Language and the Border Between Perception and Cognition

Steven Gross
Johns Hopkins University

I. Introduction

Ned Block's *The Border Between Seeing and Thinking* synthesizes a vast array of experimental results to argue that there is a "joint" – a fundamental explanatory difference – between perception and cognition. Perceptual states, on his view, are constitutively iconic, non-conceptual, and non-propositional; cognitive states do not possess these features constitutively, to the extent they possess them at all. It's argued that these constitutive features mesh with perception's function to quickly but reliably represent aspects of the here and now and help explain why clear cases of perception exhibit various empirical marks, such as adaptation, rivalry, pop-out, and illusory contours.[1] Notably, Block's characterization does not preclude (limited) top-down effects of cognition on perception.

But where does language fall in relation to this joint? Might it challenge the distinction? Language—more specifically, for our purposes, utterance comprehension—is complex, involving multiple kinds of processing and representations. Where some aspects or stages fall is fairly clear: the extraction of phonemes, on the perceptual side; judgments of what was asserted, on the cognitive. But things are less clear with others. Representations of syntax in comprehension are, like typical cognitive states, amodal and not iconic. But, like typical perceptual processes, parsing is stimulus-driven, fast, (nearly) automatic, and to some extent modular (Ferreira & Nye 2017). Fodor (1983, p. 44) thus grouped parsing (and some other aspects of linguistic comprehension) with "traditional" cases of perception, arguing that these "input systems" form an important natural kind. Some go further: They claim that utterance comprehension more generally is perceptual, even the assignment of meaning or content. Brogaard (2017; 2020) supports this by adverting to the kind of empirical marks—susceptibility to pop-out, adaptation, etc.—that Block argues are diagnostic of clear cases of perception. (Gross (forthcoming) critically discusses empirical arguments for "perceiving meaning.")

Block sees "systems that appear to constitute a mental kind that is both paradigmatically perceptual and paradigmatically cognitive" (16) as posing the biggest threat to the joint as he conceives it. He worries in particular about core cognition—specialized systems implicated in the mental representation of causation, approximate numerosity, and agency. Block's strategy in response (335-345) is to argue that, though core cognition

---

[1] For more on perception's function(s), see Phillips and Firestone's contribution to this symposium.

involves both perceptual and cognitive representations, there is no call to posit representations intermediate between them (at least regarding causation and approximate numerosity—agency is set aside). His discussion focuses on Nick Shea's (2014) development of the charge. Interestingly, Shea also cites parsing as problematically blending perceptual and cognitive features.

If Block deems language less of a threat to the joint than core cognition, it may be because he thinks a parallel response—that there is no call to posit "a third type of mental representation intermediate between percepts and concepts" (335)—is more easily made in the case of language. Block (53-57) adverts to adaptation results to argue that phonemes are perceived as such, but words—construed as individuated in part by their syntactic and semantic features—are not. This would establish a clear divide between the perceptual and the cognitive in linguistic comprehension. A takeaway would be that features like speed (at least a coarse-grained introspective sense of speed), automaticity, stimulus-dependence, and modularity do not differentiate perception and cognition. There may be another joint that separates systems that share such features from others. But, on his view, it wouldn't be the joint between perception and cognition.

In what follows, I raise some questions regarding the joint and language. Even if there is an important divide between phonemic and lexical attribution, certain visual word form representations may pose a problem for Block's conception of the joint: they seem both susceptible to adaptation and non-iconic. I'll offer several possible strategies in response; one would illuminate and tighten the connection between iconicity and adaptation. Consideration of language may also lead to other refinements. Much of Block's discussion suggests at least a rough alignment of the transition from the perceptual to the cognitive with a transition from the perceptual to the conceptual, as well as with presence in working memory and global broadcasting.[2] Language comprehension may provide reason to cleave some of these alignments. Relatedly, language comprehension might challenge a thesis, endorsed by Burge, that, when functioning properly, there must be a non-ampliative transition from perception to conception. It is unclear whether Block would endorse this thesis. If not, this would be another interesting difference, beyond those Block notes himself, between his and Burge's otherwise fairly similar views.[3] In any event, I hope to draw Block out further on the case of language.

