Seeing and Conceptualizing: Modularity and the Shallow Contents of Perception

ERIC MANDELBAUM

CUNY Graduate Center and Baruch College

After presenting evidence about categorization behavior, this paper argues for the following theses: 1) that there is a border between perception and cognition; 2) that the border is to be characterized by perception being modular (and cognition not being so); 3) that perception outputs conceptualized representations, so views that posit that the output of perception is solely non-conceptual are false; and 4) that perceptual content consists of basic-level categories and not richer contents.

1. Categorization and Architecture

Where does conceptualization occur? When categorizing a visually presented stimulus, does categorization happen during perception, or is it post-perceptual? As I will use the term, "conceptualists" are theorists who maintain that categorizing a visually presented stimulus (henceforth: categorizing) occurs perceptually, so that perception outputs an already conceptualized representation. In contrast, "non-conceptualists" propose that the outputs of perception are non-conceptual representations, which entails that categorization must occur post-perceptually.

As normally understood by psychologists, categorization just is the process of applying[1] (/predicating) a concept to a preexisting representation. The question of where categorization occurs thus bears on some aspects of the non-conceptual content debate. If categorization occurs intraperceptually, then perception outputs representations with conceptual content. Hence my use of "conceptualist" and "non-conceptualist." But questions of categorization also bear a surprisingly overlooked relation to another enormous topic in cognitive science—the architectural question of the autonomy of perception from cognition.

Recently there has been renewed support for the idea that there is a joint in nature separating perception from cognition (Carey 2009; Burge 2010; Block 2014; Firestone and Scholl 2014, 2015, forthcoming; Raftapolous 2015). The properties that are posited to distinguish perception and cognition differ depending on the theorist. Recent well-publicized proposals have posited that the hallmark of perception is that it utilizes iconic

---

[1] Perhaps it is better to say that concepts are applied to things in the world, but predicated of other representations. Nevertheless, I will use "apply" instead of "predicate" as to not beg any questions against those who think that all thought is purely associative (and thus non-predicative).

representations (Carey 2009; Burge 2014; Block 2014). In particular, this position amounts to claiming both that intraperceptual computations only utilize iconic representations, and that the representations that perception creates, those which serve as inputs to cognition, are iconic. It is also often claimed that perceptual representations are non-conceptual. If so, then identification—that is, visual classification—has to occur post-perceptually.

Of course, another venerable view distinguishes perception from cognition by holding that perception is modular.[2] Here perception and cognition are distinguished not, per se, by the format of the representations in the systems, but instead via informational criteria (*inter alia*). In classical modularity theory, perception is accomplished via domain-specific modules which contain proprietary information. The modules are domain-specific because they implement computations which are specific to the contents that they process. The module's information and computations are proprietary because they are informationally encapsulated from the rest of cognition. That is, each module's specific computational processes only have access to the module's proprietary database (along with the input to the module) and nothing more. In contrast, cognition has no proprietary algorithms, and no informational restrictions on its contents.[3] Computational processes in central cognition can compute any non-modularized information.

In the classic presentation of modularity theory Fodor (1983) speculated that modules have "shallow contents." The idea is that the outputs of modules tend to be representations of the "basic-level" objects that Rosch discovered (Rosch 1978). A question that naturally arises is whether those basic-level representations are conceptual. At first glance they certainly seem to be: after all, if the outputs of perception are showing the same prototype and basic-level naming effects of Rosch (ibid.), then they are acting like concepts (indeed many theorists hold that prototypes just are concepts, Prinz 2002; Smith and Minda 2002; Jonsson and Hampton 2007). But this needn't necessarily be so: the non-conceptualist could hold that applying concepts to iconic perceptual representations happens post-perceptually in cognition.

We are now in a position to restate the question that started this article: if there is a perception/cognition border, then where does the application of concepts to perceptual representations occur? That is, how does perceptual categorization work (or is "perceptual categorization," strictly speaking, an empty category because perception does not subserve categorization)? One can perhaps better understand the debate by thinking of categorization as consisting of a "matching function" which matches stored concepts to incoming percepts. Conceptualists will hold that for at least some representations, the matching function is deployed within perception proper so that perception outputs

---

[2]   Those who think that the border between perception and cognition is identifiable via representational format also sometimes think that there are informational restrictions on perceptual representations (and tend to be sympathetic to modularity in general). It is in their right to do so, since there is no inconsistency between the two views. There may be further consilience between the two views because one might posit that the format difference can be used to explain some aspects of informational encapsulation.

[3]   At least none that are architecturally specified. Perhaps one might want to hold that certain contents are ipso facto restricted from activation in central cognition merely because of their content—e.g., perhaps contradictory thoughts of the form P and not-P could not arise merely because of the contradictory content (see Mandelbaum 2013, 2014, forthcoming).

conceptualized representations.[4] Non-conceptualists, on the other hand, posit that perception outputs representations that then have to be matched to concepts stored in central cognition, in which case central cognition performs the matching function.

Note that the question of where categorization occurs does not arise for theorists who flat out reject a perception/cognition border (Churchland 1998; Prinz 2006; Bar and Bubic 2013; Lupyan forthcoming; Shea forthcoming). If "perceptual" and "cognitive" processes are always interacting, then there should be a free-flowing exchange of information at all levels of processing. So, one might reason, the earlier we see categorization, the less likely it is that we have a perception/cognition border. This is because those who hold that there is a border also tend to hold that at early levels of processing there are informational restrictions on the processes that can transform perceptual representations. Thus, border theorists are naturally aligned with a feedforward model of cognition, one where most of cognition isn't available to early processing. So, prima facie, one might reason that if there were lightning-fast categorization, then that would have to be accomplished through the aid of cognition, in which case there would be either no border or very early cognitive penetration (see, e.g., Macpherson 2015).

