

Seeing-As in the Light of Vision Science

Ned Block
NYU

Introduction

I will discuss the issue of what we see, using the framework of Tyler Burge's wonderful and monumental book *Origins of Objectivity* (Burge 2010).

On Burge's view, every percept is constituted by a "perceptual attributive" (that represents an attribute) and a singular element (that represents an individual). The format of a percept is map-like or iconic, and could be symbolized as "That F".¹ The most basic perceptual judgments and beliefs are conceptualized and propositionalized versions of percepts in which the attributive need not be bound to a time and place.

What are the actual perceptual attributives? Vision science picks out a small set of basic low level attributives that are products of sensory transduction and are causally involved in the production of other visual attributives: shape, spatial relations (including position and size), geometrical motion, texture, brightness and color. Burge discusses a higher level attributive for objects ("integrated body") and considers that there may be some higher level attributives for some biologically important properties like food, danger and shelter. However, he notes (and I agree) that probably there are no culture-specific higher level attributives for teacups and recessions (p. 101). Terminological note: I will abbreviate the list of low level environmental attributives with "color, shape and texture" and I will refer to all attributives that are not low level as "high level".

Seeing and Concepts

One tradition in philosophy holds that all seeing is conceptual. Brian O'Shaughnessy (O'Shaughnessy 2012) argues (p. 42) that according to Wittgenstein, "...the work of the Understanding lies at the center of visual perceptual experience." This point of view appears in a different form in Jerry Fodor's forthcoming review of Burge's book (paraphrased with permission). Fodor's argument is: (1) No seeing without seeing as; (2) No seeing as without conceptualization; (3) No conceptualization without concepts. So seeing is inherently conceptual. Fodor has justified (2) on the ground that "...perception involves constancies and constancies require inferences, so those visual attributions involve concepts." (Fodor 2007).

Since Helmholtz, computations whose successful operation depends on facts about the world have been called "inferences" but this is a metaphorical use of the term that does not justify the attribution of concepts to the visual system. Premise 2 is false. Burge refutes the idea by appealing to the fact that we share

¹ More fully, the singular element in the percept is an occurrent context-bound application of 'that' referring to a non-repeatable property-instance such as an object or event or a "trope".

some perceptual attributives with lower animals, even insects. (Note that insects have color vision.) Further, given that we know that some insects can see but we do not know (and indeed doubt) that they are capable of conceptual thought, we should not suppose that there is any a priori argument that visual attribution requires concepts.

'Concept' is used in different ways so this disagreement may seem verbal. However, Burge, Fodor and I would agree that a concept is constitutively apt to function in a propositional structure. And we would agree on a difference in computational role between percepts and concepts. Roughly: perception is informationally encapsulated and concepts play an inferentially promiscuous role in cognition. Burge and I think there is a format difference as well: perception is iconic. In what follows I will be less concerned with characterizing the difference between concepts and percepts than with explaining how to tell whether there is a form of seeing-as that is comfortably on the perception side of the perception cognition joint in nature.

I do not have the space to discuss controversies over whether there is in fact a joint in nature between perception and cognition (but see (Firestone 2013; Firestone and Scholl 2013)) except to note that such a joint is compatible with causation from one to the other and with the existence of borderline cases. There is a joint in nature between living and non-living things but also borderline cases such as viruses and prions.

Acquiring New Visual Attributives?

Endre Begby describes Burge's view as a "peculiarly reductive account of what we are capable of perceiving, properly speaking" (Begby 2011). One thing that is supposed to be reductive is that there are no visual attributives specific to teacups and CD players.

The claim that there are culture-specific perceptual attributives is often based on the phenomenological consequences of perceptual expertise—e.g. learning to recognize pine trees changes our experience of them (Siegel 2010). However, there is an alternative explanation of this effect in terms of *low level* attributives. Early vision is highly influenced by contours in the environment. For example, Mary Peterson has shown that viewers are more likely to see the white part of the left item of Figure 1 as figure (rather than ground) compared to the white part on the right item (Peterson, Harvey et al. 1991). (The border appears to shape the figure but the ground appears shapeless, which suggests something for which there is ample additional evidence (Peterson and Skow 2008), that the figure-shape suppresses the ground-shape.) The difference between the left and right no doubt reflects greater familiarity with contours of standing women in western dress than inverted standing women .