II. Adaptation and Language

Block takes up core cognition because of the challenge it poses to his view of the joint. Block's reason for discussing perception and cognition in language is different: it's presented as a case study that exhibits the utility of adaptation in determining where representations fall.

---

[2] Regarding the cognitive and the conceptual, note for example the quotes from Block in paragraphs 3 and 4 above.

[3] The differences Block discusses include his acceptance, and Burge's rejection, of perceptions lacking singular content (103-119) and of adaptation-evidence for visual perception of higher-level properties (68-80).

Susceptibility to adaptation is among the most powerful indicators of a perceptual phenomenon. Adaptation typically involves a "repulsive" effect on perception after repeated or prolonged exposure to a stimulus. Subsequent perception is biased away from its previous response, as when prolonged exposure to red leads to a green after-image. That phonemes adapt has been well-established—e.g., Eimas & Corbit (1973). For example, subjects repeatedly exposed to /da/ are more likely to hear a subsequent /da/ as /ta/—and vice versa. (/d/ and /t/ are opposed along a dimension of phonological space.)

Block focuses on work by Arthur Samuel (1997) that well demonstrates how adaptation can settle difficult cases. Samuel replicated phoneme adaptation in the course of extending the phenomenon to phonemic restoration. In phonemic restoration, subjects report hearing a phoneme that has been replaced by white noise. For example, when presented with /legi#lature/, they report hearing /legislature/; and they cannot reliably distinguish between stimuli with and without the phoneme. A natural question—not readily answered from the armchair—is whether these reports reflect a perceptual effect or whether phonemes are restored rather as part of a perception-based judgment. Samuel answered this question by showing that restored phonemes, like phonemes in fact in the stimulus, generate adaptation effects.[4] Repeated exposure to stimuli like /legi#lature/ can bias subsequent perception away from the restored phoneme. This gives strong reason to treat the restored phonemes as themselves perceptual.

Samuel also investigated whether attributions of lexicality or other non-phonological features of words can contribute to this adaptation effect. For example, does repeated exposure to 'kiss' generate a greater phonemic adaptation effect than repeated exposure to the pseudoword 'giss'? Samuel found no difference, suggesting that the adaptation that is generated in both cases arises solely from phonemic features. (Samuel's larger interest is to determine the extent of top-down lexical effects on phonological processes. He argues for effects on phoneme restoration but not on subsequent adaptation.) Block describes these results as showing "a strong adaptation effect for phonemes but none for words, suggesting that words are not perceptually represented (though of course they are cognitively represented)" (54). But this formulation (cf. Samuel 2001, p. 349) could mislead: Samuel's results show that lexicality and other non-phonological features of words do not contribute to *phonemic* adaptation. They do not show that lexicality or other non-phonological features of words do not adapt—at least not straight-forwardly. This would require a bridging hypothesis to the effect that, if words adapt, then that adaptation should also affect phonemic adaptation.

Is there reason to think that words, or non-phonological features of words, adapt? As mentioned, some argue that there is semantic adaptation (Nes 2016, Brogaard 2020). Their argument adverts to semantic satiation, suggesting that it is a kind of adaptation. Semantic satiation is the familiar phenomenon of words seeming to be drained of

---

[4] The phrase 'phonemes in fact in the stimulus' would need to be rewritten to accommodate fictionalism about phonemes—e.g., Rey (2020). Note that Block's characterization of iconicity—provided below—seems to preclude fictionalism about instances of properties represented in perception.

meaning after repetition. The phenomenon has been experimentally investigated by assuming it is related to decreased response times on various tasks.

