On An Ambiguity in Mandatory Perceptual Processing; Or, Finding Out What's So Special about Unimodal Perceptual Processes

Abstract:

Perceptual processes, in particular modular processes, have been understood as being *mandatory*. But exactly what mandatoriness amounts to is left to intuition. This paper identifies a crucial ambiguity in the notion of mandatoriness. Discussions of mandatory processes have run together notions of *automaticity* and *ballisticity*. Teasing apart these notions helps to taxonomize different potential natural kinds of perceptual processes; in particular, we can individuate unimodal, multimodal, and purely cognitive processes according to how they process stimuli. These distinctions can help us to better evaluate the evidentiary status of modularity theory.

Keywords: Modularity, Cognitive Penetration, Perceptual Processing, Cognitive Architecture, Psychological Kinds

***

1) The Putative Autonomy of Perceptual (and Perhaps Other) Mental Processes

One of the most fraught debates in philosophy of mind and cognitive science concerns whether perceptual processes are inherently *modular*. In philosophy the debate connects to foundational issues in the philosophy of science and epistemology, as well as to debates about perception. For instance, one may be interested in how much the theories one holds

affects one's perceptions (Churchland 1988, Fodor 1988) or in how top-down penetration would affect justification by perception (Siegel 2012). Investigations of the theory-neutrality of perception generally take a stand on the question of how modular perceptual processes are. If you claim, as Siegel (2010) does, that knowing that you are seeing an elm and not just a tree can change the phenomenology of your percept, you are—prima facie, at least—committed to a claim about how cognition can interact with our perceptual faculties.[1] Indeed, modularity is at heart a claim about how little interaction there is between our conceptual stores and our perceptual apparati.

Because of their connection to foundational issues in the philosophy of mind, science, and epistemology, the modularity wars unsurprisingly have been important to philosophers of perception. But they are also important to cognitive scientists concerned with the science of perception, in part because of how advanced the study of perception has become. If one wants to see the state of the art of cognitive science, one can hardly do better than focusing on perceptual processes. The study of perception has been at the cutting edge of cognitive science since the field's inception, offering basic models of perceptual processes and mechanisms with far greater frequency and accuracy than does the study of cognition. Modularity theory is often trotted out as the theoretical framework unifying the perceptual faculties as a natural kind of sorts (Fodor 1983; Scholl and Leslie 1999; Scholl 2005; Carey 2009).

[1] Siegel's claim, like Macpherson's (2012) pertains to how perceptual *experience*, that is phenomenology, can change based on one's cognitive and conative states. These discussions differ from traditional debates about modularity which are focused on the modularity of perception and not experience. It is the modularity of perception proper that will be the focus of the arguments in this essay.

Modules are mental processes that are fast, domain specific, informationally encapsulated, and mandatory (Fodor 1998, Carruthers 2006, Mandelbaum 2012). Modules are also thought to be innate (though Karmiloff-Smith [1992] challenges this criterion), to have regularized and predictable development, and to exhibit characteristic breakdowns.[2] In its original presentation (Fodor 1983), modularity theory hypothesized that the perceptual processing underwriting the five basic senses, in addition to language processing, were modular. Since then the groupings of what count as modules have ballooned to include non-peripheral elements of the mind, such as a number module, a letter-recognition module, a visual-postural module, and a cheater detection module.

Some Evolutionary Psychologists (e.g., Tooby and Cosmides 1992; Pinker 1997; Barrett and Kurzban 2006; Carruthers 2006) have tried to push the notion even further, arguing that every mental process is in fact modular, thus destroying the alluring hypothesis that modules constitute a natural kind that is perception-specific. In the hands of these theorists, unimodal perceptual processing, linguistic parsing, and cheater detection constitute their own single natural kind, in that they are all equally modular. This idea should be (and has been, see Fodor 2000) upsetting for both philosophers and cognitive scientists. For one thing, many of the considerations that can be used to explain why perception should be modular don't carry over to non-perceptual mental processes. From an evolutionary standpoint, perception should be relatively autonomous from our goals and desires since, sadly, the world often

