here both groups overestimate the size of the coins, but poor children

by as much as 50%. Bruner and Goodman interpret the data to suggest that subjects' values or desires for money are influencing how the coins are perceived. Put simply, the high value of money results in seeing money as bigger (than it actually is).

Following this early New Look study was a barrage of similarly spirited experimentation, some of it apparently supportive of the central New Look claim--that perception and cognition are "continuous"--some of it not. A number of theorists, both in philosophy and psychology, have revived attention to the New Look approach.³ One very recent example involves valenced images and size perception (van Ulzen et. al 2008). Here subjects were presented, on a computer screen, with circles containing either negatively, positively, or neutrally valenced images (e.g. respectively: an aimed gun, kittens, mushrooms). The task was then to adjust a report circle, located on the bottom right corner of the computer screen, by pushing or pulling a computer mouse. Leaving out various details, subjects consistently reported circles containing negatively valenced images to be larger than circles containing images of positive or neutral valence (where the actual size of the image-containing circle is static across trials). On the face of it, these are cases where evaluative attitudes held by subjects (perhaps a fear or aversion to the objects depicted in the negative images) are influencing visual size perception.

These data are certainly compelling. But the question to be asked here is whether they are to be explained as instances of cognitive penetration, or something else. Consider, in turn, the two definitions discussed in the previous section. Recall that Pylyshyn's (SC) requires, at least, an identifiable representational coherence between the casually antecedent mental state and the resultant perception. In both the Bruner and Goodman studies, and the van Ulzen et al studies, the antecedent mental state is, plausibly, some orrectic attitude (for example, a desire for money in the first case, and a negative affective attitude towards images

³ See Balcetis and Dunning 2010; van Ulzen et al 2008; Stokes 2012.

like guns and spiders in the second). In both studies, it appears that size perception is affected: subjects see coins or circles with negative images as bigger. However, application of (SC) undercuts this interpretation: it is far from clear that the semantic coherence criterion is satisfied. Consider the Bruner and Goodman results. Here, Bruner and Goodman conjecture, there is some background desire or value for money. Candidate cognitive states/contents would include *S* desires that [I have money] or *S* evaluates that [Money is good]. And the (allegedly) resultant experience is one where coins are experienced as bigger (than they, objectively, are). Put crudely, and begging important questions about the content of experience for the moment, suppose that this experience should be specified as something like: *S* sees that [the coin is size *n*]. (Here it does not matter how the variable, *n*, is filled; it simply placeholders a size-specific slot in the content of the subject's experience. And as per the experimental data, this quantification is such that the coin is experienced as bigger than it, objectively, is.) Here then is the trouble: there is no clear way to specify how contents of the first sort would cohere with a content of the second sort. As it is sometimes put in discussion of perceptual content and/or perceptual justification, there is nothing in the first content to "hook up" with the content of the (allegedly) resultant experience. One way to put this is in terms of inference, one could not infer [the coin is size *n*], for any *n*, from, say, [Money is good]. And this is true no matter the attitude taken towards the latter content. (For example, the inference would be no better if the attitude was doxastic rather than orrectic.) Even on the weaker, non-inferential, interpretation of (SC), the verdict is the same: the content of the antecedent state (the "meaning" as Pylyshyn puts it) bears no intelligible connection to the content of the second state. Therefore, these data do not provide evidence for cognitive penetration of perception.

This interpretation of the relevant experimental cases is further clarified by identifying some alternative interpretations of these or similar data (*alternative* to a cognitive

penetration interpretation). Some have, in considering these and other New Look(ish) experiments, suggested that the background mental states have an effect on how the perceptual stimulus is remembered, but not how it is perceived. Call this the *memory interpretation*.⁴ Similarly, one might argue that the subjects across control and experimental circumstances have normal, veridical visual experiences (as of the sizes of the target stimuli), but then make judgments about these perceived stimuli in a way informed by the background mental states. Like the previous interpretation, according to this *judgment interpretation* we have, at most, evidence for cognitive effects on (other) cognitive states. And this is uncontroversial.⁵ Finally, one might allow that the perceptual experiences of experimental subjects are affected (by contrast to control subjects), but argue that this effect is enabled by active changes in attention. Thus, as a result of the subject's background mental states (say, the subject's desire for money), she attends differently to the target stimulus, and in ways that alter experience. This, the critic will urge, is like in kind to instances where one knows how attending differently will alter experience. Thus one sees a figure as a duck or as a rabbit, by attending to different parts of the figure. And this attentional shifting depends upon "knowing the trick", as it were.⁶ But this is not the phenomenon of interest, since the effect on perception is indirect: attentional *acts* mediate between background mental states and the resultant experience. Call this the *attention-shift interpretation*. One or other of these interpretations, an advocate of (SC) might urge, better explain the experimental data. So, we have not instances of cognitive penetration, but something else: either cognition affecting cognition, or action-guided perceptual change.

⁴ See McCurdy 1956.

⁵ See Pylyshyn 1999.

⁶ See Fodor 1983; 1988. See also Macpherson 2012 and Stokes 2012; 2013 for discussion of these interpretive strategies.

What of the second definition, (CP), and its application to the New Look cases?

Recall that this definition eschewed the semantic criterion, maintaining instead that cognitive penetration requires both an effect on the phenomenal character of perceptual experience, and an internal, mental causal link between antecedent mental state (a belief, value, desire, etc) and resultant experience. This definition has the advantage of incompatibility with the above-named alternative interpretations: if a phenomenon or case meets (CP), then it is not an instance of mere cognitive effects on cognition (by appeal to clause (1) of (CP)) nor an instance of action-guided perceptual change (by appeal to clause (2) of (CP)). And plausibly, these New Look cases do meet the conditions of (CP). Taking again the Bruner and Goodman experiment as the example, these subjects have a desire or value for money which affects, directly, the perceptual experience of the size of coins (by contrast with the control stimuli, where no such effect is recorded). Given the online nature of the task--subjects inspect the stimulus *as* they adjust the light patch to match--the memory interpretation is implausible. Similarly, there is little reason to think that the judgment interpretation is apt, since it would require a consistent mismatch between veridical visual experience of the coins and an erroneous, online report of the size of the coins. Subjects give no indication (for example, surprise or confusion) that their reports deviate in this way from their current experience. Finally, the attention-shift interpretation is implausible, since there seems to be no relevant attentional difference between controls and experimental subjects, and thus no attentional explanation of the respective differences between these subjects' reports. This, anyway, is how an advocate of (CP) might defend its application to the case.

The critic may be unpersuaded by this line of reasoning. No matter (for now). The important point for the present discussion is that, as some have argued, appeal to (CP) plausibly yields a pro-cognitive penetration verdict on this set of data. Result: by appeal to one definition, (SC), a set of data is judged *not* to evidence cognitive penetration. By appeal

to another definition, (CP), the same data is judged as good evidence for cognitive penetration. Put simply: (SC) plausibly yields a ‘NO’ verdict, (CP), a ‘YES’ verdict. This is clear case of theoretical cross-talk. And no matter one’s side of the debate, this scenario does not bode well for theory adjudication. A second type of case further reveals the problem.