Block also discusses semantic satiation, but for a different reason. Nes and Brogaard, like Block, treat adaptation as a mark of the perceptual. They then argue that, if semantic satiation is a kind of adaptation, it provides reason to class meaning comprehension as perceptual. Block, on the other hand, sees representations of word meaning as obviously cognitive. He therefore worries that, if semantic satiation is a kind of adaptation, there would be cognitive adaptation and so adaptation might not be a good mark of the perceptual after all. Block's reply is that semantic satiation is too different from clear cases of adaptation to be grouped with it. He cites studies apparently showing, for example, age-specificity and task-dependence (86). In my discussion of alleged evidence for semantic perception (Gross forthcoming), I provide a different reason for not treating semantic satiation as a kind of adaptation. I want to take one aspect of my remarks there a bit further, as doing so will lead to a *prima facie* problem for Block's conception of the joint.

Tian & Huber (2010) ask whether semantic satiation involves satiation of meaning, satiation of lower-level lexical features (e.g., orthographic or phonetic features), or satiation of the "association" between meaning and lower-level lexical features.[5] They had subjects perform speeded matching tasks involving visually-presented labels for categories (FRUIT) and sub-categories thereof (APPLE). In one, subjects across multiple trials saw a category label followed by a sub-category label and then judged whether the sub-category fell under the category. In another, subjects saw sub-category labels followed by distinct sub-category labels and judged whether they fell under the same category. In the last experiment, the task was simply to judge whether the first and second label were identical. In all three experiments, an adaptor was repeated across half of the trials. In the first and last experiment, a specific label was repeated. In the second experiment, the adaptor was a specific category, effected by using in half the trials sub-category labels drawn from that specific category. Lower-level lexical satiation predicts slower responses for the repeated label in the first and last experiment. Meaning satiation predicts slower responses for the first two experiments. Association satiation predicts slower responses only for the first experiment, which is what they found.

The rejection of meaning satiation may seem to be a blow to defenders of semantic perception. But the question of word adaptation remains. The "association" between meaning and lower-level lexical features can be interpreted as part of what a (representation of a) word is. Words, on the construal at issue, link phonological, semantic, and syntactic features—as well as other features, such as orthography if one is

---

[5] Tian & Huber use 'lexical representation' for the orthographic and phonemic components of word representations. Their usage does not include the representation of syntactic and semantic features. Block's use of 'word' does (there are distinct words with the phonology /bank/ and the orthography 'bank', represented as such by language-users). I speak of 'lower-level lexical features' to signal this difference. Similarly, a visual word form is not a word in Block's sense. Block use of 'word' is consonant with linguists' use of 'lexeme' for representations that link sound, syntax, and (aspects of) meaning. (The *vernacular* 'word' doesn't exactly align with what language users represent: 'chair' and '-s' may be explicitly represented in the lexicon, but not the complex 'chairs' which may be constructed as needed.)

literate. (And defenders of semantic perception can say that what they claim is perceived is a phonemic or orthographic form *having* a particular meaning.) If semantic satiation is association satiation, and if association satiation is a kind of adaptation, then arguably words adapt.

But perhaps satiation is not a kind of adaptation. The reason Gross (forthcoming) provides for thinking it is not is that, in standard cases of adaptation, if a feature adapts, then so do lower-level features on which it depends (Webster 2015). Indeed, this is why demonstrating adaptation of higher-level phenomena can be so hard. If satiation is a kind of adaptation, we should expect to find lower-level lexical satiation as well. But, as reported above, Tian & Huber did not find lower-level lexical satiation.

It's important to note, however, that, had lower-level lexical satiation been found, this *also* would have undermined the evidence for semantic perception. For satiation, in this case, could be attributed to satiation of lower-level features. This is worth pointing out for several reasons. First, there are some questions of experimental design one could raise about Tian & Huber's third experiment (see Gross forthcoming). Second, since phonemes *do* adapt, it's natural to wonder whether phonemic lower-level lexical satiation would be found if Tian & Huber's experiment could be adapted to auditory stimuli. Third, another study found that visual word forms *do* adapt. It's this study that leads to a *prima facie* problem for Block.