But such reasoning would be too quick. Consider again those who defend the little discussed "shallow contents" aspect of modularity theory. They are conceptualists, for they hold that perception outputs conceptual representations corresponding to basic-level categories. Since they also maintain that modules are severely informationally restricted, only the most basic objects should be categorizable intramodularly. That is, even if one thinks that there is a perception/cognition border, and that perception subserves categorization, one would still only expect a relatively impoverished range of classificatory ability. To put it plainly, if there is a visual module (or set of modules) conceptualist border theorists would hold that perception has access to a relatively scant amount of information—just enough to allow perception to (at most) specify basic-level objects (e.g., cars) with their perceptual form.

Putting this all together, it seems that if we found that we could perform incredibly quick visual categorization of rapidly presented stimuli, we would have motivation to reject the non-conceptualist's reason for drawing the perception/cognition border, for either a) there is no such border or b) there is a border, but perception outputs conceptualized representations. A precisification of what sort of speed matters for carving up mental architecture (and why it matters) will be given in short order. In particular, what will be most probative is relative speed and processing order: non-conceptualists hold that first we perceive and then we categorize, in which case categorization should be slower than perception. Conceptualists will instead hold that since perception underwrites categorization, there should be a class of representations that have no delays at all between the output of perception and categorization.

In what follows, I argue that there is immediate categorization of visually presented stimuli and yet, nonetheless, we should still posit the existence of a perception/cognition border. Thus, I will hold that the non-conceptualists border theorists are wrong, for (at least part of) the output of perception is already conceptualized—that is,

---

[4]   Claiming that perception outputs a conceptualized representation does not amount to claiming that all of perceptual content is conceptual (see McDowell 1994; Brewer 1999; Mandik 2012). For all the categorization claim cares, it could turn out that perception also outputs representations that have non-conceptual content.

conceptualization via visual identification occurs intramodularly. I will then end by noting how the shallow contents outputted by perception affect debates about non-conceptual content and the admissible contents of perception.

2. Some Marks of the Conceptual

How could we tell whether a given representation was conceptual? Although the question is vexed there are some generally agreed upon properties that must count for a representation to be deemed conceptual. I take it that it is non-negotiable that whatever is conceptual is repeatable. Concept identity has to be construed in such a way that concepts persist over time at least in order to explain such mundane behavior as: categorizing a visually presented tiger as a tiger. The idea that concepts must be repeatable is a theme running through theorists as different as Evans (1982), Prinz (2002), Fodor (1998), and every cognitive psychologist I can think of. Roughly, the reasoning goes as such: say you like the generality constraint as a mark of the conceptual. There are multiple ways to cash out such a constraint, but all of them involve the representation at hand carrying its identity conditions over time; that is, they involve the representation being repeatable. For example, one reading of the generality constraint has a representation C counting as a concept if it embeds in contexts such as A IS C and B IS C and C IS A, ad infinitum.[5] But note, in each iteration it is the same (type identical) representation, C, in each context, so C must be repeatable.

Another way of cashing out the generality constraint is to say that concepts must be productive and systematic (Fodor 1998). In this case C is a concept only if it embeds in compositional contexts. But for a representation to do so is for it to ipso facto keep its identity conditions across time. According to this criterion, if C is a concept then one can think LARGE C, EDIBLE LARGE C, LOOMING OMNIPRESENT EDIBLE LARGE C, and so on. Again, it is the same (type identical) C repeating in each context.

Lastly, one might think that a characteristic property of concepts is that they are available for use by multiple mental systems. On this reading, we have prima facie reason to think that C is a concept if C can be utilized by mental processes as diverse as (e.g.,) reasoning, long-term planning, memory, articulatory, and motor processes. Concepts are hypothesized to be the building blocks of thought, and to be the vehicles through which we achieve rational thinking. The paradigmatic examples of rational thinking are stimulus-independent, decision-theoretic thought. For concepts to achieve this role, they will have to be poised for use by multiple mental systems.[6] Thus being used in disparate mental activities is highly suggestive of a given representation's being conceptual.[7] But again, this model of the functional role of concepts presupposes the repeatability of the

---

[5] Typographical note: small caps will be used to denote (stipulated structural descriptions of) concepts.

[6] Conveniently, the multiple systems constraint also serves to rule out intramodular representations as counting as concepts, which is the verdict given by many theories of concepts (because, e.g., intramodular representations fail the generality constraint test).

[7] 'Highly suggestive' does not, of course, equate to 'apodictic.' Perhaps one has reason to think that certain non-conceptual contents can also guide some limited action (Peacocke 1992). I don't intend to engage with that debate here. Instead we can use the heuristic that the more mental processes a representation is available for, and the more decision-theoretic the computations that utilize the representation are, the more canonically conceptual the representation appears to be. Overall, my goal in the paper is to allow for a broad use of 'concept' so as to not get caught up in specific subtleties of different theories of concepts and instead just highlight the main, core features of concepts.

representation under discussion. It is the (type identical) representation C that appears in both planning and verbal report. One decides that it is asparagus that one wants for dinner in part by reasoning with the concept ASPARAGUS and then reports that yes, they will have the asparagus by using the same ASPARAGUS concept. This constraint—the idea that conceptual representations are available to many different cognitive processes—is the idea underlying cognitive psychology's use of categorization. In (e.g.,) investigations of prototype structure, one looks at what factors affect categorization in diverse contexts (see, e.g., Hampton 1998; Prinz 2002). One looks to see how we categorize cats having first thought about pets, and then having thought about animals in general. But it is the same concept, CATS, that is purported to be under investigation in both contexts.

So, we now have a short list of probative properties a representation should have to be analyzed as a concept. The representation should be primed for use by diverse mental processes, with processes underlying compositionality and categorization given a particularly prominent post.[8] A necessary condition is that such a representation needs to be repeatable.

We can also go a bit further and say what it would take to show that perception has a conceptual element: it would be to show that the output of perception is used directly in thought. That would mean that the representation is not just used to cause the activation of a concept stored in central cognition, but is used by other mental processes such as decision theoretic and motoric mental processes, *inter alia*. That is, if perception outputs representations that are used to not just feed into a matching function, but to actually guide action by entering into other cognitive processes, then the outputs of perception are conceptual. So, now we have a test for the outputs of perception.