Some recent philosophers of perception have argued that perceptual experience represents *high-level* properties. If vision, for example, is representational then it represents basic low-level properties like colour and shape. Traditionally, however, properties like being of a natural or artificial kind, being caused/a cause, and being expressive of emotion are understood as the result of post-perceptual cognitive processes like judgment or belief; these high-level properties, tradition has it, are not “picked up” by perception. Both the phenomenology of experience and empirical evidence provide reason to doubt the traditional view.

Susanna Siegel (2006) argues from phenomenal contrast. It is introspectively plausible that one’s overall experience of an object or event of kind *K*, after one learns to recognize instances of *K*, changes. Taking one of Siegel’s central examples, as one learns to recognize pine trees and thereby learns to pick them out in an array of different types of tree, the phenomenology of one’s overall experience changes (when seeing pine trees). So one’s overall experience, when seeing pine trees is a part of that overall experience, eventually differs phenomenally from overall experiences (involving pine tree seeings) had prior to acquisition of the recognitional capacity. Siegel then claims that this contrast in overall phenomenology is best explained as a difference in the representational content of perceptual experience, the latter of which is a part of the overall experience. Finally, the change in perceptual representation is best explained as a change in high-level property representation. It is not that one’s experience of colours or shapes changes upon acquisition

of the capacity to recognize pine trees. Instead, one has *visual* experiences as of pine trees; one perceptually represents the relevant objects *as* pine trees. The argument concludes, then, by identifying kind-properties, as well as other high-level properties, as admissible contents of perceptual experience.⁷

High-level perceptual representation is particularly plausible in contexts of perceiving art. It is clear that different perceivers of artworks make different judgments, and also that judgments often enough co-vary with the art-knowledge of those perceivers. Additionally, one might think that it is not only judgments that co-vary with knowledge in this way, but in fact that experiences co-vary. A perceiver *S* with sufficient knowledge of abstract expressionism will know that (nearly) monochromatic colours and rough rectangular shapes are *standard* features for many of Mark Rothko's paintings (what are known as his "multiform" works), while there will be some *variability* within this narrow range of colour and shape. Suppose another perceiver, *T*, lacks any such knowledge of abstract expressionism or Rothko, knowing only that standard for paintings, as such, is paint on canvas, while shapes and colours can vary widely. Suppose that *S* and *T* both perceive one of Rothko's multiform pieces (the reader might consider, for example, *Four Darks in Red*, 1958). Given *T*'s lack of knowledge, he is likely to aesthetically judge the work to be dull or lifeless, since it is so colourless and simple in its shapes relative to the broad category of painting. *S*, however, has more precise knowledge about this kind of work, and will perceive the work not just as a painting, but in the category of Rothko's multiforms. Again, standard for this category are nearly monochromatic colours and rough rectangles, but these features do vary across various Rothko pieces. *S* will likely attribute different aesthetic properties; he might judge the work to be striking or vivacious. Now, one might maintain that these aesthetic

⁷ Tim Bayne (2009) argues to the same conclusion by appeal to empirical research on subjects with associative agnosia. See also van Gulick 1994, Siewart 1998, Siegel 2009, and Hawley and Macpherson 2011.

property attributions are merely post-perceptual judgments. But the following possibility is sufficiently plausible for serious consideration: these perceivers visually perceive the works differently, by virtue of their differing art-knowledge. *S*, for example, does not just judge the work to be striking or vivacious, she perceptually represents the work as instantiating these high-level aesthetic properties. And this is a consequence of learning and knowing about art.⁸

These possibilities are both examples of what we might call diachronic cognitive effects on perception. On the face of it, as one learns to recognize certain kinds of things--be it pine trees or Rothko paintings--one's phenomenal experience of these things changes. And this change takes time. Here is one final example of a case of diachronic effects on perception, this one not requiring any appeal to high-level property representation. Early research on inverting lenses--goggles or lenses that, in short, turn one's world upside down--revealed remarkable human capacity for relatively quick adaptation. Initially, subjects wearing the lenses are radically disoriented, failing at both tasks of descriptive judgment and vision-guided action. But after a training period, often a week or so, subjects adjust, performing the same tasks in statistically successful ways (Stratton 1897; Kottenhoff 1957; Taylor 1962). Paul Churchland (1988) interprets these data as an application of learning--namely, about novel relations between movement and resultant experience, about expectations and what experience actually delivers, and so on--to seeing. In short, the subjects have distinct visual

⁸ See Walton 1970. See Stokes (unpublished manuscript) for discussion of both cognitive penetration and the perception of art, and a defence of diachronic cases as genuine examples of cognitive penetration.

experiences--of motion, orientation, shape--before and after learning how to “use” the inverting lenses successfully.⁹

Each of these cases involves some learning--to spot pine trees, to spot categories of art or aesthetic properties, learning how to use a radically distorting visual apparatus. And, apparently, this learning affects the phenomenology of perceptual experience. What is crucial to note, is that this effect (if it is a genuine effect) takes place over time. This is what one would expect, insofar as the learning also takes place over time. So, plausibly, the look of pine trees, say, gradually changes as one better learns to recognize pine trees. The question to ask now is whether this diachronic feature of all of the relevant cases, whatever their differences may be, excludes such cases from being instances of cognitive penetration. Consider once more the two definitions articulated in the previous section, (SC) and (CP).

Upon first glance, one might apply (SC) to get the verdict that these diachronic cases are indeed cases of cognitive penetration. Specifying the relevant contents here, as is often the case, is a complicated (if not strained) matter. But things might go as follows. Take the first example of high-level property representation. Here a subject *S* gradually learns that (or, if one prefers, forms a belief that) [Pine trees have features *F*], where ‘*F*’ placeholders the perceivable shape and colour features typically possessed by pine trees. According to the high-level content theory, and simplifying, *S*’s experience of pine trees will also gradually change, where eventually *S* will (in the presence of pine trees) token an experience of the type: *S* sees (or sees that) [There is a tree with features *F*]. Alternatively, *S* sees (or sees that) [There is a pine tree], where the high-level property of being a pine tree is at least partly

⁹ There are worries, however, about whether the evidence is sufficient to suggest changes in the phenomenology of subjects’ experiences, rather than just a cognitive adjustment to a stable (but upside down) environment. See Prinz 2006 and Schwitzgebel (unpublished manuscript). For present purposes, however, this detail is of no consequence. Grant that there is an apparent change in the phenomenology of experience, the question is whether the diachronic nature of this type of case, and of the high-level property cases, is incompatible with cognitive penetration. Distinct definitions deliver distinct verdicts here, or at least, verdicts for different reasons.

characterized by certain low-level colour and shape properties, *F*. Here the coherence is much clearer than in the previous set of cases: one learns about pine-tree looks and the content of these concepts or beliefs informs, in a coherent way, the consequent perceptual experiences of pine trees. A similar story could be told about the cases involving perception of art.