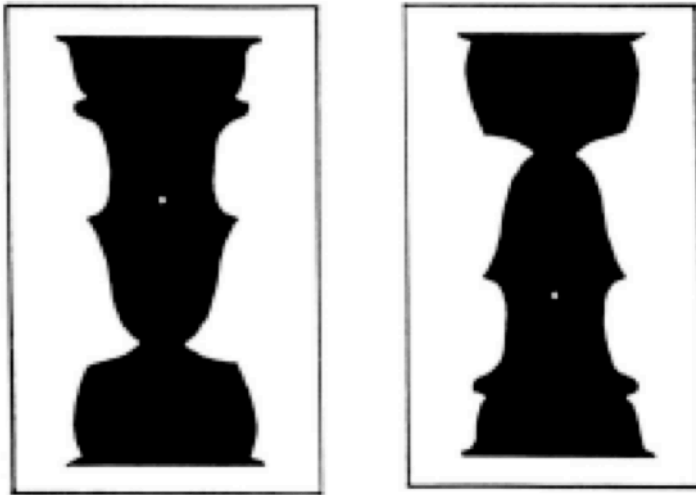


Figure 1

Caption to Figure 1: Viewers are more likely to see the white parts of the left item as figure than for the inverted version on the right though Gestalt principles favor the small, closed, convex and symmetric black portions in both cases. From (Peterson, Harvey et al. 1991) courtesy of the American Psychological Association.

End of caption-----

Recognitional Coextension

If moderately complex pictures—e.g. scenes or faces—are presented one after another with no advance information, subjects can perceive them moderately well at a rate of 6 per second and better than chance at 12 per second (Potter, Staub et al. 2004; Potter, Wyble et al. 2012). We must do this on the basis of bottom-up processing of low level features since there is no time for a substantial top-down influence (VanRullen and Koch 2003). Are there clusters of low level features that are what I will call “*recognitionally coextensive*” with, e.g. faceness that is, coextensive with faceness-to-the-extent-that-we-can-recognize-it? If so—as it must be for quick context-less recognition to be possible—the question arises as to how we can possibly distinguish the use of a face-attributive from a recognitionally coextensive conger of low level attributives.

Susanna Siegel (Siegel 2010) has argued for culture-specific perceptual attributives and even perceptual attributives for named individuals, e.g. being John Malkovich. Her “method of phenomenal contrast” appeals to the best explanation of pairs that differ phenomenally, and so in principle allows for experimental evidence as part of the best explanation, but when she actually applies the method, only armchair considerations are used. However, only empirical evidence can distinguish among the following: (1) complexes of low level attributives, (2) the high level attributives that are recognitionally coextensive with them, (3) conceptual attributives that are part of perceptual judgments rather than perceptions themselves.

I will discuss three empirical phenomena. The first will provide prima facie

evidence for facial expressions being perceptual attributives but will not settle the recognitional coextension issue. The second will show that there are high level attributives for some aspects of face perception but will not settle the perceptual rather than cognitive nature of those attributives. The final case will consider the issue of whether the technique used in the first two discussions really isolates perceptual attributives as opposed to cognitive attributives instantiated in an occurrent perceptual judgment. Unfortunately these three cases concern three different classes of attributives, the last is not high level and none of the high level attributives are cultural along the lines of teacups and CD-cases. My concern is more with the methodology of answering the questions rather than the actual answers.

That methodology is based on perceptual adaptation, what John Frisby (from an idea of John Mollon) called the “psychophysicists’ microelectrode” (Mollon 1974; Frisby 1979). Just as the neuroscientist can first raise, then reduce a neuron’s firing rate by direct stimulation with a microelectrode, the psychophysicist can first raise, then reduce a neural system’s activity by stimulating it with its preferred stimulus. Perceptual adaptation was known to Aristotle who described (in “On Dreams”) what we now call the “waterfall illusion” in which “...when persons turn away from looking at objects in motion, e.g. rivers, and especially those which flow very rapidly, things really at rest are seen as moving” (Aristotle 1955, p. 731). Staring at something moving down raises the threshold for detecting downward motion, biasing the percept towards upwards motion, so stationary things look like they are moving upwards .