It is worth noting that some prominent theorists predict that perception would fail such a test. Take Burge, who introduces the idea of a "perceptual attributive" (2010). Burge analyzes the content of perception as *that f*, which contains a non-repeatable singular element (the *that* part) and a repeatable "general attributive" (the *f* part).[9] Burge is explicit in stating that 1) perceptual representations of the form *that f* do not involve deployment of the concept *f*, and 2) that perceptual representations do not enter into diverse cognitive processes (such as categorization) as described above. Thus, for Burge (and Block 2014), perception has no conceptual element, at least not in any of the senses of conceptual in the literature. As a consequence, Burge and Block must say that the matching function is deployed by cognition proper.

3. The Absurd Rapidity of Perceptual Categorization

As mentioned above, if categorization happens extremely quickly, then non-conceptualists are in trouble, for non-conceptualists hold that perception has to first output a wholly non-conceptualized representation and then match that representation to a concept in central cognition. So the quicker categorization happens, the worse it seems for

---

[8]   Categorizing strikes me (and others, see Prinz 2002) as sufficient for concepthood. More concretely, if representation C is used to categorize X as a C, then C is a concept. But I do not need the full sufficiency claim in order for my arguments to go through.

[9]   A typical Burge quote to that effect: "Perception involves a context-dependent element and a general repeatable element" (Burge, ibid., p. 232). It is unclear what the relation between these two elements and their iconicity is (e.g., whether Burge takes both elements to be part of a single icon, or separate iconic representations).

non-conceptualists, and the better things look for conceptualist border theorists or those who deny a border completely. And categorization happens at eye-popping speeds.

Take, for instance, the findings coming out of Mary Potter's laboratory. In Potter et al. (2014), subjects were rapidly presented with a series of visual images (in a rapid serial visual processing task, henceforth: RSVP). The subjects' goal was to detect and categorize a specified target from the sequence of the rapidly presented images. For example, a subject would be shown the category name by seeing the word(s) "flower" or "box of vegetables." Subjects would then be shown the sequence of images,[10] followed by what I will call the detection task, where subjects had to answer whether or not the target appeared in the images. If the target was missing, the subjects were told this after their response and the trial ended. But in trials where the subjects correctly detected a present target, subjects then received an extra question: they had to then pick the target out of a forced choice presentation of two pictures of the category, only one of which had been previously presented.

So, as a subject in a target-present trial, you might see the word "flowers" followed by a series of images, followed by a question "Yes/No?" If the subject correctly responded "Yes," they would then see two pictures of flowers, and they had to choose which of the pictures had been presented earlier in the RSVP task. As described so far, the task sounds blissfully easy until one considers the presentation times of the serial images. In one condition, the images in the RSVP task were presented for only 13 ms per image. To illustrate how absurdly short 13 ms is, consider that the shortest recorded blink times (i.e., the amount of time the eyelid envelopes the eye) are around 100 ms. The temporary blackout caused by blinking lasts between 40-200 ms (Volkman et al. 1980). And yet even at 13 ms presentation rate per picture (and even for 12 consecutive pictures!), subjects could correctly detect the target at levels significantly above chance. Moreover, the same held for the forced choice task on the target-present trials: subjects correctly detected which picture they were shown at above chance levels.[11]

Subjects' performances were, of course, affected by the quick durations. If the images were presented at 80 ms per picture, subjects' performance improved. Nonetheless, subjects were able to correctly complete the task at 13 ms presentations. Even more astonishing is that these presentations were forward and backward masked.[12] Masking normally ensures that the processing of the stimulus has to be stopped because the next stimulus is already appearing to be processed. This is especially true for the stimuli in the Potter experiments, for they were meaningful masks (as opposed to visual noise), which have particularly disruptive effects, often completely wiping out consciousness of the masked stimulus, as in the "attentional blink" (e.g., Loftus et al. 1988; Dehaene et al. 2001; Asplund et al. 2014). But regardless how one interprets masking the moral is the same: a perceptual process is able to subserve detection and categorization even under extreme conditions.

---

[10] Depending on the experiment subjects would see either 6 or 12 images. The difference between these variants will not matter for the current discussion.

[11] The reader may wonder why the experimenters chose 13 ms as the stimulus presentation time. The answer is instructive: the experimenters could not find a presentation time so fast that subjects could not succeed at the task, so they kept shortening presentation times. They stopped at 13 ms only because *that was the quickest refresh rate that their screen could display*. This raises the very real possibility that we can categorize objects more quickly than we can currently make technology to display such objects.

[12] Targets were never presented as the first or last image, thus targets were always masked on both sides.

That last sentence may appear to beg the question as it claimed that a *perceptual* process subserved *categorization*. Both claims need arguing. I will address the latter first.

The reason to think that honest-to-god categorization is occurring is that the subjects' task in responding isn't to merely parrot the stimulus back immediately. Instead, subjects must first change the format of the target. For instance, subjects are given the word "flowers" and then have to visually identify whether one of the pictures contained an image of flowers. Of course, they then also have to remember what image they saw, and then use this information to guide their response. Thus whatever representation is caused by the 13 ms presentation must be robust enough to subserve matching between word and pictorial formats, and be available for use by reasoning, linguistic, and motor systems. Thus, following the argument in section 2, the representation has to be conceptual because it appears primed for immediate use by a host of different mental processes. In which case it appears that the super short presentation times were long enough to allow the stimulus to be categorized *as such*.

But why should we see such categorization as being underwritten by perceptual processes? Why not just take this datum as evidence for top-down penetration? The main reason not to is based in neurological wiring: 13 ms is just too short a time to allow for top-down connections to take hold. Reentrant loops in the visual system are estimated to take at least 50 ms to make a round trip (Macknik and Martinez-Conde 2007; Dicarlo et al. 2012; Potter et al. 2014). And that is for reentrant loops *within vision*, nevermind central cognition.[13] If one held the view that vision is modular and consists of a series of submodules, 50 ms would be the shortest time it could take for feedback from higher vision. But the timing needed for *cognitive* penetration would be considerably longer. Thus, the 13 ms presentation rates are just too short a time for the categorization to be accomplished by cognition. Instead, the processing appears to be accomplished via a mental process that is super fast acting, whose processing starts automatically once it receives its proper input, and processes ballistically and without the need of awareness; in other words, a modular processor (see Mandelbaum 2015). Thus, Potter's work is evidence for the modularity of perception and for perception underwriting some forms of categorization.