And finally, a similar story may be told about the inverting lenses cases. One simple way to characterize these cases concerns the shift in the apparent orientation of the world, as experienced by subjects. Suppose that ‘rightside up’ denotes the way the world actually is, and appears to be, when we have normal visual experience. And ‘upside down’ denotes the way the world would be if the initial experiences of inverting-lens wearers were accurate. As it is sometimes put in discussions of perceptual contents, ‘rightside up’ designates a proper part of the accuracy conditions for normal human visual perception; ‘upside down’ designates a proper part of the accuracy conditions for lens-inverted human visual perception. These designations give the content, or part of it, of the relevant perceptual experiences. Now, the inverting-lens wearer will initially experience the world as upside down but will also, because she understands the nature of the apparatus, have a belief that [The world is rightside up]. (The apparatus is vision inverting, but not world inverting! And subjects understand this.) The same subject also acknowledges that her current perceptual experience is, so to speak, at tension with this general belief regarding world-orientation: things initially appear upside down. And this belief (that the world is rightside up) is maintained and applied as the subject better and better copes with her disoriented visual array. Eventually, there is a shift in the subject’s perceptual experience, part of the content of which would be specified as [The world is rightside up]. Here the coherence between background cognitive state and resultant perceptual experience could not be clearer.

As it turns out, advocates of (SC), or something like it, resist the claim that these diachronic cases are genuine cases of cognitive penetration. This is not entirely unreasonable, since semantic coherence as specified by Pylyshyn (1999) is merely a necessary condition. So perhaps there are other conditions required but not met by these cases. Alternatively, an advocate might maintain that these cases do not even meet (SC) as baldly characterized, since (SC) requires a coherence relation between a cognitive state and perception. And, one might maintain, the cases in question do not involve a relation between belief, say, and perception, but instead just some kind of non-cognitive perceptual change. Jerry Fodor gives this reply to Churchland's discussion of the inverting lenses. "For there are, after all, good ecological reasons why you might expect plasticity of this sort.... what needs to be kept open for re-calibration is whatever mechanisms compute the appropriate motor commands for getting to (or pointing to, or grasping) a visible object on the basis of its perceived location. Adaptation to inverted (and otherwise spatially distorting) lenses is plausibly an extreme case of this sort of recalibration" (Fodor 1988: 193). So this *intra-perceptual interpretation* may well be invoked for the diachronic cases, which alleges that the changes in perceptual systems, are made *by* perceptual systems, over time. So, no cognitive penetration because one necessary relatum, a cognitive state, is absent in the relevant causal story.

So, it is less than perfectly clear what verdict (SC) delivers. On some interpretations of the cases, semantic coherence obtains, and obtains between a cognitive state and perception. But on different interpretations, there is coherence, but not between the appropriate mental states. How are these same cases treated by (CP)?

Grant for the moment that these diachronic cases involve a cognitive effect on the phenomenal character of perceptual experience. Learning to recognize pine trees or Rothko works, for example, is a cognitive achievement and one that results in changes in perceptual

representation. Grant, then, that these cases meet clause (1) of (CP). However, one may worry that the diachronic nature of the cases precludes them from meeting clause (2) of the definition. This condition requires that the causal link between cognitive state *C* and resultant experience *E* is entirely internal to the perceiving agent and, moreover, mental. But one might worry that various actions mediate, over time, between *C* (or *Cs*) as it takes hold and changes, and *E* (or *Es*) as it changes in correlative ways. For example, as one hones the capacity to recognize pine trees, one will actively use attention and explore one's environment. One will look at pine trees from different angles, distances, and under various viewing conditions, perhaps asking questions of the trained pine-tree spotter along the way. And these actions are required, one might infer, to bring about the relevant changes in perception of pine trees. If this is the appropriate characterization of these cases, then the causal chain from *C* to *E* violates clause (2): some events in that chain are external to the perceiver and, therefore, non-mental. Interpreted this way, diachronic cases are not cognitive penetration per (CP).

Of course things may not be this simple. Another plausible characterization of the diachronic cases locates the just mentioned attentional and exploratory acts in the (stages of) acquisition and honing of the recognitional capacity, but not in the causal story about how that changing capacity influences perceptual experience (in ways that change, accordingly, over time). So one might grant that learning to spot pine trees requires physically exploring one's space and actively using attention, while maintaining that the resultant changes in perception are internal, mental causal consequents of the pine-tree spotting capacity. In short, learning the look of pine trees as such, involves action. But perceptually representing the property of pine trees is a (relatively) direct effect of that learning (or, once learned, of the execution of that capacity). On this interpretation, then, clause (2) of (CP) is also satisfied, and in spite of the diachronicity of the cases.

As with (SC), then, a verdict from (CP) on diachronic cases is less than straightforward. Here is not the place for adjudication or extended analysis on how these cases should be understood in terms of one or both definitions. Again, the task here is not to make a conclusive case that these cases are (or are not) instances of cognitive penetration.¹⁰ Instead, the lesson to be gleaned is that application of the definitions is challenging, and the two definitions may again yield divergent verdicts. Or, they may yield the same verdict, but for diverging reasons.

One reply to this situation is to conclude that (SC) and (CP), and the theorists that advocate these definitions (or something like them), target distinct *definienda*. If the definitions are used in such divergent ways, or deliver clearly distinct verdicts on the same observed phenomenon, then perhaps they are simply not talking about the same thing. There is some truth to this reply: it acknowledges what may be a mistake endemic to the cognitive penetration literature. Attempts to define or characterise the phenomenon have erred towards defining cognitive penetration as such, and in a way that has lost sight of the supposed consequences of the (possible) phenomenon. That is to say, there are reasons that philosophers and cognitive scientists began discussing the possibility of something like cognitive penetration, and those reasons all concern the consequences that a phenomenon would have *if* that phenomenon occurred (and, perhaps, with some frequency). These consequences are, it will now be argued, what is of common interest to both parties of the cognitive penetration debate, and no matter what definitions such parties offer. The fix, then, is to understand the target phenomenon in terms of its consequences and, in turn, to abandon essentialist definitions of ‘cognitive penetration’.

¹⁰ As it stands, the diachronic cases appear mostly disregarded, sometimes only implicitly, as not cognitive penetration; but as will become clear in the discussion of the two sections that follow, this dismissal may be unmotivated.

III. Consequences of cognitive penetration

There are three central consequences of interest. They concern, respectively, theory-ladenness of empirical observation, the epistemic role of perception, and (modular) architectures of the mind. These are briefly discussed in turn.

Traditional empiricist models of scientific inquiry came under fire in the middle of the 20th century. (Empiricism here includes, but is not exclusive to, logical positivists and logical empiricists.) A number of philosophers challenged the traditional assumption that scientific observation is theory-neutral in a way that would support rational theory choice (Hanson 1958, 1969; Kuhn 1962; Feyerabend 1962). At least one version of this worry understands the relevant observation as, simply, perceptual observation. The corresponding worry is epistemic: if perceptual observation is laden with theory (in particular the theory or theories being tested), then it will not provide a means for rationally choosing or adjudicating between scientific theories.