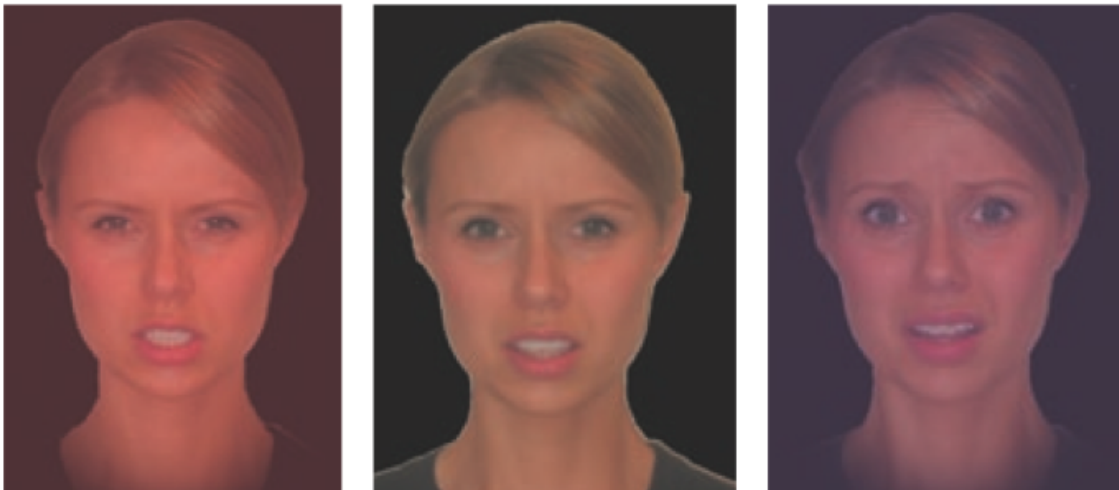


Figure 2

Caption to Figure 2: Cover the two pictures on the left with a blank piece of paper. Stare intently at the picture on the right for 1 minute. Then look at the center picture. Now cover the two pictures on the right and stare intently at the picture on the left for one minute. Now look at the center picture again. It will appear to have a different expression. Similar experiments show adaptation that is at least coextensive with facial aspects of race, gender and individual identity. From (Butler, Oruc et al. 2008) with

permission from Elsevier.

End of caption-----

The middle picture in Figure 2 is ambiguous between anger and fear. When one stares at the fearful face on the right, the threshold for firing of neural systems (Butler, Oruc et al. 2008)(Butler, Oruc et al. 2008)(Butler, Oruc et al. 2008) that code for fearfulness is raised, so the perception of the middle picture is biased towards anger. And the opposite happens when one stares at the angry face first. Similar phenomena occur for other facial expressions, face identity and face gender.

This phenomenon grounds a prima facie case that we have visual attributives for facial expressions. And there is additional evidence that the effect is not mainly due to recognitionally equivalent low level attributives. One can vary the low level properties and so long as the emotional expressions are kept constant, adaptation obtains, though diminished somewhat, suggesting that face perception utilizes both low and high level attributives. It would take baroque congeries of low level properties to explain this fact (Butler, Oruc et al. 2008). (See also Rossion and Boremanse (2011)).

A further line of evidence (Susilo, McKone et al. 2010) is based on two ideas. The first idea is to compare adaptation effects between inverted and upright faces on the assumption that adaptation effects that work for inverted faces are likely to derive from low level attributives, whereas extra adaptation for upright faces is likely to involve high level perceptual attributives specific to faces. Susilo, et. al. note that adaptation for height transfers from one shape to another. For example, staring at an elongated ellipse (the “adaptor”) makes a rectangle (“adaptee”) look shorter and staring at a short adaptor ellipse makes an adaptee rectangle look longer. Susilo et. al. designed an experiment that examined transfer of adaptation from the letter ‘T’ of various heights to faces whose eye to mouth distance also varied as shown in Figure 3 and also the reverse transfer of adaptation. They used adaptor items (both ‘T’s and faces) of three different elongations, testing the effects of these differences in adaptors on a variety of elongations of “adaptee” faces and ‘T’s.