[2] Initially, modules were also posited to be neurally localized and to have 'shallow outputs,' but these properties have since dropped out of most presentations of modularity. To my eyes, the neural localization criterion has been dropped both because of the increasing amount of anti-localization evidence (see, e.g., Anderson 2010) and also because of all of the confusion it engendered between neurological modularity and psychological modularity, the latter being the topic at issue here. I suspect that discussions of shallow outputs have ceased in part because people had trouble making sense of exactly what constituted a shallow output and in part because, insofar as one does have an idea of what it is, it's hard to see how what would constitute a shallow output of, say, a audition or gustation module.

isn't as we hope and expect it to be. If I open the door to my house and expect to see a happy tail-wagging dog but instead stumble upon a ravenous wolf, it's best that I indeed see the aggressive animal as a wolf and not the family pet—that way I might be able to keep my arm intact.

It's also important that the ambient environment is not just processed correctly, but also quickly: it'll do me little good to see the animal as a wolf if it takes five minutes to do so. By then I'll be supper. In order to speed up processing, it's a reasonable strategy to ignore most of the beliefs we have about the world, instead mainly using bottom-up processes to let the environment determine the percept. This strategy saves time: instead of searching through one's full belief store and figuring out which beliefs are germane and which aren't, modules come preprogrammed with a domain-specific proprietary database. All other beliefs are ignored.

The result is a highly efficient perceptual apparatus. However, the very features that make for efficient perception also impede efficient cognitive processing. Using bottom-up data-driven strategies for perception is a pretty good idea, mostly because the evidence we need really is often right in front of our faces. This is rarely the case for cognition. The evidence for how the weather is at this moment really is outside; the evidence for what career paths we should follow…not so much. The truths that cognition is built to uncover are often best attacked by utilizing *all* available information, as opposed to some predetermined subset.

So, by extending the notions of modules beyond mere perceptual ones, we end up losing a sense of what makes the perceptual faculties a natural kind, and we lose some of the

evolutionary considerations that could help to explain how we ended up with the cognitive architecture we have. Moreover, many cognitive scientists share the intuition that there is something a bit odd about lumping unimodal perceptual processes, multimodal processes, and cognitive processes together. Even the sternest proponents of massive modularity have suggested that something doesn't quite sit right with grouping together unimodal, multimodal, and cognitive modules (see, e.g., Barrett and Kurzban 2006). If we want a firm theoretical grounding for what makes perceptual faculties an interesting class of mental process, we must differentiate between types of mental processes. By doing so, we can clarify the debates over whether perception really is modular and in turn illuminate the question of whether there is a single natural kind unifying certain perceptual, linguistic, and cognitive processes.

2) The Ambiguities of Mandatory Processing

Roughly, one can understand mandatoriness in two ways, which I label *automatic* and *ballistic*.[3] A mandatory process is *automatic* in this sense just in case if a module encounters a stimulus that is in its proper domain, the module *must* process the stimulus regardless of what else is occurring in one's mind (e.g., regardless of the allocation of other mental processes). In essence, automaticity says that all that it takes to start a perceptual process is the mere presence of the domain-specific input.

In contrast, a mandatory process is *ballistic*, just in case when the processor starts it cannot be stopped by any endogenous means. The proper input is not necessarily processed every time

[3] There is also one further sort of ambiguity here that won't be the focus of the discussion; see footnote 7 for further details.

the input reaches the module, but once the processing starts, one cannot stunt it at will, either through top-down effort or via other roughly psychological means.[4] Identifying something as a purely ballistic process does not explain how the processing is initiated. It simply describes how the processing unfolds, viz. in an uninterrupted fashion.

The contrast can be summed up like this: automatic processing can't help but start whenever the proper stimulus is encountered, whereas in ballistic processing, the mere contiguity of the stimulus with the processor isn't necessarily enough to set off the module. Likewise, an automatic but not ballistic process would be one where processing begins anytime a proper stimulus is encountered but may be stopped short by any number factors after the onset of the processing.