Hanif et al. (2013) investigated visual word form adaptation using a task that parallels one used for faces (cf. Butler et al. 2008, which Block cites—both papers come from Jason Barton's group). They created ambiguous visual word forms by morphing two unambiguous visual word forms. Across a series of experiments, subjects adapted to one of the unambiguous word forms and then were presented with an ambiguous morph. Their task was to report which of two test items most resembled the ambiguous morph. The test items were visual word forms orthographically identical to those used in creating the ambiguous morphs. Across the experiments, the test items differed from the adaptor and morph regarding various lower-level features: handwriting style, computer font, or case. (See Figure 1.) [INSERT Fig. 1 SOMEWHERE AROUND HERE] The main result was that reports were biased away from the adaptor. Hanif et al. thus found an adaptation effect at a higher level than the level at which Tian & Huber failed to find a satiation effect, since Tian & Huber did not vary case. In other words, it seems that case-insensitive visual word form representations adapt. Hanif et al.'s result is consistent with studies that have found case-insensitive *neural* adaptation—decreased activity following repeated or prolonged exposure to a stimulus—in the visual word form area (Dehaene et al. 2001).

Here's the *prima facie* problem for Block. Case-insensitive visual word forms adapt. If adaptation is a compelling mark of the perceptual, then they are perceived. But perceptual representations, according to Block, are constitutively iconic. It is not obvious, on Block's view, that representations of case-insensitive visual word forms are iconic.

On Block's favored account of iconicity, iconic representations exhibit analog tracking and mirroring. As he elaborates:

> Analog tracking and mirroring obtains when there is a set of environmental properties and a set of representations of those environmental properties such that:
>
> 1. Certain differences in representations function as responses to differences in environmental properties in a way that is sensitive to the degree of environmental differences. …
>
> 2. Certain differences in representations function to alter the situation that is represented in a way that depends on the degree of representational change.
>
> 3. Certain relations (including temporal relations) among the environmental properties are mirrored by representations that instantiate analogs of those relations. (182-3)

Block emphasizes the importance of degrees in difference in this characterization. But it is not obvious how this applies to case-insensitive orthography. It is also not obvious what relations among case-insensitive orthographic forms are mirrored by relations that instantiate analogs of those relations—perhaps transitional probabilities. Burge's (2018, pp. 80-82) somewhat similar conception of iconicity builds in a requirement that the relation between representation and represented environmental property be in some sense "natural". It is unclear there is a reasonable sense of naturalness in which case-insensitive orthographic forms and our representations of them would satisfy such a requirement. Perhaps naturalness is not a requirement Block would endorse. If so, that would be yet another difference between Block and Burge's otherwise similar views.

Case-insensitive orthography briefly comes up in Block's discussion of E. J. Green and Jake Quilty-Dunn's (2021) argument that perceptual object representations are conceptual. Green and Quilty-Dunn advert to the object-specific preview benefit: seeing some feature instantiated on an object improves speed and accuracy when identifying a later presentation of that feature instantiated on the same object in comparison to a presentation not instantiated on that object. The effect is found across changes in case, as when 'bread' is previewed and 'BREAD' is the test item. Block ascribes to Green and Quilty-Dunn the claim that the representation, since it's case-insensitive, abstracts from spatial features and is thus non-iconic (cf. Quilty-Dunn 2019). Block does not question this claim but rather argues that what's responsible for the object-specific preview benefit are conceptual aspects of working memory representations, not would-be conceptual aspects of perceptual representations. Block's reply might indicate agreement that case-insensitive visual word form representations are not—or not obviously—iconic.[6]

---

[6] One might relate the abstraction from spatial features to another aspect of case-insensitive visual word form representation that can seem to violate a necessary condition of perception—viz., that they are not viewpoint-specific. Block (215) states that perception is viewpoint-specific. But, as he notes elsewhere (98), there is evidence of viewpoint-neutral perceptual representations of faces, which would show that