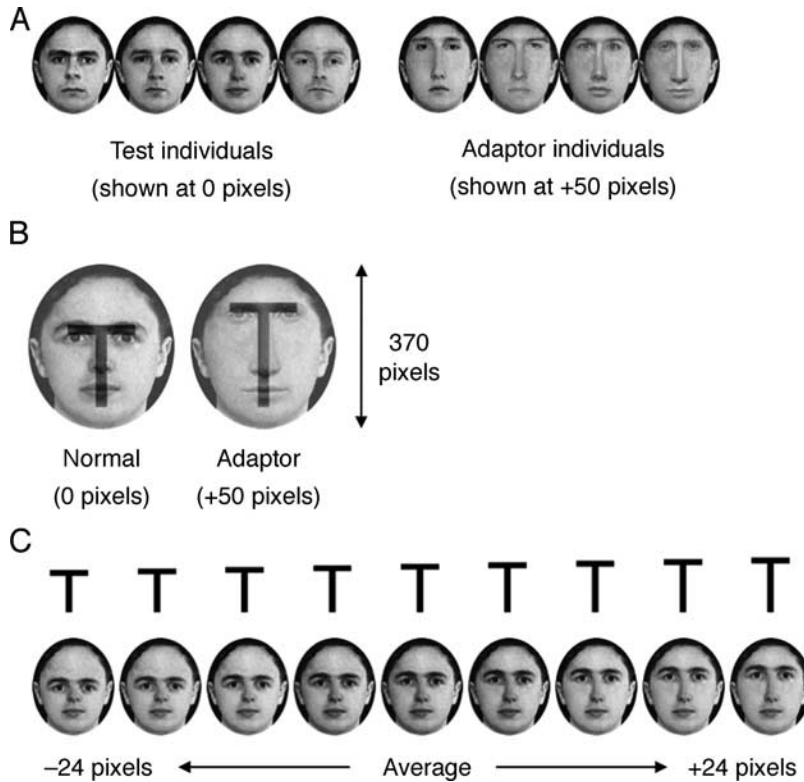


Figure 3

Caption to Figure 3: The most elongated faces are shown in right side of (A). The pixels mentioned corresponded to .29% of full head height. A short and long face and 'T' are shown in (B). (C) has 'T's and faces of various different elongations used as the adaptees. From (Susilo, McKone et al. 2010) with permission of the Association For Research in Vision and Ophthalmology.

End of caption-----

The technique was to ask the subject to stare at a face or a 'T' of one of the three elongations for 4 seconds, then to view a face or a 'T' of one or another elongation, judging whether the test item was longer or shorter. A subject who has adapted to an elongated stimulus will see another stimulus as shorter than it would otherwise have looked, so to the extent that perception of upright and inverted faces is low level, transfer of adaptation *within* stimulus types should be the same as *between* stimulus types. Using 'F-F' to mean transfer from face to face, there are four types of transfers that should all be the same if perception is entirely low level, i.e. $F-F = F-T = T-T = T-F$. But if upright face perception and aftereffects derive only from face-specific attributives, there should be no transfer in either direction between faces and 'T's.

The results were that for inverted stimuli, F-F, F-T, T-T and T-F were nearly the same, only 8% of the aftereffect was face-specific. Inverted faces are seen almost entirely via low level shape-general attributives. In the case of upright faces, 55% of the aftereffect was face-specific and 45% was low level. So high

level face-specific attributives play a slightly larger role than low level attributives. Of course no single result can rule out that the result is due to differences between low level features of right-side-up and upside-down faces (e.g. the downward rather than upward curve of the eyebrows). These results show us how evidence can bear on what attributives the visual system actually uses.

But do these results concern a technical sense of 'see'? Tim Williamson asked me in a presentation of this material whether in this sense of 'see' one can see New College. It is useful to distinguish primary from secondary seeing. We can take primary seeing to be the application of a visual attributive to a "visual object", i.e. an object that is itself picked out by a demonstrative element in a percept of the sort described in the 2nd paragraph of this article. Secondary seeing involves hybrids of visual attributives and concepts applied to objects of primary seeing and complexes of them in states that put together perception with perceptual judgment. Such hybrids can represent things that are not visual objects (i.e. not the referents of the demonstrative elements in perceptions) on the basis of visual objects that compose them. Thus the notion of seeing in which we see New College is secondary seeing and it can be reconstructed in terms of the sciences of perception and cognition.

Perception vs Cognition

The face in the middle in Figure 2 *looks* first angry then fearful. But can we be sure from introspection that those "looks" are really perceptual, as opposed to primarily the "cognitive phenomenology" of a conceptual overlay on perception, that is, partly or wholly a matter of a conscious episode of perceptual judgment rather than pure perception? Moving to the Susilo experiment, if a subject sees the adaptee as less elongated than the adaptor, how do we know whether this case of seeing-as is purely perceptual as opposed to partly or wholly a matter of perceptual judgment? (One might raise this worry separately for the adaptor and adaptee.)