So two theories--say, an earth-centred theory vs. a sun-centred theory of our galaxy--may be equally successful in terms of elegance, parsimony, internal coherence and other (non-observational) criteria of theoretical success. The plausible empirical method of adjudicating between these competing theories is to test them against perceptual observations of the world. However, if the observations made by the respective theorists are imbued with theoretical commitment, then those observations fail to provide a neutral form of adjudication. Extending the toy example, the earth-centred theorist might report, upon watching the sunrise, that he sees a sun moving across a stationary horizon. And, predictably, the sun-centred theorist reports seeing a horizon moving to expose a stationary sun. Accordingly, both theorists will report empirical corroboration of their respective predictions, and no progress is made in motivating a rational choice between theories. This

is, on one characterization, a perceptual phenomenon. Differences in perception potentially undermine the rational, theoretical role of empirical observation.¹¹

Although theory-ladenness is typically presented as problematizing the empiricist picture of science and, more generally, the role of perception in scientific theory choice, it is worth noting possible instances of “good” theory-laden observation. Kuhn, discussing the effects of “normal science” on its practitioners, writes:

What were ducks in the scientist’s world before the revolution are rabbits afterwards... Transformations like these, though usually more gradual and almost always irreversible, are common concomitants of scientific training. Looking at a contour map, the student sees lines on paper, the cartographer a picture of a terrain. Looking at a bubble-chamber photograph, the student sees lines on paper, the physicist a record of familiar subnuclear events. Only after a number of such transformations of vision does the student become an inhabitant of the scientist’s world seeing what the scientist sees... (Kuhn 1962: 111).

These kinds of perceptual achievements, supposing for the moment that they could occur, might be beneficial in some way. The cartographer more efficiently (perceptually) identifies the map’s features, the sonograph technician, the sex of a fetus. This kind of accuracy and efficiency gives the expert in science and medicine an advantage over the non-expert. And this advantage would depend upon the background theoretical and technological understanding of the perceivers.

Questions remain here, of course, about whether this theory-ladenness would be epistemically good. By many criteria for epistemic normativity, theory-ladenness remains an epistemic bad even if it sometimes produces good results. So, for example, just as unreliable belief forming mechanisms can sometimes produce true beliefs, theory-laden observation may occasionally produce more efficient or accurate perceptual observation. But a true belief formed by unreliable mechanisms is in no way justified by the mere fact that it is true.

¹¹ This example is modeled on Hanson 1959: 5-8.

Analogously, one might maintain, theory-ladenness is generally epistemically pernicious even if it sometimes improves observation and testing. Suffice it to say, then, that there are open questions about the epistemic status of theory-laden observation. What is clear, nonetheless, is that if observation is theory-laden, this is of epistemic import. Theory-ladenness then, no matter how the open questions are answered, is an important possible consequence of cognitive effects on perception.¹²

The second consequence is also epistemic, generalizing from the first consequence. Perceptual experience, on the most intuitive picture, provides us with knowledge about the world. But this epistemic role for perception is threatened if there are circumstances where background cognitive states influence perceptual experience that, in turn, influence belief formation. The most obviously pernicious cases are ones that involve a rough causal schema of the form:

$$\textit{Belief that } P \rightarrow \textit{Experience that } P \rightarrow \textit{Belief that } P$$

Here the causal history of the consequent belief involves a circularity that, plausibly, undermines reason for that very belief (Siegel 2011, forthcoming). The problem with this circularity, or anything like it, is clear given the supposed role for perception and action. Cognitive impenetrability theorists are no less clear on this point. As Fodor suggests, the “function of perception is to deliver to thought a *representation* of the world” (1985: 5). And since this representation is supposed to inform belief and action (about or in response to the here-and-now), it should track not what the agent believes, wants, or otherwise thinks about the non-present; it should instead track the present environment, here-and-now. Indeed, it is for this very reason that Fodor argues that perceptual systems are cognitively impenetrable, claiming that “isolation of perceptual analysis from certain effects of background belief and

¹² For more recent discussion see, among others, Churchland 1979, 1988; Fodor 1984; 1988; Brewer and Lambert 2001; Raftopoulos 2001, 2006.

set... has implications for both the speed and *objectivity* of perceptual integration” (Fodor 1983: 43; emphasis added; see also Pylyshyn 1980). Whether perceptual processing is isolated (and thus, cognitively impenetrable) is an empirical question, and up for current debate. But the consequence here is clear: *if* perception is influenced by background cognitive states in these ways, then the objectivity of perceptual representation is threatened.

Here too there are important open questions. One such question concerns the scope of the epistemic consequence. If a belief has a circular etiology (following the schema above), does this result in a mere case of *local* unjustified belief, or does some kind of *global* scepticism follow? Jack Lyons frames the question this way. “Suppose...that top-down influence is merely probabilistic in the sense that theory-consonant observations are more likely than they would otherwise be but that objective facts are still significant determinants of what is observed...Many factors keep our perceptual access to the world from being infallible---poor observation conditions, camouflage, distraction, sleepiness, etc.---why should prior beliefs be more than another such factor?” (Lyons 2011). Lyons is suggesting that these other factors are insufficient to motivate global scepticism, so why should something like cognitive penetration motivate global scepticism? A response might go: factors like camouflage, distraction, and the like are perceptually *muddying*, but they are epistemically neutral with respect to some proposition *P*. Belief etiologies like that schematized above, however, present a perceptual *bias* towards some *P*. In some of the alleged cases, the top-down influence is biasing in ways that raise the probability that the subject infers that *P*. So it is not the confounding of infallibility—the muddiness--that is epistemically pernicious; it is the apparent biasing effect towards experiences that encourage certain inferences. But this brings Lyons’ point regarding local versus global results to the

fore: are these biasing effects sufficiently frequent to motivate a general epistemic problem, or are they just another example (even if slightly different) of human fallibility?¹³

The first consequence, then, is epistemic, concerning the scientific-theoretical role of perception. The second consequence concerns, as we might say, the “everyday” knowledge-providing role of perception. This general epistemic consequence is importantly relevant to the third possible consequence of cognitive penetration, namely, architectures of the mind. Modularity theorists (at least of Fodorian strength) claim that perceptual systems are *informationally encapsulated*. A core motivation (perhaps the *primary* motivation) for this claim is that the processing of such systems would be immune to error introduced by the broader cognitive system. Since computations performed by these modules are supposed to be insensitive to what the organism knows, expects, or wants, the resulting perceptual representations more reliably inform the organism about its environment (Fodor 1983: 68-70; see also Pylyshyn 1980). The modularity thesis is an empirical one, and one that has set substantial research agendas in cognitive science. So, if it turns out that cognitive penetration of perceptual systems occurs (that is, perceptual systems are *unencapsulated* relative to cognitive systems), then this alleged feature of modules, even if epistemically desirable, is not actual.

Although this last consequence is often put this way, it is not exclusively a concern about perceptual *processing*.¹⁴ If perceptual experience is directly influenced by background cognitive states, modularity is no less threatened. Here are two reasons. First, although higher

¹³ Another open question concerns whether the general epistemic consequences are, by necessity, bad. Just as with theory-ladenness, one can conceive of everyday scenarios where expertise or knowledge improves perceptual accuracy.