And Potter's work is not alone. Keysers et al. (2001) showed that subjects could categorize stimuli presented at 14 ms (more on this particular experiment later). Similarly Grill-Spector and Kanwisher (2005) presented subjects images (such as a dog or a boat) at 17 ms and then masked them (with visual noise). Subjects were again significantly above chance even at these presentation rates. But a closer look at their data lends even more credence to a modular conceptualist view of vision. Subjects had one of three tasks. In the first, they had to detect whether or not they saw a target image (vs. seeing visual noise—a texture with no object). The second had them detect which of ten possible targets (face, bird, dog, fish, flower, house, car, boat, guitar, or trumpet) appeared in the images. The final task was similar to the first: subjects just had to detect whether a subordinate category did or did not appear in the images.[14] Note that the second task is,

---

[13]   "Within vision" meaning (e.g.,) to have a signal propagate from V1 to V4 back to V1. Also, these times are based around normal measurements of axonal propagation (which is the canonical measurements, as seen by, e.g., fMRI), but not based on neural oscillations (see, e.g., Neuling, et al. 2012) for which I know of no known estimates.

[14]   Subordinate categories are just ones that are more specific than the basic-level categories (such as those used in the second task). Thus subjects would be asked to detect whether the images had (e.g.,) a Jeep (and not just a car). Unlike the first task distractors were other subordinate (within category) images and not visual noise.

probabilistically speaking, considerably harder than the other two. In the first and last task, subjects have only two alternatives—either the images contained an object or not in the former, and either the images contained an exemplar of a subordinate category or not in the latter.

What the experimenters discovered was that the moment that subjects were able to detect the presence of a basic-level object (vs. e.g., visual noise) they could also categorize those objects. The tight connection between basic-level categorization and detection can be better seen by looking at failures: on the trials where subjects failed to correctly categorize the basic-level object, they were also at chance levels in detecting it. Perhaps even more surprisingly, response times were identical for the two tasks: categorizing a basic-level object took no longer than merely detecting the presence of that object. Thus we can safely say that performance was equivalent for basic-level detection and categorizations. However, this was not the case for the third task. Subjects needed considerably more time to identify subordinate-level exemplars and their performance was much more error filled than either of the earlier tasks. Furthermore, unlike in the categorization task, subjects could successfully detect stimuli without successfully identifying these subordinate-level categories, though subjects could never do the inverse. Thus, identifying subordinate-level categories appears to be the outlier here.

These data suggest a certain picture of perceptual processing: 17 ms presentation rates are too quick for reentrant loops, even just reentrant loops within vision, so the subjects' performance must be accomplished by a feedforward process, just as modularity suggests. Furthermore, we now have some idea how to understand the output of that process: the representation the process generates guides responses equally in the detection or categorization of basic-level categories. What would explain why detecting or categorizing a basic-level stimulus is equally easy, while merely detecting a subordinate exemplar is considerably more demanding?

This pattern of behavior should be expected if the output of the modular process is in fact a basic-level concept, like DOG, FLOWER, and HOUSE. If perception outputs basic-level concepts then detecting the presence of the basic-level content would be no harder than categorizing because perception itself would accomplish both tasks. Since one has no top down access to the workings of the modules, even though detection would (presumably) occur first, response times and accuracy rates would not be affected because the rest of cognition only has access to the outputs of perception and those outputs would arrive already categorized for basic-level concepts. Conversely, response times spike and performances decline on the subordinate task because that task involves identifying subordinate-level categories and these subordinate-level categories are *not* the proper outputs of perception. That is, they are outside the bounds of perceptual content. That is why one can detect and categorize the basic-level object without identifying it at a subordinate level: subordinate-level identification isn't a task for perception but is one for cognition proper.

The emerging picture is suggestive of the classic view of modularity, where the outputs of vision are "shallow contents"—concepts of basic-level categories. But discussion of shallow contents has been non-existent since the original presentation of modularity (Fodor 1983). Perhaps what has stalled discussion of it was the lack of recognition of what shallow contents entailed: if perception is outputting shallow contents, then it is ipso facto subserving categorization. In which case, the matching function must be accomplished intramodularly. Of course, other sorts of matching—such as categorizing to

subordinate levels—would be accomplished by central cognition, and such contents would be precluded from appearing in perception proper.[15]

4. Objections (and More Evidence)

The idea that perception is modular and outputs conceptual representations is not a particularly popular view. Thus, there are bound to be skeptical readers, whose suspicions I will now try to assuage.

Maybe the reader is skeptical of the Potter study because subjects are told what target to detect (e.g., they see "flowers") before the rapidly presented images appear. But in fact such a worry would be misplaced: subjects are just as accurate regardless of whether they get the target word before or after they see the images. That is, subjects who first see 12 images presented at 13 ms an image and are only then asked if they saw (e.g.,) flowers are still significantly above chance at detecting flowers. It appears that we process the rapidly presented images and then hold them at least long enough to perform certain computations on the information they contain well after the stimulus has disappeared.

The skeptical reader may also wonder about overall response times, and not just presentation times. Such worries are also misplaced. For one thing, overall response times involve times that are not due to perceptual processing and categorization, for overall response times include decision-theoretic and motoric elements—one has to decide which way to respond and then actually make the relevant motions. Thus, the cleanest way to examine perceptual processes is through stimulus presentation rates.