There is reason to think that concepts don't adapt in the way percepts do. Ambiguous figure-ground stimuli (see Figure 1) engender oscillating percepts that change every few seconds. Both behavioral and neural evidence indicate that the figure representation suppresses the ground representation, with the strongest suppression directed towards the strongest competitors (Peterson and Skow 2008; Salvagio, Cacciamani et al. 2012). The suppression is short-lived, putting a switch of figure and ground within the realm of possibility, but the switching itself is a result of adaptation (Alais, O'Shea et al. 2010). Such forms of multistable perception have three fundamental properties (Logothetis and Leopold 1999): exclusivity (the multistable percepts are not present simultaneously, although transitions can have elements of both), inevitability (one interpretation will eventually replace another) and randomness (the duration of one alternation is not a function of previous durations). Of course figure-ground alternation can be modulated somewhat via intentions, and there are ambiguous figures where cognition makes a difference to perception (Potter 1964), but these facts do not show conceptual adaptation.

In displays like those in Figure 1, subjects obtained—as figure-- the meaningful shapes like the white shape on the left faster and were able to intentionally maintain them longer only when the figures were roughly upright rather than rotated or inverted—and crucially, *whether or not they knew what the shapes were* (Peterson and Gibson 1994). The orientation-dependence shows the adaptation is at least partly perceptual, and the independence of what the subjects know suggests the adaptation is not cognitive. So at least some kinds of figure-ground alternations are perceptual, not conceptual. Further, the point of perception is to register news, and adaptation contributes to that by filtering out old news; old news however, is important to cognition, so one would not expect adaptation in cognitive systems.

Other things equal, if there were conceptual adaptation one would expect conceptual-without-perceptual adaptation—e.g. conceptually ambiguous situations would show alternations. For example, a morally ambiguous situation might lead to alternating interpretations that show the same three fundamental properties. No such phenomena have been reported to my knowledge.

I conclude that there is evidence that the alternations observed for facial expressions and other aspects of faces (e.g. identity and gender) are perceptual in nature. And there are other types of evidence that can at least in principle be used in deciding the question in a particular case of whether adaptation is perceptual.

If the locus of adaptation is in cortical visual areas, that would provide some evidence that it is perceptual, but even then there could be a cognitive element. For example, working memory, an at least partly cognitive capacity, makes use of neural coalitions involving perceptual representations that provide part of the contents of working memory representations (Curtis and D'Esposito 2003).

I will describe evidence that in one case adaptation is wholly perceptual. The technique compares adaptation with priming, an effect that works in the opposite direction from adaptation. Adaptation to a feature makes that feature harder to see; priming makes it easier. For example, exposure of a tenth of a second to a stimulus moving to the left biases the viewer towards seeing an ambiguous stimulus (ambiguous between rightward and leftward motion) as moving leftwards, so long as the ambiguous stimulus is presented immediately afterward (Kanai and Verstraten 2005). Priming of this sort is thought to result from briefly spreading activation in the visual system.

Caspar Schwiedrzik and Lucia Melloni investigated adaptation and priming simultaneously in perception of tilt. Subjects first saw oriented grids of the sort of the leftmost or rightmost grid in Figure 4 for 4 seconds, then indicated which direction of tilt they saw. Then they saw a grid with ambiguous tilt (like the middle one) and indicated its direction of tilt.

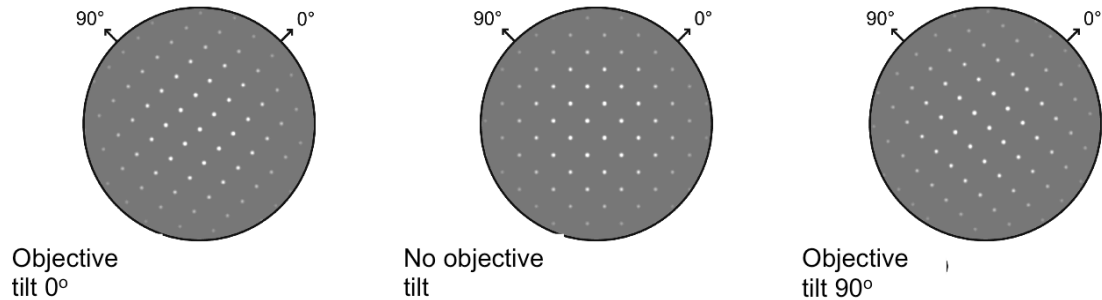


Figure 4

Caption to Figure 4: The leftmost grid is more likely to be seen as oriented to the northeast (stipulated to be 0°) than the northwest whereas the rightmost grid is northeast (90°), and the middle grid is multi-stable as between the two. This figure is similar to part of Figure 1 in (Schwiedrzik, Ruff et al. 2013). With permission from Oxford University Press, publishers of *Cerebral Cortex*.