¹⁴ For instance, Fodor emphasizes computational mechanisms of perceptual “input systems” as informationally encapsulated from “central cognitive processors” (Fodor 1983). And Pylyshyn defends (typically) the cognitive impenetrability of perception with evidence for the impenetrability of early vision: a functionally defined component in visual processing that computes 3D shape descriptions of objects (Pylyshyn 1999).

level effects on some components of perceptual processing do not imply the cognitive penetration of experience; and higher level effects on perceptual experience do not imply (by themselves) that a particular stage of perceptual processing is penetrated; higher level effects on perceptual experience imply an effect on perceptual processing at *some* stage. At least this follows if we assume any kind of physicalism, according to which experience is identified with, constituted by, (metaphysically) determined by, or the output of perceptual processes. Second, as mentioned above, the modularity thesis is largely motivated by epistemic concerns. The idea is that since the envisioned encapsulated perceptual module rigidly performs its computational function and with no interference from extraneous higher-level information--what the organism knows, expects, or wants--the resulting perceptual representations more reliably inform the organism about its environment. And a concern with the reliability or accuracy of perception is a concern with perceptual *representation or experience*, not merely processing. One forms beliefs on the basis of what one sees, hears, and so on. So modularity theorists have to be concerned with perceptual experiences, not just processing.¹⁵

These three consequences are substantial, concerning the epistemology of science, everyday reasoning and rationality, and broad cognitive scientific theories about the architecture of the mind. Some of the details of such consequences require further analysis, and it is unclear which alleged cases of cognitive penetration imply which, if any, of the consequences. But what is clear is that agreement converges on these consequences. Fine details aside, theorists on either side of the cognitive penetration debate agree that the importance of the possible phenomenon--cognitive penetration--consists in its bearing these consequences. And so, it will now be argued, the question--Is perception cognitively

¹⁵ For more on relations between cognitive penetration and modularity, see Deroy 2013 and forthcoming; Wu, forthcoming; Stokes and Bergeron, unpublished manuscript.

penetrable?-- should be revised (or revived) accordingly. We should instead ask: does cognition affect perception in such a way that one or more of these consequences is realized?

IV. Consequentialism

The basic prescriptive thesis is this: Any analysis of cognitive penetration should be constrained by its consequences. Therefore, an analysis (or if one prefers, a definition) of cognitive penetrability will be successful just in case and to the degree that it describes a phenomenon (or class of phenomena) that has implications for: the rationality of scientific theory choice; the epistemic role of perception; mental architecture.¹⁶ Call this the *consequentialist constraint* on analyses of cognitive penetration. The reader will note that as stated, the consequences are presented as a list; the logical relation between the analysis (or definition) and the three general consequences are in need of elucidation. And indeed the constraint could be satisfied in a number of ways. The remainder of the paper will articulate three such options, and then give reasons for favouring the third. The resulting characterization of cognitive penetration will then be applied to the cases discussed in §II. Finally, the definitions from §I will be re-considered in the light of the consequentialist constraint.

The first option is to stick with traditional conceptual analysis. One constructs an essential definition that describes a phenomenon that results in one or more of the relevant consequences. Here the conditions specified in the definition need not make any one consequence explicit, but instead just describe the relevant mental and causal structures. So long as any one satisfaction of these conditions is co-extensive with a phenomenon that

¹⁶ 'Mental architecture' concerns, in this context, more specifically, the nature and structure of perceptual systems and how they relate to non-perceptual systems.

bears at least one of the relevant consequences, then the consequentialist constraint is met.¹⁷ This option may be tempting, but there are reasons to doubt both the need for and probable success of an essentialist definition. Contemporary philosophy is rife with examples of controversial if not failed attempts at conceptual analysis. So although it might be ideal for some purposes, we have reason to be sceptical that ‘cognitive penetrability’ is amenable to this kind of definition. Perhaps more importantly, if we can do the work by appeal to the consequences, then why bother with the conceptual analysis and corresponding counterexample game?

The second option is what we might call *conjunctive consequentialism*. It says that ψ is cognitive penetration if and only if ψ is a cognitive-perceptual relation that implies consequences for theory-ladenness and the epistemic role of perception and mental architecture. In some ways, this is an improvement over the previous option, since it makes the consequences explicit. However, this is no less an instance of traditional conceptual analysis and so any worries that applied to this logical feature of the first option will apply here as well. More substantively, the conjunctivist option makes an assumption: cognitive penetration is a phenomenon that has all (and only) three general consequences. Accordingly, it rules out the possibility that there are distinct phenomena of interest that imply distinct consequences. There are ways that this deviates from the spirit of the consequentialist proposal. Each consequence matters and, by the very nature of the consequences themselves, for different reasons. And it is not implausible that there are distinct cognitive-perceptual phenomena that imply one consequence but not the other.

¹⁷ A distinct gloss on this option is to make implication of all consequences necessary. This could be done non-explicitly, where non-consequentialist conditions are specified, and any satisfaction of those conditions is co-extensive with implication of all three consequences. Or, it might be done explicitly, as represented by the second option discussed just below.

Here is one such possible scenario. Strong modularity theorists maintain that perceptual modules are encapsulated from other systems in the overall organism. As such, these modules are supposed to perform their computational functions in a way that is independent from, and thus not penetrated by, information in other systems. So, cognitive impenetrability is supposed to be a categorical feature of perceptual modules. Therefore, a single genuine case of cognition affecting, say, perceptual processing would provide a counterexample to encapsulated modules. Now suppose such a case (or even several of them) is found, but the higher level information tokened by the cognitive state is entirely theoretically neutral. In this scenario, one consequence is implied but not another: mental architecture is not as the modularity theorist claims while perceptual observation is, for all we know, theory-neutral. The conjunctivist option rules out this scenario by definition.

Finally, a third option is *disjunctive consequentialism*. It says that ψ is cognitive penetration if and only if ψ is a cognitive-perceptual relation that implies consequences for theory-ladenness or the epistemic role of perception or mental architecture. This best captures the spirit of the consequentialist constraint. Theorists on both sides of the debate (or involved in related, but distinct debates) are interested in these consequences. Moreover, it leaves open the possibility that there are distinct phenomena that meet distinct disjunctive conditions. Of course, one may maintain that what unifies the phenomena is simply that they bear one consequence or more; thus, the consequences unify. And accordingly, one can call any phenomenon that satisfies the cluster condition ‘cognitive penetration’. A virtue of disjunctive consequentialism is that it makes no commitment regarding unified explananda beyond the appeal to consequences. This both makes explicit the most important issues of concern (namely, the consequences of concern to both parties of the debate) but without commitment to a single unified mental phenomenon of interest.