There are various ways one might reply to this *prima facie* problem. One might accept that there is adaptation and that case-insensitive visual word forms are not iconic, but shrug them off as edge-cases that no more challenge the joint than twilight challenges the distinction between night and day (cf. 21). Alternatively, one might grant the adaptation and the non-iconicity but suggest that the bearing of the case-insensitive visual word forms on the adaptation is indirect: given the overlearned associations between orthography and phonology, perhaps the adaptation actually reflects a phonological effect. A different appeal to association would suggest that representation of a letter of one case tends to activate representation of that letter in the other case, allowing adaptation to spread across cases. Appeal to case-insensitive visual word form representations would then not be necessary to explain the adaptation results, except insofar as they mediate the association.[7] Presumably, though, a different story would be needed for adaptation across fonts and handwriting styles. Finally, one might try denying either that the effect is in fact an *adaptation* effect or that case-insensitive visual word forms are in fact non-iconic. Interestingly, there may be a way of pursuing this strategy that would tie the possibility of adaptation more generally to iconicity. This would provide a further way that constitutive features of perception (here, iconicity) explain why various empirical marks (here, adaptation) cluster with clear cases of perception. The idea is that, for the effect to count as truly an *adaptation*, the bias away from the adaptor must be a repulsion along some dimension of a feature space. But if one allows that there is a space with such dimensions, then one has at least removed some obstacles towards defending the iconicity of case-insensitive visual word forms. It might be that the degrees of difference crucial for tracking and analog mirroring live in this more abstract space. This idea, if it could be plausibly developed, might yield a dilemma for the proponent of the problem case: either the effect is not adaptational, or (if naturalness is not required) the possibility of iconicity has not been precluded.

III. Language and the Transition from Perception to Cognition

Much of Block's discussion can be read as limning (at least for "paradigm cases" (214)) roughly the following picture of the transition from perception to cognition. With attention, perceptual states transition to content-related cognitive states in working memory. Neurophysiologically, this transition is realized in a transition from locally recurrent patterns of neural activity in sensory cortex to more globally recurrent patterns involving both sensory cortex and areas implicated in cognition. The resulting working memory states are thus broadcast to cognitive systems more generally that are implicated in reasoning, planning, and the like. Because of their availability as premises and conclusions in reasoning, such states are conceptual. (They can contain "remnants" of the

---

perception can be viewpoint-neutral. I believe the latter is his considered view. Or perhaps the two remarks reflect some uncertainty.

[7] This would not require treating case-insensitive visual word form representations as disjunctive. (If they were disjunctive, that would also provide reason for denying they are perceptual: Block argues there are no disjunctive percepts (146).)

perceptual states, however—iconic elements in a "cognitive envelope" that are nonetheless conceptual in virtue of the role these conceptual states play.)

This picture aligns the transition from the perceptual to the cognitive with a transition to presence in working memory, global broadcasting, and being conceptual. (Block rejects the global workspace theory of *consciousness* but thinks it works as a theory of conceptualization. (11)) Among the reasons the picture is rough is that Block is open to the possibility that a three-way distinction is needed between the perceptual, the (non-conceptual) cognitive, and the conceptual. He notes that mental maps, for example, might be cognitive but not conceptual (102, 140, and 243). In this section, I suggest that language might provide further reason for thinking the alignments rough. Drawing further distinctions need not the threaten the joint Block defends, as he notes in discussing mental maps (243). But it complicates the picture.

We can start where the last section left off. Suppose case-insensitive visual word form representations are not perceptual. The transition to them from case-*sensitive* visual word form representations is thus a transition from the perceptual to the non-perceptual—indeed, let's suppose, to the cognitive. But it is doubtful that this transition, held to occur within the visual word form area itself (Dehaene et al. 2001), is a transition to working memory. Nor is there reason to think it involves conceptualization—a transition to a representation available generally for inference.