End of caption-----

The idea of the experimental design is that an attractive effect of perception of the first grid on perception of the second (ambiguous) grid indicates priming, whereas a repulsive effect indicates adaptation. The first result is unsurprising: the more 0° -ish the first grid was objectively, the more likely subjects were to classify it as 0° . See Figure 5A which maps the probability of reporting 0° on the first grid (vertical axis) against the objective tilt of the first grid (horizontal axis).

Priming is exhibited in Figure 5B by the black line being above the gray line: subjects were more likely to classify *the second (ambiguous) grid* as 0° if they classified the first grid as 0° . Higher on the vertical axis indicates a greater probability of reporting 0° on the second grid and the black line reflects subjects for whom that the first grid was reported as zero. This difference between the black line and the gray line shows attraction between the first report and the second report: priming.

Most importantly and surprisingly: the more 90° -ish the *first grid* was *objectively*, the *more* likely subjects were to classify the *second* grid as 0° . This is an adaptation effect and is indicated by the upward slope of both lines in Figure 5B as compared with the downward slope of the line in Figure 5A. Crucially this effect held both for those who classified the first grid as 0° and those who classified the first grid as 90° . The black and gray lines are parallel, so the adaptation effect worked in the same way independently of *what the subjects thought they saw*: what was important for adaptation was not how the subjects took themselves to have seen the first grid but the actual objective tilt of the first grid as registered in the visual system.