That said, some response times are shockingly fast, as long as one uses a dependent variable that allows for a quickly deployable response. For example, Kirchner and Thorpe (2006) showed subjects two images of natural scenes at once on a computer screen. One image contained animals in their natural environment, and the other contained natural scenes without animals (e.g., mountains, forests, buildings). The images were flashed for only 20 ms. Subjects were instructed to saccade to the image that contained the animal. Subjects could accurately complete the task in 120 ms. That's 120 ms to perceive, categorize, and respond; in other words, that 120 ms includes not just perceptual processes, but also decision theoretic and motoric processes (such as merely preparing the saccade, which takes 20-25 ms; Schiller and Kendall 2004). Again, such times are too fast for feedback loops from cognition to help decode the stimuli.[16]

---

[15]   In addition to how high (or low) level the contents of perception are, there is a related debate about the format of the vehicles that carry such content. Because of the ease of translation between visual and verbal formats, we have reason to declare that the output of perception is conceptualized. But does that entail it could not be iconic? No. But it is worth mentioning that some of Potter's stimuli weren't merely basic-level conceptual activations. For example, some of the stimuli depicted relations between basic-level concepts (e.g., children holding hands, a bear catching fish, boxes of vegetables). So it at least appears that the output of perception needn't be the activation of a simple concept, but can be complex, which opens up the possibility that it might in fact have propositional structure (see Quilty-Dunn, MS).

[16]   Similar evidence can be found using time-sensitive neural recording techniques. For instance, experiments utilizing ERP have found ultra-rapid categorization in under 50 ms (Mouchetant-Rostaing et al. 2000; VanRullen 2007). Single-cell recordings in patients with pharamacologically intractable epilepsy have showed category-specific neural firing (to categories of animals, chairs, faces, fruits, vegetables, and vehicles) at 100 ms (Liu, et al. 2009). The firings were specific enough that one could decode which category the subject was looking at merely by which cells were firing (at rates as quick as 100 ms). Here (as elsewhere) specificity was stimulus invariant—neuronal responses were the same within category regardless of changes to stimulus orientation and size (and regardless of whether different stimuli within the same category were used).

Note that the Kirchner and Thorpe can be used for dual purposes. First, successful completion of the task requires subjects to categorize the images as animals. The fact that subjects succeed at the task, therefore, means that they were categorizing the images as animals. Second, the 120 ms response times ensure that the response is the result of a feedforward perceptual process, just as modularity theorists would have it. Using our time estimates above that leaves approximately 100 ms for perception, categorization, and decision theoretic processes to work. If categorization is in fact being accomplished via central cognitive processes, then cognition has to solve the frame problem in that time, since cognition is totally unencapsulated. That is, cognition would have to identify objects as animals while sifting through and ignoring all that we know about animals (not to mention the particular animals seen). If, on the other hand, modular perceptual processes are responsible for categorizing then there is no such problem to solve because modular processes are by definition highly informationally restricted. Cognition stores everything we know about animals, but at least for the modularist, the only thing perception knows about animals is what they look like. Thus, the plight of the non-conceptualist is to carve out enough time for encapsulated cognitive processes to successfully deploy matching functions. Of course, this isn't a knockdown argument—this is cognitive science after all—but it is a challenge that is far from trivial to meet.

The skeptical reader might worry that it would be arbitrary from a design perspective for perception to output basic-level categories as opposed to more or less specific categories. Consider the sage advice "If called by a panther, don't anther" (Ogden Nash, as quoted by Fodor 1983, p. 70). One of the reasons theorists have been drawn to modularity theory is its evolutionary rationale: we want some processes that are informationally encapsulated because in perception we are willing to trade some amount of accuracy for increases in speed. Roughly, the intuition is that during panther identification what really matters is accomplishing such identification quickly. What matters less is being wrong as long as we err on the side of false positives. Searching through everything we know about panthers in order to make an identification would be extremely time consuming. The informational encapsulation of modular processes solves this problem by severely restricting what panther-related information vision would have access to in making panther identifications. That is, part of the proprietary information that vision has is what panthers look like.[17] Of course, there cannot be too much proprietary information or vision would be slowed. Seen in this light it makes sense that basic-level categories would be outputted by perception for that is the level of categorization that allows for easy action. Knowing something is an animal does not tell you whether to run away or not, but knowing it is a *lion* is self-evidently motivational.[18]

---

[17] Fodor (1983, p. 97) once opined that the outputs of modules are basic-level categories because these categories are the most abstract categories that are predicted by the distal stimuli (e.g., that are visually identifiable by their shape alone). It turns out, that basic-level categories are in fact recognized via general shape properties and can be detected via low-spatial frequencies (unlike subordinate-category identification which is dependent on high-spatial frequencies, Collin and Mcmullen 2005).

[18] The "easy action" claim is in part shorthand for all of the well-known advantages basic-level categories have for cognition—basic-level categories are taken to be maximally informative (and to maximally afford action; see Rosch 1978 for review). And importantly, categorizing with more specificity than just the basic-level—e.g., knowing that it is Jake's favorite lion that is in front of you—does not lend much more fitness enhancing helpful information than merely knowing it is a lion full stop. This is just a consequence of basic-level categories being maximally informative—one gains very little extra information in identifying subordinate categories below the basic-level.

It is somewhat odd that this consequence of modularity and its evolutionary benefits hasn't been noted previously. After all, the evolutionary benefits of modularity have often been touted (see, e.g., Pinker 1997; Tooby and Cosmides 2005; Barrett and Kurzban 2006). Moreover, we presuppose such categorization in other candidate modules, even though it is rarely discussed under those terms. Take the putative language module for instance. Part of the module's function is to recover phonological, syntactic, and semantic properties from (e.g.,) acoustic waveforms. The idea is that the module outputs something like the logical form of a sentence in addition to its phonological and syntactic properties. In contrast, implicatures and other pragmatic factors aren't meant to be specified in the parse—for that one needs to do some reasoning in central cognition (even if this reasoning is unconscious, see Mandelbaum forthcoming). But the parse itself is supposed to give us the meaning of the sentence. In which case the parse—the output of the language module—is subserving a categorizing function. At the very least it is producing the categorization of (e.g.,) word forms and parts of speech from transduced acoustic streams.

The current suggestion is that the same types of categorization that happen in linguistic perception happen in visual perception. Just as basic syntactic and semantic properties are produced by the language parser, so too are basic-level properties produced by the visual system. And just as certain pragmatic aspects of language are not specified in linguistic perception, so too are certain subordinate (and superordinate, for that matter) properties not outputted by visual perception.