To illustrate, suppose that attention can affect unconscious modular processes. If the modular process is automatic, then no reallocations of attention could stop its processing from starting once it has encountered its domain-specific stimuli, but for all automaticity cares, reallocations of attention could be able to stop the processing from completing its route after the module has been triggered. However, when mandatory processes are ballistic, reallocations of attention (among presumably other factors) could shortcut whether the module would start its processing, but once the processing has been started, the processing would have to continue to completion.

[4] I'll leave out the 'through psychological means' modifier from here on out, assuming that it's taken as read. Of course, non-psychological factors could shortcut either of these notions; a mental process might be stunted because of number of 'lower-level' and non-psychological factors such as aneurysms, unfortunately aimed projectiles, or untimely death. This is just to say that here, as elsewhere in the special sciences, *ceteris paribus* clauses abound.

In principle, an automatic process could also be ballistic. I'll argue that an interesting set of perceptual processes is both. But I'll also argue that, upon examination, these two types of mandatory processes dissociate, not just conceptually, but also empirically, in taxonomically interesting ways.

Let's use vision as the paradigm example of a unimodal perceptual process. Consider the ease with which we parse a normal visual scene. We open our eyes and poof, the world just appears before us. Visual processors thus seem to be automatic: one doesn't expend any cognitive energy to see the ambient environment, one just opens one's eyes and puts the lights on. Vision is automatic in the sense that whenever the visual apparatus interacts with its proprietary stimuli (i.e., the transduction of impingements of light arrays on the retina), the apparatus must begin its processing.

Vision appears to be ballistic too: when encountering an illusion, such as an illusion like the young girl/old woman illusion, we can flip the illusion back and forth, seeing the image sometimes as an old woman, sometimes as a young girl. But we know that the illusion is neither: it is just an ambiguous representation. Yet the visual system cannot recover this information, it can only specify the percept to be stable as either the young woman or the old woman.[5] Likewise, when watching the hollow-mask illusion, we can't stop the stream of processing by introducing, say, the information that the mask does not have two convex sides. Such knowledge doesn't affect visual processing, because once the visual processing starts, it cannot stop. Thus, ballisticity helps to ensure a certain sort of informational

[5] One might object that in fact we can choose to change our fixation point, but that sort of early selective attention is pre-modular. One has the choice of where to move one's eyes, but once the fixation point is set the percept arises automatically.

encapsulation: knowledge that could bear on the resolution of a percept cannot affect perceptual processing, in part because the modular process cannot stop once it has started. If the module were not ballistic, however, then there could be a halting point where germane information could be introduced.

In multimodal perceptual processes, we do seem to find such halting points. Language perception develops in a multimodal fashion: children generally learn language not just by hearing linguistic input but also by seeing the shape of the speaker's lips. The visual cues help to decode the sentence. The fact that language comprehension is multimodal can be seen in the McGurk effect, in which an illusory percept is created by one hearing 'ba-ba' while seeing lips that are producing the sound 'ga-ga,' thus producing a multimodally determined percept that sounds like 'da-da' (McGurk and MacDonald 1976).

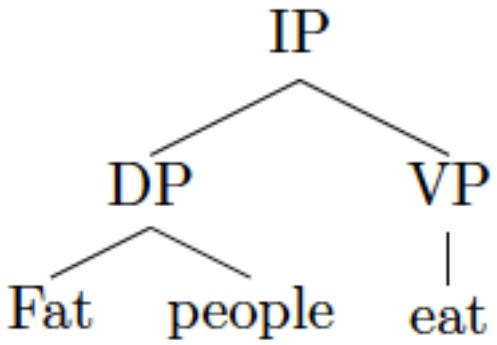
Language perception is, no doubt, automatic: one doesn't have to try and hear one's native language as language.[6] In fact, in normal conditions one cannot help but hear a sentence in one's native language as *language*; one cannot hear it merely as an acoustic stream of sounds. The semantic properties of a sentence just pop out in normal conditions.