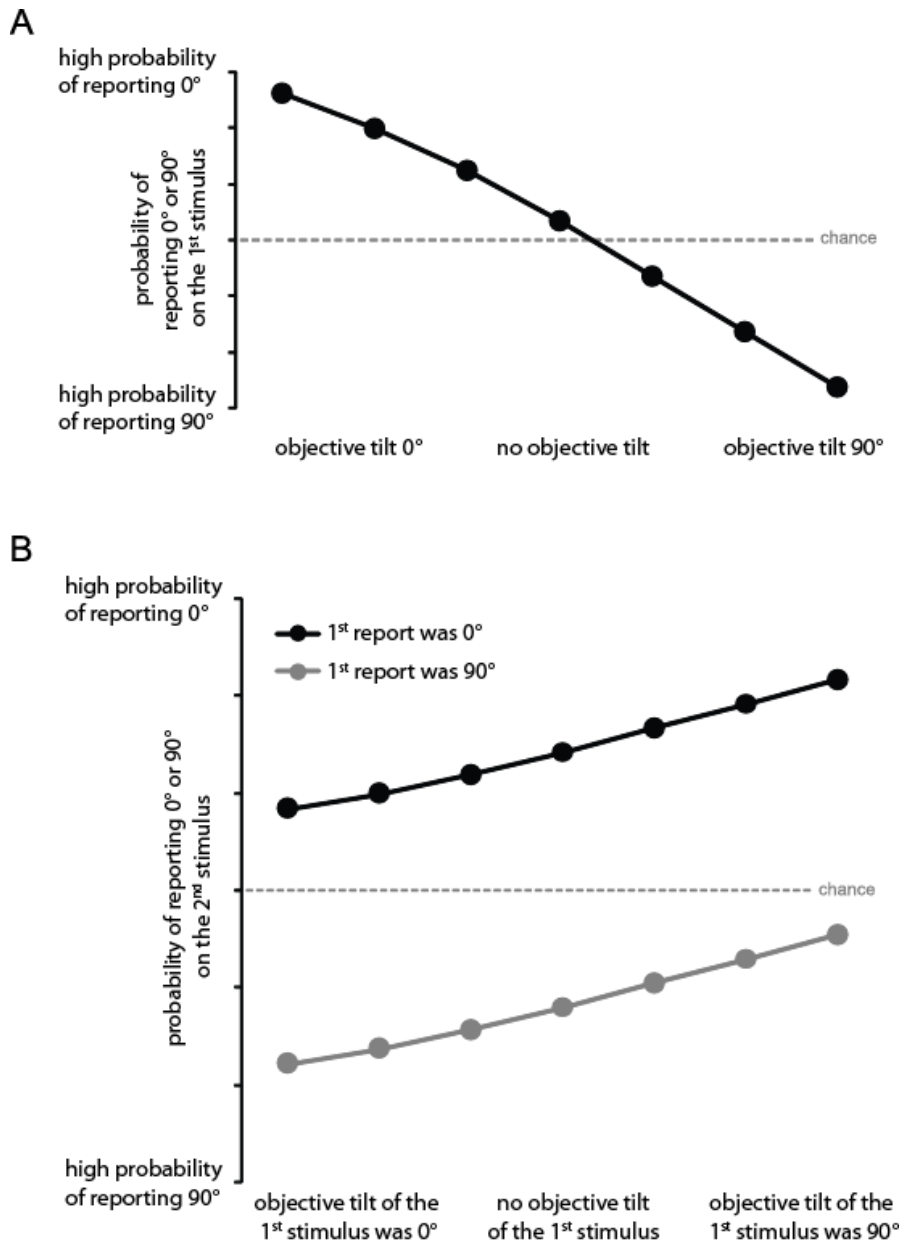


Figure 5

Caption to Figure 5: See text. This figure is similar to Figure 2 in (Schwiedrzik, Ruff et al. 2013). I am grateful to Caspar Schwiedrzik for adapting this figure to the needs of this article. With permission from Oxford University Press.

End of caption-----

Since adaptation does not depend on what the perceiver judges, it is hard to see how it could be a cognitive effect. The upshot for our discussion of face adaptation is that at least one kind of adaptation is probably perceptual and not cognitive.

Schwiedrzik and Melloni also scanned the subjects, finding that priming engaged a wide range of brain areas in the front and middle of the brain whereas adaptation engaged only early visual areas (V2 and V3) in the back of the brain. What is significant here—given the earlier remark about working memory—is that the adaptation does not involve the frontal and parietal areas that are thought to underlie concepts and cognition. They conclude that “adaptation is a purely local phenomenon...” What I like about this experiment is that there is a mesh between the psychological and neural evidence and both suggest that—at least for the admittedly simple stimuli used in this experiment— adaptation is perceptual rather than cognitive.

Thus we can distinguish empirically among (1) a high level perceptual attributive, (2) the holistic constellation of color, shape and texture attributives that are recognitionally coextensive with that attributive and (3) the conceptualized attributives that play a role in conscious episodes of thought and judgment rather than perception.

Direct realists often claim, on a priori grounds, that representations in perception are all sub-personal (Travis 2004, p. 59). This view is a mistaken product of a deeply wrongheaded armchair methodology. I have described—unfortunately only briefly—evidence that we have visual representations of fearfulness of faces and similar experiments show visual representations of other emotions and of gender and race. These representations are both perceptual and play a direct role in relation to personal level interests and values. That makes them personal, not sub-personal. Painting with an infamously broad brush, Wittgenstein (1958, p. 232) complains of the “confusion and barrenness of psychology” saying “... in psychology there are experimental methods and conceptual confusion.” (As with many of Wittgenstein’s less wise remarks, Wittgensteinians (Hacker forthcoming) have tried to show that this quotation means something other than what it says.) However, the real confusion lies with philosophers who try to settle the nature of seeing-as in ignorance of the sophisticated conceptual apparatus that has been developed in psychology to approach such issues.

Is this Philosophy?

This article reports the results of a number of psychological experiments and so it may occur to philosophers reading it to wonder whether this article—and the chapters of Burge’s book that appeal to facts about perception—are philosophy rather than psychology. Both are both! Philosophy of mind is a mix of empirical and relatively a priori considerations, both of which can be seen both as philosophy and as theoretical psychology. However, it would be a mistake to

think that those who know nothing of the science of the mind can just stick to the relatively a priori parts of philosophy of mind, since one needs to understand the empirical facts to even know where there is room for relatively a priori philosophy. Burge's book is masterful in matching a priori and empirical methods to the issues that are appropriate for them.²