As one may have predicted given some of the discussion above, one does not simply apply the disjunctive analysis to any one case to deliver a verdict. Instead, the inferential procedure remains abductive. For any given case, one considers alternative interpretations and asks what best explains the data. So, one asks, is this data best interpreted in accordance with the memory interpretation, judgement interpretation, attention-shift interpretation, intra-perceptual interpretation, and so on, *or* as a phenomenon that satisfies the cluster condition specified by disjunctive consequentialism? It is important to note that these alternatives are *alternatives* to any interpretation that characterizes a phenomenon as implying one or more of the relevant consequences. So, for example, if a phenomenon is best interpreted as involving an overt shift in attention, or a cognitive effect on memory (but not perception), then there are no relevant consequent concerns about the epistemic role of perception nor about mental architecture vis-a-vis perceptual systems. The standard alternatives thus preclude the consequences. Making this relationship explicit is another virtue of the account. This will become clearer upon application of the consequentialist account to the two empirical cases types discussed in §II.

It should be noted at the outset that the goal here is not to employ the consequentialist account to deliver conclusive verdicts. This is of course the final goal, and for any one case or set of data, extended analysis would be required. But here the ambition is more modest, namely, to offer a sketch of the debate-neutral value of the consequentialist line. Accordingly, discussion of the two types of cases will be brief.

Consider first the New Look and New Look inspired cases involving apparent effects of evaluative attitudes on perception. Here the question is whether this apparent effect is actual: is perception affected by value *in such a way that* more than one of the relevant consequences follow? And, recall, if the empirical results are best explained in one of the alternative ways, then these consequences do not follow. Working through these alternatives,

then, begin with the memory and judgment interpretations. In both control and experimental circumstances, these experiments are online: the perceptual target is available while reports (of a match) are made. Accordingly, it is implausible that the reports are just of subjects' memories. Although more plausible than the memory interpretation, it also seems unlikely that perception is wholly unaffected (e.g. coins are seen accurately) while judgements consistently are erroneous (e.g. coins are reported as significantly larger), since size judgements are made on the basis of current perceptual experience. If this were the effect, we would expect some kind of confusion on the part of the subjects, since they would on this (judgment) interpretation be seeing the two objects (e.g. coin and matching light patch) veridically (where, given the results, the adjusted light patch is significantly larger than the coin), while reporting that the two objects match. No indication of this kind of confusion or surprise shows up in the relevant experiments. There also seems to be insufficient differences between control and experimental circumstances for an explanation involving overt shifts in attention. In the Bruner and Goodman studies, for example, the task is relevantly the same, with target (coin or cardboard cutout analogues) and report circle displaced by 6 inches along the horizontal plane. Any attentional performance that occurs in the experimental case would presumably occur in the control case. Finally, the target stimuli in these experiments--coins, or things like guns in the van Ulzen experiments--are not likely those for which we have a mere perceptually learned response or some kind of perceptual evolution or plasticity. Instead, we learn about these artificial kinds. And this higher-level learning, and the accordant values of those kinds, may affect how we perceive such kinds. Although only a brief treatment, this data looks very much like evidence for a genuine cognitive (or, if one prefers, higher-level) effect on perception. The final question is whether this effect is of relevant consequence.

If the brief analysis just above is apt, then the cases in question do appear to bear the relevant consequences. If non-doxastic states like desires or values influence experience in these ways--by contrast with the alternative interpretations--then both epistemic consequences follow. Perceptual observation in scientific consequences will not be theory-neutral, and the general knowledge providing role of perception will be threatened. Put more neutrally: at the very least, the epistemic role of perception is cast in a new light, and will accordingly require new analysis. The same goes for mental architecture. Strong modularity theorists like Fodor are emphatic on at least this point: the computational processing of perceptual modules is supposed not to be influenced by the organism's goals and values (in addition to beliefs). And, plausibly, this is precisely what is occurring in the relevant cases. The consequence then is that any strong form of perceptual modularity is threatened.

Consider the second kind of case, involving some apparent diachronic change to cognitive-perceptual systems. Here one's perceptual experience apparently changes, but over time, as one learns in some way (for example, one learns how pine trees look). Two points should be stressed here. Appeal to the two considered definitions offers neither clear verdicts on these cases, nor verdicts driven by the same considerations. Second, an often implicit assumption made in the current literature is that these cases are not cases of cognitive penetration simply by virtue of their diachronic nature. But the first point suggests that this assumption is not obviously motivated by appeal to the definitions. So at the very least, the diachronic cases deserve careful consideration.

The alternative interpretations most relevant here are the intra-perceptual interpretation and an attention-based interpretation. As discussed above, Fodor favours an intra-perceptual interpretation of the inverting lenses data. And he might plausibly invoke a similar story for high-level perceptual content changes, perhaps appealing to a story about *perceptual learning* where, on one gloss, repeated exposure to a stimulus kind improves one's

ability (in some way) to perceive that kind. Thus repeated exposure to pine trees results in an increased capacity for distinguishing pine trees from other kinds of stimuli. This explanation can be given entirely, the thought goes, in terms of perceptual systems.¹⁸ Or, one may think that what happens in the diachronic cases is simply that one learns how to better attend to the environment. This kind of change will plausibly often not involve overt attentional action; thus the attention-*shift* interpretation is not appropriate. However, it may be that attentional mechanisms change in their sensitivities as one acquires new information about kinds of objects like pine trees. And so one, rather naturally, attends to those objects differently and as a result of the acquisition of new information (say, concepts or beliefs about pine trees). We might call this, simply, an *attention-based interpretation*. How can the consequentialist approach help in these interpretative contexts?

If Fodor is right about the inverting-lens cases, and this intra-perceptual explanation given can be generalized to the other diachronic cases, then disjunctive consequentialism delivers a clear verdict: lacking a cognitive influence on perception, none of these cases imply the relevant consequences for epistemology or architectures of mind. Note, however, that this verdict has nothing to do with the diachronicity of the cases.

It is less clear, however, that more nuanced attention-based interpretation precludes cognitive penetration. Indeed, a very important question, and one insufficiently discussed in the literature, is whether all kinds of attentional mediation between cognitive state and experience counts against cognitive penetration, or, is there some kind (or kinds) of attentional mediation that is neutral with respect to cognitive penetration? A conclusive verdict, then, is not forthcoming absent substantial further analysis of attention and its bearing on cognitive penetration. But nonetheless some indication of how consequentialism

¹⁸ ‘Perceptual learning’ is, like ‘cognitive penetration’ a commonly used concept but not a singularly defined one. Much could be said here, but space will not allow it in this paper. See E.J. Gibson 1969; Goldstone 1998; Goldstone et al 2010.

might be applied can be offered. Suppose then, for the moment, that some attention-based interpretation of diachronic cases is correct; do any of the relevant consequences follow?

On the attention-based interpretation, new concepts or beliefs determine, in an at least partly sub-personal and non-active way, how one attends to stimuli. This can be understood in at least a couple ways. It may be that when one learns about pine trees, one becomes more attuned to certain perceptual cues. These cues “grab” attention and aid in distinguishing pine trees from other stimuli. Or, put slightly differently, it may be that attentional selection mechanisms change, such that a subconscious mechanism (rather than the agent) selects some features as more important than others, and experience changes accordingly. The first thing to note is that the basic condition of disjunctive consequentialism is satisfied by this picture: there is a cognitive-perceptual relation. The question is whether the mediating attentional mechanisms preclude cognitive penetration. And here the answer should ultimately be given by appeal to the consequences.