But with eccentric stimuli the situation looks a bit different. When given a garden-path sentence, perceivers still cannot help but hear the sentence as a *sentence* and not, say, as a purely acoustic object, yet the syntax and semantics of the sentence are often not specified—the parsing of the sentence 'crashes.' Traditional models of what happens during such crashes work as follows: a context analyzer is continually monitoring the structural

[6] Some dichotic listening experiments show that you needn't even be conscious of the stream in order to process it as language and encode its semantics.

descriptions that are created during language perception (Crain and Steedman 1985, Ni et al. 1996). Since the module is mostly informationally encapsulated, the contextual analyzer cannot introduce all the beliefs that could be relevant for parsing the sentence as the sentence is being processed. Instead, all the contextual analyzer does is tell the module when it should discontinue its current processing stream.

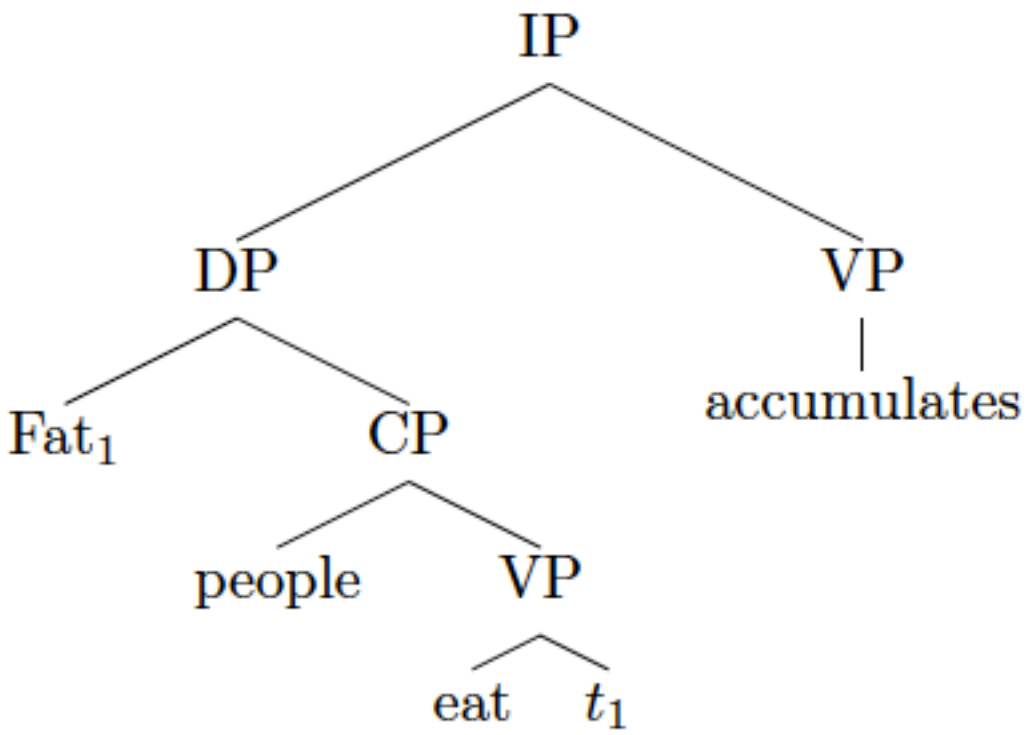
For instance, when one encounters 'fat people eat accumulates' one first builds the structure like in figure 1.



(Figure 1)

This structure then crashes once the system encounters 'accumulates.' What happens then is that the contextual analyzer tells the module to stop its processing and opens up the online information to central cognition so that all of one's beliefs can be used to form the sentence. This is a much slower way of creating a structural description of a sentence and is rarely used; it is only brought to bear in situations where the modular processing has been stopped. Say we were just having a conversation about how environmental factors affect obesity, and you utter, 'Fat people eat accumulates.' Once the contextual analyzer starts the halting (after

'eat') the contextual information can be used to create the correct structure of the sentence (figure 2), thereby specifying its correct meaning.