Bibliography

- Alais, D., R. P. O'Shea, et al. (2010). "Visual sensitivity underlying changes in visual consciousness." Current Biology **20**: 1362-1367.
- Aristotle (1955). On Dreams. Aristotle. Parva naturalia. W. D. Ross. Oxford, Clarendon Press.
- Begby, E. (2011) "Review of Origins of Objectivity." Notre Dame Philosophical Reviews **2011**.
- Burge, T. (2010). Origins of Objectivity. Oxford, Oxford University Press.
- Butler, A., I. Oruc, et al. (2008). "Factors contributing to the adaptation aftereffects of facial expression." Brain Research **1191**: 116-126.
- Curtis, C. and M. D'Esposito (2003). "Persistent activity in the prefrontal cortex during working memory." Trends in Cognitive Sciences **7**(9): 415-423.
- Firestone, C. (2013). "How "Paternalistic" Is Spatial Perception? Why Wearing a Heavy Backpack Doesn't--and Couldn't-- Make Hills Look Steeper." Perspectives on Psychological Science **8**(4): 455-473.
- Firestone, C. and B. J. Scholl (2013). ""Top-down" effects where none should be found: The El Greco fallacy in perception research." Psychological Science.
- Fodor, J. (2007). The Revenge of the Given. Contemporary Debates in Philosophy of Mind. B. McLaughlin and J. Cohen. Oxford, Blackwell: 105-117.
- Frisby, J. (1979). Seeing: Illusion, Brain and Mind. Oxford, Oxford University Press.
- Hacker, P. M. S. (forthcoming). The relevance of Wittgenstein's philosophy of psychology to the psychological sciences. Proceedings of the Leipzig Conference on Wittgenstein and Science, 2007.
- Kanai, R. and F. Verstraten (2005). "Perceptual manifestations of fast neural plasticity: motion priming, rapid motion aftereffect and perceptual sensitization." Vision Research **45**(25-26): 3109-3116.
- Logothetis, N. and D. Leopold (1999). "Multistable phenomena: changing views in perception." Trends in Cognitive Sciences **3**(7): 254-264.
- Mollon, J. (1974). "After-effects and the brain." New Scientist **61**(886): 479-482.
- O'Shaughnessy, B. (2012). Seeing an Aspect and Seeing Under an Aspect. Wittgenstein and Philosophy of Mind. J. Ellis and D. Guevara. Oxford, Oxford University Press: 37-60.
- Peterson, M. and B. Gibson (1994). "Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours." Perception and Psychophysics **56**(5): 551-564.

² Thanks to Jake Beck, Endre Begby, Tyler Burge, Susan Carey, Grace Helton, Rob Hopkins, Caspar Schwiederzick and Susanna Siegel for comments on an earlier draft.

- Peterson, M., E. Harvey, et al. (1991). "Shape Recognition Contributions to Figure-Ground Reversal: Which Route Counts?" Journal of Experimental Psychology: Human Perception and Performance **17**(4): 1075-1089.
- Peterson, M. and E. Skow (2008). "Inhibitory Competition Between Shape Properties in Figure-Ground Perception." Journal of Experimental Psychology **34**(2): 251-267.
- Potter, M. C. (1964). "Inference in Visual Recognition." Science **144**(3617): 424-425.
- Potter, M. C., A. Staub, et al. (2004). "Pictorial and conceptual representation of glimpsed pictures." Journal of Experimental Psychology: Human Perception and Performance **30**: 478-489.
- Potter, M. C., B. Wyble, et al. (2012). "Post-detection at 13 ms/picture in RSVP." Journal of Vision **12**(118).
- Rossion, B. and A. Boremanse (2011). "Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady state visual-evoked potentials." Journal of Vision **11**(2): 1-21.
- Salvagio, E., L. Cacciamani, et al. (2012). "Competition-strength-dependent ground suppression in figure-ground perception." Attention, Perception & Psychophysics **74**: 964-978.
- Schwiedrzik, C., C. C. Ruff, et al. (2013). "Untangling Perceptual Memory: Hysteresis and Adaptation Map into Separate Cortical Networks." Cerebral Cortex.
- Siegel, S. (2010). The Contents of Visual Experience. Oxford, Oxford University Press.
- Susilo, T., E. McKone, et al. (2010). "Solving the upside-down puzzle: Why do upright and inverted face aftereffects look alike?" Journal of Vision **10**(13): 1-16.
- Travis, C. (2004). "The Silence of the Senses." Mind **113**(449): 57-94.
- VanRullen, R. and C. Koch (2003). "Visual Selective Behavior can be Triggered by a Feed-Foward Process." Journal of Cognitive Neuroscience **15**(2): 209-217.
- Wittgenstein, L. (1958). Philosophical Investigations, Second Edition. Oxford, Blackwell.