There is nothing in this picture that insures negative epistemic consequences--for example a problematic bias towards background scientific theory or towards one's general desires or values about the world. However, if indeed experience is malleable in these ways, even if through the mediation of attentional changes, then epistemic consequences remain on the table. Perceptual modalities, simply, are on this picture more susceptible to cognitive-driven (rather than purely stimulus-driven) change than traditionally theorized. And this at least leaves open the possibility that scientific observation is theory-laden, and that the way that perception provides knowledge is less clean than usually supposed. Neither consequence need be negative--for example, perhaps perception generally becomes more reliable as a result of these attentional changes--but they are plausible consequences no less. More analysis is needed here, but the simple point is that the attention-based interpretation (of

diachronic cases) is not, unlike the standard alternative interpretations, clearly incompatible with cognitive penetration, *so long as* the latter is understood by its consequences.

The same is true for the mental architecture consequence. Modularity has received countless characterizations since it was first given voice by Fodor in the early 1980s. And so whether modularity is threatened by these diachronic cases, interpreted in terms of basic attentional changes, will depend upon the working notion of modularity. Two features that Fodor took to typify perceptual modules is that they were informationally encapsulated and innately specified. Perhaps only implicitly, it is the former feature that may be invoked to dismiss the diachronic cases, since there seems to be no immediate effect on the information processing of perceptual modules in such cases. But it is not clear why an immediate effect should be the only one of concern. For example, if the affected attentional mechanisms are part of the perceptual module, then their being changed over time is still relevant. Put simply, a change to the computational processing of a module is no less a change for its taking place over time. And if the information processing activity of modules change in these ways, then the second mentioned feature is threatened: modules do not appear innately specified in the sense of being *hardwired*, as it is often put. The modularity theorist could of course defend himself at various places here. But the simple lesson is the same as just above: although more analysis is required, there is nothing in the diachronicity of these cases, even when interpreted as involving attention, that precludes the relevant consequence.

Here are two concluding lessons on the two kinds of empirical cases considered, and how they are to be treated by consequentialism. First, as stated at the start of this discussion, neither kind of case has been given a conclusive verdict by disjunctive consequentialism. Instead, a clear debate-neutral strategy has been countenanced, and additional points for analysis have been identified. For example, a complete analysis of the diachronic cases should involve some further discussion of perceptual learning and attention, with one eye

always on how the cases, thus interpreted, do or do not bear the relevant consequences. Second, it should be emphasized that the consequentialist line does deliver the following clear result: whatever else one says about either kind of case, it is not the distinguishing feature of the kind that drives the verdict on cognitive penetration. In both the evaluative cases and the diachronic cases, interpretations are offered whereby the cases come out as cognitive penetration in spite of, respectively, the penetrating state being non-doxastic and the mental effect being a diachronic one. This is progress: it reveals that certain kinds of cases are still relevant to the debate. And they are relevant to the debate not because they clearly satisfy some extant essential definition, but instead because they have prima facie bearing on the debate-neutral consequences. Consequentialism can claim this result as a virtue.

One final way to motivate consequentialism is to see how it relates to existing definitions of cognitive penetration. Recall Pylyshyn's semantic criterion (SC). It claims that cognitive penetration must be an inferentially or representationally coherent relation between penetrating state and penetrated experience. Consequentialism encourages us to reject this criterion, since it is not motivated by the consequences. The epistemic concerns certainly do not require this kind of coherence. Indeed, in the extreme cases, one can imagine scenarios where background cognitive states affect experience in wholly semantically incoherent ways and, accordingly, the epistemic worries would be all the more pressing. Put simply, at least some of the epistemic concern is that perception may be influenced in ways that are irrational, and so a rationality constraint on the phenomenon is ill-placed. The same is true for the mental architecture consequence. Focusing again on modularity, the question is whether perceptual systems are modular in the sense that their information processing is not affected by the processing of independent, cognitive systems. And although it may be difficult for us to understand this kind of situation, perceptual processing *could be* affected in

non-modular ways without this effect being one that is semantically coherent. So here again, SC is simply not motivated by the consequences of concern. Of course one may, as Pylyshyn apparently does, motivate SC by appeal to a broad computationalism about the mind. But this motivation is independent of, and insufficiently neutral with respect to, the concerns of existing debate in cognitive science. Consequentialism helps us see this and, hopefully, better enables progress on the topic.

Conclusion

The virtue of this consequentialist approach is simple: once such an analysis is in hand, empirical studies can be devised, executed, and interpreted accordingly. Testing for cognitive penetration becomes testing for a phenomenon that bears the relevant consequences for the epistemology and architecture of mind. This (re)identifies the phenomenon of interest as one common to both sides of the debate. And it provides a unified metric for assessing the relevance of particular mental phenomena. The motivation for consequentialism, then, is two fold. It provides a clearer alternative to data interpretation--like the two kinds considered in §II and §IV above--than extant definitions. And second, it is debate-neutral, sensitive to the concerns of all relevant theorists.

Indeed, early in the cognitive penetration literature, there is fairly clear precedent for an appeal to consequences.

[I]f you consider the sort of background information that penetrates perception (according to modularity theory), it turns out that perception is neutral, de facto, *with respect to most of the scientific (and, for that matter, practical) disagreements that observation is called upon to resolve*. According to standard versions of modularity theory...perceptual processing has access only to background information about certain pervasive features of the relations between distal layouts and their proximal projections. ...reliance on such information constitutes a perceptual bias...But this bias leaves perception neutral with respect to *almost all* theoretical disputes, so it couldn't ground any *general* argument for the unreliability of observation. (Fodor 1988: 189).

Here Fodor is appealing to epistemic consequences to disarm certain instances of perceptual processing as *non-cognitive* penetration (some of the very cases that Churchland 1988 provides as examples of cognitive penetration). Although not labeled as such, this is consequentialism at work.^{19,20} Similarly, the second definition discussed in §I above is not entirely uninformed by consequences. As it was originally introduced (Stokes 2012), the definition was constructed to describe a phenomenon that is not well-interpreted in any of the standard alternative ways given by critics. And one way to understand these interpretations as *alternatives* is that if they are apt for a given case, then that case implies none of the relevant consequences. For example, a case best interpreted as a cognitive effect on judgement implies no relevant consequences for theories of perception and knowledge, or for modularity of perceptual systems. So although the prescription here is to abandon such essentialist definitions, some of them are more informed than others by consequentialist concerns.

None of this, of course, settles the debate. This was not the ambition of this paper. The ambition was instead to motivate a conceptual strategy for refining theoretical and empirical research on cognitive-perceptual relations of interest to a variety of theorists of

¹⁹ Note that here Fodor is employing the two epistemic consequences to dismiss certain kinds of data. At the same time, however, he is acknowledging some kind of influence on perceptual processing. So whether consequences for modularity follow will importantly depend upon the strength of the posited modules, whether, for example, they must be informationally encapsulated (with respect to the mentioned background information).