(Figure 2)

Thus, we can begin to see a real difference between unimodal and multimodal perceptual processes—unimodal perceptual processes are both ballistic and automatic, whereas multimodal processes may just be automatic and not ballistic.

Are there perceptual processes that are ballistic but not automatic? The evidence here is even thinner, but one can speculate. It seems plausible that the 'cognitive modules' that evolutionary psychologists discuss are, in fact, ballistic but not automatic. Take the classic example of a cognitive module: cheater detection. When given a 'social conditional'—that is,

a conditional in which there is a benefit received without a requirement met, such as, 'if one is going out at night, then one must tie a small piece of red volcanic rock around one's ankle' (Cosmides et al. 2010)—people often can correctly identify the situations that would falsify the conditional. However, people have serious trouble parsing the truth conditions of conditionals that aren't this type of social conditional—e.g., 'if one is taking out the garbage, then one must tie a small piece of red volcanic rock around one's ankle' (ibid.). For example, 80% of participants correctly identified the truth conditions of the former sentence, whereas only 44% correctly identified the conditions of the latter one (which has no benefit involved, so isn't considered a 'social conditional'; ibid.).

Cosmides et al. want to claim that the processing of social conditionals is modular. This may indeed be true in some sense, but is it both automatic and ballistic? I suggest that the processor may be ballistic but not automatic. There has been no evidence uncovered that merely encountering these sentences makes their truth values (i.e., the conditions under which the conditional is false) pop out. One might immediately parse the sentences as a bit of language, but that in and of itself doesn't seem to produce apprehension of the falsifying conditions. However, once one attempts to assess the sentence, then the falsifying conditions do seem to pop out. I suspect cheater detection isn't automatic because the mere encountering of the output of the language parser (the conditional) does not in and of itself make the falsifying conditions apparent; to accomplish that one has to also trigger certain concepts having to do with benefits received and unmet requirements (Cosmides et al 2010).

The current suggestion is that attention and motivation help to trigger certain concepts which can start cheater detection but, once the processing is started, we can just sit back and

wait for the answer to pop out: no further endogenous activity is necessary (and furthermore, subsequent endogenous activity won't be able to stop the process even if one wanted to). Endogenous activity may be necessary to start the module, but once started, the module runs on its own. If this is right, then we can start to see a real trichotomy of perceptual processes: unimodal perceptual processes are automatic and ballistic, multimodal perceptual processes are automatic but not ballistic, and central modular processes may just be ballistic without being automatic.

It is too early to make any definitive claim that such a taxonomy carves the mind at (some of) its joints. But it does appear to bring into focus a serious problem with the view that unimodal perceptual processes, linguistic parsing, and cheater detection form a single natural kind: some of these processes are ballistic, others automatic, and still others are both. Perhaps as opposed to searching for a single set of mental processes that form a psychological kind, we should instead be searching for which processes are automatic, ballistic, or both.

Even if these distinctions don't prove to create true psychological kinds on their own, they may aid our search in understanding what makes unimodal and multimodal processes interestingly different—at the very least they differ in terms of mandatoriness.[7] One can

[7] Perhaps it's worth adding that there is at least one further disambiguation of the notion of mandatoriness, at least as it's used in the literature. Most confusingly, Fodor not only ran both the automaticity and ballisiticty reading together in the original presentations of modularity, but he also introduced a third, wholly separate notion and claimed that it was the most conservative one, even though it appears to be the most tendentious. When discussing mandatoriness, Fodor writes, 'Perhaps the most conservative claim is this: input analysis [i.e., modular processing] is mandatory in that it provides the *only* route by which transducer outputs can gain access to central processes; if transduced information is to affect thought at all, it must do so via the computations that input systems perform' (Fodor 1983, 54). Fodor's basic idea is that no raw sensations (i.e., transduced information) can get into central cognition without first moving through a modular process. I suspect that this supposedly conservative test of mandatoriness has been the source of much confusion in debates over

hope that separating the automatic from the ballistic can inform future debates over the modularity of perceptual processing.