5. How Conceptualism Saves Modularity

Taking this all in, it seems that the default position for a modularist should have always been to be a (partial) conceptualist, for without the conceptualism one loses some of the most seductive explanatory goods of modularity. But there's another reason to be a conceptualist: it allows one to respond to seemingly damning evidence against modularity theory. Take the work of Mary Peterson, who has been at the forefront of research on figure/ground perception. A figure/ground image consists of a white and a black silhouette (see figure 1). The silhouette seen as the figure will appear shaped by the black/white border, whereas the silhouette seen as the ground will appear shapeless. Peterson has shown that the shape of the silhouette that loses the competition and is parsed as ground will be suppressed in consciousness (Peterson and Skow 2008). Unsurprisingly, if the ground in fact resembled a recognizable shape, subjects are not conscious of that fact. Nevertheless, it appears that recognizable shapes that are ultimately seen as ground and thus suppressed are still *unconsciously* perceived and categorized before figure/ground assignment (Peterson et al. 2012; Cacciamani et al. 2014; see figure 2). Of particular interest for the current discussion: the test that showed this was a lexical decision task (where the subject has to say whether a string of letters forms a word or a non-word). This means that subjects must have unconsciously categorized the ground in a format that allows for speeded responses when identifying the word that expressed the same content as the ground.

At first glance, unconscious categorization on the ground-side looks like a strike against modular models of perception. Indeed, in earlier work, Peterson et al. (2012) take their findings to be evidence for this: "Peterson & Skow's results showed that before figure assignment, representations of objects that might be perceived on opposite sides of borders are activated, at least at the level at which shape structure is represented. Therefore, their results are inconsistent with feedforward models of figure assignment in particular and of
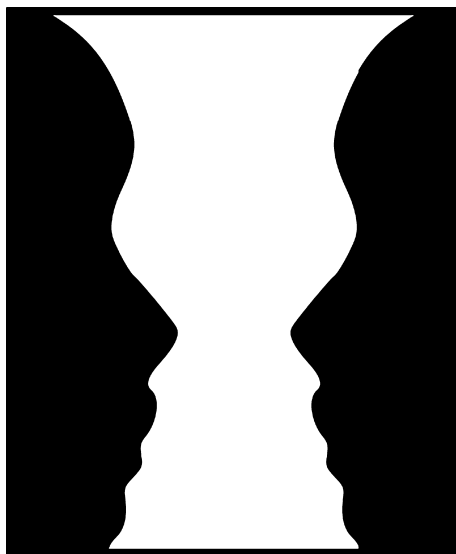
Figure 1. A typical figure/ground stimuli. One can see the faces and the white center as shapeless ground, or one can see the white vase as figure and the black periphery as shapeless ground.[19]

perception in general" (Peterson et al. 2012, p. 299–300). However, such an inference only makes sense against a backdrop where one assumes that feedforward processes do not have access to semantic information in order to output a categorized representation. Once that model is made available, semantic access on the ground-side no longer looks like an instance of top-down penetration. In part, this is because the categorization of the ground is parsed too quickly for cognitive penetration: the stimuli in Cacciamani et al. (2014) need only be presented for 50 ms (and the categorization appears to be complete after only 150 ms; see figure 2 for a sample stimulus). Peterson herself, no fan of modularity or foe of top-down effects, even explicitly recognizes as much in her most recent work on the topic. Here is Peterson on categorization happening too fast for penetration:

> That the effects of semantic access to ground regions are evident very early and diminish over time is consistent with the hypothesis that the semantic activation for ground regions is being accessed on an initial, fast, feedforward pass through the visual system, prior to the completion of figure assignment...This result is further evidence in support of a fast, nonselective evaluation of regions that could be perceived as objects, regardless of final figural status (Cacciamani et al. 2014, pp. 2543–2544).

But note this explanation—understanding how the meaning is accessed so quickly—is only available for the modularist who supposes that perception in fact subserves categorization. The modular non-conceptualist appears to have two problems. One is that the meanings of the silhouettes in the figure/ground images sometimes affect assignment of figure/ground (i.e., if only one of the silhouettes represents a common image, that silhouette is more likely to be seen as figure; Peterson and Gibson 1994). This appears to be

---

Figure 2. A sample figure/ground stimuli used in Cacciamani et al. (2014). Even when the white boots are not consciously perceived as such and are instead seen as ground, they still cause unconscious conceptual activation.

an instance of top-down penetration, but not if object meanings are available before figure/ground assignment as part of the intramodular proprietary database. Since non-conceptualists do not posit such intramodular categorization information, this explanation isn't available to them. Perhaps the modular non-conceptualist would then be inclined to ditch the modularism in favor of being a top-down theorist. But in that case they would run into a different problem: that of explaining how the effects happen before reentrant connections are available.[20]

6. Wrapping Up

So far I have argued that there is a border between perception and cognition, one that exists because vision is modular. The particular type of modularity at hand is one that interprets vision as outputting conceptualized representations but only for basic-level categories and not richer contents. Having established the evidence in favor of a modular, fast-feedforward sweep that produces conceptualized outputs, let's turn to some consequences of this model. If such a picture is right, it has a nativist bent. Although language perception might be specific to humans, visual perception absolutely isn't. In which case one might expect that we should find similar visual capacities (caused by similar visual modules) in closely related species. And it appears that we do. In the aforementioned Keysers study I said that subjects could categorize stimuli that were only visually presented for 14 ms. What I didn't specify was who the subjects were. It turns out that this performance holds for monkeys or humans. Rhesus macaques show similar acuity to humans under similarly lightning fast presentation times (14 ms; Keysers et al. 2001).[21] Likewise, new world monkeys have shown similar capacities at 25 ms presentation rates (Proctor and Brosnan 2013). So, similar findings are widespread across primates, which is what one would expect if categorization was subserved by an innate modular process.