²⁰ Similarly, Raftopoulos (2001; 2006) appeals to an epistemic criterion to evaluate certain arguments and empirical evidence for cognitive penetration. Briefly, he argues that early vision is unaffected by anything sufficiently theoretical. And perceptual experience (later vision, if one likes) is affected in a top-down way, but via the control of spatial attention. Because this latter effect is indirect, Raftopoulos argues, the epistemic consequences do not follow. Here the first and second of the three consequences of disjunctive consequentialism--theory-ladenness and the general epistemic consequence--are collapsed into one and used as *the* criterion for cognitive penetration.

mind. The hope is that this strategy--consequentialism--will enable new progress on an important set of questions about the human mind.

Acknowledgments:

Thanks first to the editors for organizing this volume, and providing feedback on this paper. Thanks to Vince Bergeron, Fiona Macpherson, and Wayne Wu for discussion of these and related ideas. And thank you the Utah Philosophy Department, where this paper was given as a department colloquium in Winter 2012; in particular, thanks to Steve Downes, Matt Haber, Eric Hutton, Kim Johnston, Lucas Matthews, Lex Newman, Anya Plutynski, Jonah Schupbach, Cindy Stark, Jim Tabery, and Blake Vernon.

Works cited

Balceris, E. and D. Dunning. 'Wishful Seeing: Desired objects are seen as closer.' *Psychological Science*, 21 (2010): 147-52.

Bayne, T. 'Perception and the Reach of Phenomenal Content' *Philosophical Quarterly* 59 (2009):385-404.

Brewer, W. F., & Lambert, B. L. 'The Theory-Ladeness of Observation and the Theory-Ladeness of the Rest of the Scientific Process.' *Philosophy of Science* 68 (2001): 176-86.

Bruner, J.S., & C.C. Goodman. 'Value and need as organizing factors in perception.' *Journal of Abnormal and Social Psychology* 42 (1947): 33-44.

Churchland, P. M. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press, 1979.

_____. 'Perceptual Plasticity and Theoretical Neutrality: A Reply to Jerry Fodor.' *Philosophy of Science* 55 (1988): 167-187.

Deroy, O. 'Object-sensitivity or cognitive penetration of perception: the case of coloured shapes.' *Philosophical Studies* 162 (2013):87-107

_____. 'Modularity of perception.' in M. Matthen (ed.) *Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press, forthcoming.

Feyerabend, P. 'Explanation, reduction and empiricism.' *Minnesota studies in philosophy of science, Vol. 3* Ed. H. Feigl & G. Maxwell. Minneapolis, MN: University of Minnesota Press, 1962. 28-97.

Fodor, J. *Modularity of Mind*. Cambridge, MA: MIT Press, 1983.

_____. 'Observation Reconsidered', *Philosophy of Science* 51 (1984) 23-43.

_____. 'Precis of The Modularity of Mind', *The Behavioural and Brain Sciences* 8 (1985): 1-5

_____. 'A reply to Churchland's 'Perceptual plasticity and theoretical neutrality.' *Philosophy of Science* 55 (1988): 188-198.

Gibson, E.J. (1969). *Principles of perceptual learning and development*. East Norwalk, CT, US: Appleton-Century-Crofts.

Goldstone, R. 'Perceptual Learning.' *Annual Review of Psychology* 49 (1998) 585-612.

Goldstone, R. L., Landy, D. H., & Son, J. Y. 'The education of perception.' *Topics in Cognitive Science* 2 (2010) 265-284.

Hanson, N. R. *Patterns of Discovery*. Cambridge: Cambridge University Press, 1958.

_____. *Perception and Discovery: An Introduction to Scientific Inquiry*. Freeman and Cooper, 1969.

Hawley K. and F. Macpherson. *The Admissible Contents of Experience*. London: Wiley-Blackwell, 2011.

Kottenhoff, H. 'Situational and Personal Influences on Space Perception With Experimental Spectacles.' *Acta Psychologica* 13 (1957): 79-97.

Kuhn, T. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press, 1962.

Lyons, J. 'Circularity, Reliability, and Cognitive Penetrability of Perception.' *Philosophical Issues* 21 (2011): 289-311

Macpherson, F. 'Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism.' *Philosophy and Phenomenological Research* 84 (2012): 24-62.

McCurdy, H. G. 'Coin perception studies and the concept of schemata,' *Psychological Review* 63 (1956): 160-168.

Prinz, J. 'Putting the Brakes on Enactivism.' *Psyche* 12 (2006) 1-19.

Pylyshyn, Z. 'Computation and Cognition: Issues in the Foundations of Cognitive Science.' *Behavioral and Brain Sciences* 3 (1980): 111-32.

_____. *Computation and Cognition*. Cambridge, MA: MIT Press, 1984.

_____. 'Is vision continuous with cognition? The case for cognitive impenetrability of visual perception.' *Behavioral and Brain Sciences* 22 (1999): 341-365.

Raftopoulos, A. 'Reentrant Pathways and the Theory-Ladenness of Perception.' *Philosophy of Science (Proceedings)* 68 (2001): 187-199.

_____. 'Defending Realism on the Proper Ground,' *Philosophical Psychology* 19 (2006): 47-77.

Schwitzgebel, E. 'The problem of known illusion and the problem of unreportable illusion.' (unpublished manuscript).

Siegel, S. 'Which Properties are Represented in Perception?' *Perceptual Experience*. Ed. T. Szabo Gendler and J. Hawthorne. Oxford: Oxford University Press, 2006. 481-503.

_____. 'The Visual Experience of Causation.' *Philosophical Quarterly* 59 (2009):519-540.

- _____. 'Cognitive Penetrability and Perceptual Justification.' *Nous* (2011).
- _____. 'The Epistemic Impact of the Etiology of Belief.' *Philosophical Studies* (forthcoming).
- Stokes, D. 'Perceiving and Desiring: A new look at the cognitive penetrability of experience.' *Philosophical Studies* 158 (2012): 479-92.
- _____. 'The cognitive penetrability of perception' *Philosophy Compass* (2013).
- _____. 'Cognitive penetration and the perception of art' (unpublished manuscript)
- Stokes, D. and Bergeron, V. 'Modular architectures and informational encapsulation: A dilemma.' (unpublished manuscript)
- Stratton, G. M. 'Vision without inversion of the retinal image.' *Psychological Review* 4(5), (1897): 463-481.
- Taylor, J. *The Behavioral Basis of Perception*. New Haven: Yale University Press, 1962.
- van Ulzen, N.R., G.R. Semin, R. Oudejans, & P. Beek. 'Affective stimulus properties influence size perception and the Ebbinghaus illusion.' *Psychological Research* 72 (2008): 304–310.
- Van Gulick, R. 'Deficit Studies and the Function of Phenomenal Consciousness.' *Philosophical Psychopathology*. Ed. G. Graham and G.L. Stephens. Cambridge, Mass: MIT Press, 1994. 25-49.
- Walton, K. 'Categories of Art' *Philosophical Review* 79 (1970): 334-67.
- Wu, W. 'Visual Spatial Constancy and Modularity: Does Intention Penetrate Vision?' *Philosophical Studies* (forthcoming).