Anderson, M. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 33: 245-313.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience* 15 (4): 600-9.

Barrett, H., and Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychological Review*, 113 (3): 628-47.

Carey, S. (2009). *The Origins of Concepts*. New York: Oxford University Press.

Carruthers, P. (2006). *The Architecture of Mind*. New York: Oxford University Press.

Churchland, P. (1988). Perceptual plasticity and theoretical neutrality: a reply to Jerry Fodor. *Philosophy of Science*, 55: 167-87.

modularity. For example, Bar (2003), Kverga et al. (2007), and Fenske et al. (2006) all make claims that modularity is false because of certain sorts of top-down facilitation of object recognition. However, these claims are made on the basis of data using low-spatial-frequency images (that is, images that are very hard to see as images of anything). The modularist might object to these theorists (as in fact many do; see, e.g., Fodor 1988) on the basis that degraded stimuli encourage guessing, and guessing is cognitive not perceptual; degraded stimuli would thus not count as proper inputs in any sense. Proper inputs would be ones that aren't wholly degraded. However, for this type of explanation to work, there has to be a bound on what can set off the module; that is, it has to be allowed that certain sorts of raw, transduced information, such as the retinal array of low-spatial-frequency images, can bypass a module completely. It seems that the 'conservative' sense is thus much more tendentious than either the automatic or the ballistic sense. Moreover, I am satisfied the work of Bar and others shows that the putatively 'conservative' sense is quite probably false, and therefore have ignored this third sense in this essay.

Cosmides, L., Barrett, H., and Tooby, J. (2010). Adaptive specialization, social exchange and the evolution of human intelligence. *Proceedings of the National Academy of Sciences*, 107 (2): 9007-12.

Crain, S. and Steedman, M. (1985). "On not being led up the garden path: the use of context by the psychological parser. In D. Dowly, L. Kartunen, and A. Zwicky (Eds.) *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives.* Cambridge: Cambridge University Press.

Fenske, M., Aminoff, E., Gronau, N., and Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Progress in Brain Research*, 155, 3-21.

Fodor, J. (1983). *The Modularity of Mind.* Cambridge, Mass: MIT Press.

Fodor, J. (1988). A reply to Churchland's 'Perceptual plasticity and theoretical neutrality'. *Philosophy of Science* 55 (2): 188-98.

Fodor, J. (1998). *In Critical Condition.* Cambridge, Mass: MIT Press.

Fodor, J. (2000). *The Mind Doesn't Work That Way*. Cambridge, Mass: MIT Press.

Karmiloff-Smith, A. (1995). *Beyond modularity: A Developmental Perspective on Cognitive Science*. Cambridge, Mass.: MIT Press.

Kverga, K., Ghuman, A., and Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, 65 145-68.

Macpherson, F. (2012) Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84 (1): 24-62.

Mandelbaum, E. (2012). "Numerical Architecture." *Topics in Cognitive Science*.

McGurk, H., and Macdonald, J. (1976). Hearing lips and seeing voices," *Nature*, *264* (5588): 746-8.

Ni, W., Crain, S., and Shankweiler, D. (1996). Sidestepping garden paths: assessing the contributions of syntax, semantics, and plausibility in resolving ambiguities. *Language and Cognitive Processes*, 11 (3): 283-334.

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

Scholl, B. (2005). Innateness and (Bayesian) visual perception. In *The Innate Mind: Structure and Contents*, P. Carruthers, S. Laurence, and S. Stich, (eds.), New York: Oxford University Press.

Scholl, B., and Leslie, A. (1999). Modularity, development, and 'theory of mind'. *Mind and Language. 13* (1): 131-153.

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

Siegel, S. (2012) Cognitive penetrability and perceptual justification. *Nous*

Tooby, J. and Cosmides, L. (1992). Cognitive adaptations for social exchange. In J Barkow, L. Cosmides and J. Tooby (Eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.