---

[20]   Modular conceptualism would also serve to blunt many of the critiques that Firestone and Scholl (forthcoming) received, since many of their critics just assume recognition is inconsistent with modularity. For example, here's Raftopolous arguing that "late vision" isn't modular: "There are several ways cognition affects late vision, such as the application of concepts on some output of early vision so that hypotheses concerning the identities of distal objects be formed and tested in order for the objects to be categorized and identified" (forthcoming); and here's Levin, Baker, and Banaji "However, we also know from many decades of research that perception integrates sensory input with reliable world-knowledge" (such as how things look; Levin et al. forthcoming).

[21]   Of course, the dependent variables differed across the species: for monkeys, the specificity of neuronal firing was used as opposed to the behavioral measure used for humans.

This isn't to say that there is no learning element involved in categorization. In fact, the data calls out for a good deal of perceptual learning. The proprietary intramodular information appears to be basic-level, but many of the stimuli aren't ones that were around in the Pleistocene (e.g., house, car, boat, guitar, and trumpet were stimuli used in Grill-Spector & Kanwisher). Thus, it appears that though there is synchronic informational encapsulation, there must also be diachronic "penetration" in the form of perceptual learning. Which is exactly as it should be. Modularists want to deny that what you currently know or desire can affect what you see *right now*, but what no modularist should deny is the possibility and existence of perceptual learning. Informational encapsulation denies synchronic penetration, but diachronic penetration in the form of perceptual learning is a different story altogether.

Outside of questions of nativism, thinking of the problem of categorization has proven a fruitful tool investigating cognitive architecture. Its power for helping move forward certain debates in the philosophy of perception has been overlooked. I will conclude by touching on some debates that it does, and does not affect.

First, the debate most clearly affected pertains to whether there is a conceptualized aspect of perception. It appears there is. Of course, that's not to deny that there might be non-conceptual content in perception too. It's just to say that any view that wants perception to have only non-conceptual content appears untenable (cf. Burge 2010; Block 2014).

This brings us to the second debate clearly affected by these arguments: that there is a joint in nature between perception and cognition. Pace Block, Carey, and Burge, it appears that the defining criteria distinguishing perception and cognition is not to be predicated on the medium of representation (iconicity for perceptual content, propositional content for thought). Instead the break between perception and cognition is to be analyzed as perception involving modular processes and cognition not doing so.

Third, the arguments here have some purchase on debate about the admissible contents of perception. In particular, we appear to have arguments that militate against positing certain types of rich contents of perception. Debates about the content of perception are often confusing, so let me be clear on what I take the evidence here to point to. If perceptual processes are modular, then the question of the contents of perception boils down to the question of what contents perceptual modules output. And for that question we've uncovered some evidence: it appears that the conceptualized contents of perception are Roschian basic-level categories. Higher (than basic) level and subordinate contents aren't properly speaking the contents of perception.

It is also important to note what debates my argument leaves unscathed: debates about the contents of perceptual *experience*. Much of the evidence I've covered pertained to the unconscious outputs of perceptual modules. The debates of (e.g.,) Bayne, Brogaard, and Siegel about the content of visual experience is thus beyond the ken of the current discussion (Bayne 2009; Brogaard 2013; Siegel 2010).[22]

So, where have we ended up? It appears that perceptual faculties subserve near instantaneous categorization of stimuli that are presented at shockingly short speeds. The way to make sense of such data is to understand perception as being modular and producing

---

[22] That said, it might still be reasonable to suppose that one could use questions of categorization for illuminating the debates about the contents of experience. Insofar as one takes perceptual experience to be determined by perceptual processing, the discussion here will inform the possible contents of perceptual experience, greatly shrinking the allowable properties of high-level theorists like Bayne and Siegel, while enlarging the contents of low-level theorists' ontology.

conceptualized representations that are primed for immediate use by central cognitive processes. The content of such representations appears to be "shallow"—just basic-level categories. And this is all as it should be, at least if the goal of perception is to allow for the immediate detection of possible dangers in the ambient environment. After all, if perception is supposed to tell us where the panthers are, then perception ought to identify the panthers to us, and not just shuttle the problem of panther identification to cognition.[23]

References

Asplund, C., Fougnie, D., Zughni, S., Martin, J., & Marois, R. (2014). "The attentional blink reveals the probabilistic nature of discrete conscious perception." *Psychological Science* 25(3), 824–831.

Bar, M. and Bubic, A. (2013). "Top-down effects in visual perception." in *The Oxford Handbook of Cognitive Neuroscience* K. Ochsner and S. Kosslyn (eds.) pp. 60–73. New York: Oxford University Press.

Barrett, H., & Kurzban, R. (2006). "Modularity in cognition: framing the debate." *Psychological Review* 113(3), 628–647.

Bayne, T. (2009). "Perception and the reach of phenomenal content." *The Philosophical Quarterly* 59(236), 385–404.

Block, N. (2014). "Seeing-as in the light of vision science." *Philosophy and Phenomenological Research* 89(3), 560–572.

Brewer, B. (1999). *Perception and Reason*. Oxford: Oxford University Press.

Brogaard, B. (2013). "Do we perceive natural kind properties?" *Philosophical Studies* 162(1), 35–42.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

——. (2014). "Reply to Block: Adaptation and the upper border of perception." *Philosophy and Phenomenological Research* 89(3), 573–583.

Cacciamani, L., Mojica, A., Sanguinetti, J., & Peterson, M. (2014). "Semantic access occurs outside of awareness for the ground side of a figure." *Attention, Perception, & Psychophysics* 76(8), 2531–2547.

Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.

Churchland, P. M. (1988). "Perceptual plasticity and theoretical neutrality: a reply to Jerry Fodor." *Philosophy of Science* 55(2), 167–187.

Collin, C., & Mcmullen, P. (2005). "Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization." *Perception & Psychophysics* 67(2), 354–364.

Dehaene, S., Naccache, L., Cohen, L., LeBihan, D., Mangin, J., Poline, J., & Rivière, D. (2001). "Cerebral mechanisms of word masking and unconscious repetition priming." *Nature Neuroscience* 4(7), 752–758.

DiCarlo, J., Zoccolan, D., & Rust, N. (2012). "How does the brain solve visual object recognition?" *Neuron* 73(3), 415–434.

Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.

Firestone, C., & Scholl, B. J. (2014). ""Top-down" effects where none should be found: The El Greco fallacy in perception research." *Psychological Science* 25(1), 38–46.

——. (2015). "Can you experience 'top-down' effects on perception?: The case of race categories and perceived lightness." *Psychonomic Bulletin & Review* 22(3), 694–700.

——. (forthcoming). "There are no top-down effects." *Behavioral and Brain Sciences*.

Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT press.

——. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Grill-Spector, K., & Kanwisher, N. (2005). "Visual recognition as soon as you know it is there, you know what it is." *Psychological Science* 16(2), 152–160.

Hampton, J. (1998). "Similarity-based categorization and fuzziness of natural categories." *Cognition* 65(2–3), 137–165.

Jonsson, M. and Hampton, J. (2008). "On prototypes as defaults (Comment on Connolly, Fodor, Gleitman and Gleitman, 2007)." *Cognition* 106(2), 913–923.

Keysers, C., Xiao, D., Földiák, P., & Perrett, D. (2001). "The speed of sight." *Journal of Cognitive Neuroscience* 13(1), 90–101.

Kirchner, H., & Thorpe, S. (2006). "Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited." *Vision Research* 46(11), 1762–1776.

Levin, D., Baker, T., & Banaji, M. (Forthcoming). "Cognition can affect perception: Restating the evidence of a 'top-down' effect." *Behavioral and Brain Sciences*.

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). "Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex." *Neuron* 62(2), 281–290.

Loftus, G. R., Hanna, A. M., & Lester, L. (1988). "Conceptual masking: How one picture captures attention from another picture." *Cognitive Psychology* 20(2), 237–282.

Lupyan, G. (forthcoming). "Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems." *Review of Philosophy and Psychology*.

Macknik, S. L., & Martinez-Conde, S. (2007). "The role of feedback in visual masking and visual processing." *Advances in Cognitive Psychology* 3(12), 125–152.

Macpherson, F. (2015). "Cognitive penetration and predictive coding: a commentary on Lupyan." *Review of Philosophy and Psychology*, (doi:10.1007/s13164-015-0254-3).

Mandelbaum, E. (2013). "Against alief." *Philosophical Studies* 165(1), 197–211.

——. (2014). "Thinking is believing." *Inquiry* 57(1), 55–96.

——. (2015). "The automatic and the ballistic: modularity beyond perceptual processes." *Philosophical Psychology* 28(8), 1147–1156.

——. (forthcoming). "Attitude, inference, association: on the propositional structure of implicit bias." *Nous*.

Mandik, P. (2012). "Color-consciousness conceptualism." *Consciousness and Cognition* 21(2), 617–631.

McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.

Mouchetant-Rostaing, Y., Giard, M., Delpuech, C., Echallier, J., & Pernier, J. (2000). "Early signs of visual categorization for biological and non-biological stimuli in humans." *Neuroreport* 11(11), 2521–2525.

Neuling, T., Rach, S., Wagner, S., Wolters, C., & Herrmann, C. S. (2012). "Good vibrations: oscillatory phase shapes perception." *Neuroimage* 63(2), 771–778.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.

Peterson, M. A., & Gibson, B. S. (1994). "Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours." *Perception & Psychophysics* 56(5), 551–564.

——, & Skow, E. (2008). "Inhibitory competition between shape properties in figure-ground perception." *Journal of Experimental Psychology: Human Perception and Performance* 34(2), 251–267.

——, Cacciamani, L., Mojica, A., & Sanguinetti, J. (2012). "Meaning can be accessed for the ground side of a figure." *Gestalt Theory* 34(34), 297–314.

Pinker, S. (1997). *How the Mind Works*. New York: W. W. Norton and Co.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

——. (2006). "Is the mind really modular." in R. J. Stainton (ed.), *Contemporary Debates in Cognitive Science* (pp. 22–36). Malden, MA: Blackwell Publishing.

Potter, M., Wyble, B., Hagmann, C., & McCourt, E. (2014). "Detecting meaning in RSVP at 13 ms per picture." *Attention, Perception, & Psychophysics* 76(2), 270–279.

Proctor, D., & Brosnan, S. F. (2013). "Visual processing speed in capuchin monkeys (Cebus apella) and rhesus macaques (Macaca mulatta)." *International Journal of Comparative Psychology* 26(2), 166–175.

Quilty-Dunn, J. (MS). "The syntax and semantics of perceptual representation."

Raftopoulos, A. (2015). "The cognitive impenetrability of perception and theory-ladenness." *Journal for General Philosophy of Science* 46(1), 87–103.

——. (Forthcoming). "Studies on cognitively-driven attention suggest that late vision is cognitively penetrated, whereas early vision is not." *Behavioral and Brain Sciences*.

Rosch, E. (1978). "Principles of categorization." in B. E. Rosch & B. B. Lloyd (eds.), *Cognition and Categorization* (pp. 28–49). Hillsdale, NJ: Erlbaum.

Schiller, P. H., & Kendall, J. (2004). "Temporal factors in target selection with saccadic eye movements." *Experimental Brain Research* 154(2), 154–159.

Shea, N. (forthcoming), "Distinguishing top-down from bottom-up effects." in S. Biggs, M. Matthen and D. Stokes (eds.), *Perception and Its Modalities*. Oxford: Oxford University Press.

Siegel, S. 2010. *The Contents of Visual Experience*, New York: Oxford University Press.

Smith, J. D., & Minda, J. P. (2002). "Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(4), 800–811.

Tooby, J., & Cosmides, L. (2005). "Conceptual foundations of evolutionary psychology" in David Buss (ed.), *The Handbook of Evolutionary Psychology* (p5–p6). Hoboken, NJ: Wiley and Sons.

VanRullen, R. (2007). "The power of the feed-forward sweep." *Advances in Cognitive Psychology* 3(1–2), 167–176.

Volkmann, F. C., Riggs, L. A., & Moore, R. K. (1980). "Eyeblinks and visual suppression." *Science* 207(4433), 900–902.