This might just provide further reason for thinking case-insensitive visual word form representations are perceptual. So, let's consider what comes next: the transition from word form representations—whether phonological or orthographic—to representations of words, which link these forms to syntactic and semantic information, and thence to parsing and interpretation.[8] These representations are clearly cognitive on Block's view; but are they conceptual? Conceptual representations "constitutively function in propositional thought, reasoning, problem-solving, evaluating, deciding, and other cognitive processes and states" (139). The representations of syntactic features in the lexicon play no such role in thought and inference. (Even in syntax seminar, when one attributes a syntactic feature to a word, it is not the attribution in the lexicon one is expressing: the attribution in the lexicon is what one is hoping to get right.) Moreover, the memory store used in parsing to maintain and manipulate syntactic representations—which seems rather different from standard visual and verbal working memory models (Caplan & Waters 2013)—does not make these representations generally available for reasoning. The same could be said about semantic features, especially on views that emphasize the distance between semantic features and conceptual content (cf. Petroski 2018, according to whom semantics provides instructions for concept construction).

---

[8] My talk of what's next is not meant to entail that word form perception must be complete before word determination begins, nor that there can't be interaction across levels. For example, Gwilliams et al. (2018) present evidence that phonetic features incrementally activate candidate lexical representations prior to completion of word-form level phoneme assignment and that, relatedly, this lexical information can help resolve phoneme ambiguity. Recall also Samuel 1997 mentioned above.

These remarks also bear on a further thesis one might add to the picture—namely, that the transition from perception to cognition is never ampliative (never adds content), at least when functioning properly. This thesis seems to be endorsed by Burge. It is unclear whether Block would endorse it. If not, this is a further divergence in their views.

Ampliative transitions certainly occur *within* cognition. The range of ampliative inference in human cognition is among its most glorious and mysterious features. Ampliative transitions occur within perception as well, as when it transits from representations of geometric configurations of certain sorts to representations as of a face. For both Block and Burge, these intra-perceptual transitions are not inferential: perceptual representations lack the logical structure necessary to serve as premises and conclusions in inference (139-156). For the same reason, transitions from the perceptual to the cognitive are not inferential on their view. But this does not yet preclude their being ampliative. To claim that they are not is a substantive thesis, perhaps encouraged by the idea that the function of memory—in this case, working memory—is preservative, even if, having entered working memory, the representation may then be available for manipulation and ampliation.

Burge (2020, pp. 55-59) suggests the non-ampliation thesis in the course of characterizing basic perceptual beliefs and their relation to perceptions. Simplifying for reasons of space, the content of basic perceptual beliefs is fully derived from the content of a perceptual state. Attributive aspects of the belief are conceptualizations of perceptual attributions. Using the copula to mark the propositional structure of a belief, a basic propositional belief derived from a perception with the content <That, red square> could be <That square is red>. Burge holds that "all empirically supported beliefs either are or are supported by [basic] perceptual beliefs" (2020, p. 57). That is, they all require a non-ampliative transition from perception. This does not preclude cases where perception gives rise to a belief with a different content: misconceptualizations and other epistemologically relevant errors can occur. Note also that not all of a perception's content need be taken over by the basic perceptual belief. Indeed, if we consider a total perceptual state, or even just a total visual state, the basic perceptual belief *could* not take over all the content if one accepts Block's view that the content of perception "overflows" that of working memory (35—cf. Gross and Flombaum 2017). Finally, we may want to carve out space for ways that a conceptualization may alter the content without it counting as ampliation in the sense at issue. I have in mind, for example, the general point that perceptual representations are more fine-grained than conceptual representations (Evans 1982), as well as empirical results showing, for example, a bias in the transition to working memory of color representations towards the categorical center (Bae et al. 2105).

Block introduces a notion of minimal immediate direct perceptual judgment (14) which he suggests could serve as a useful precisification of Burge's notion (101). These judgments are minimal in conceptualizing each representational aspect of a perception and no more. (We can assume perceptions are being individuated more narrowly than total perceptual states.) They are immediate in being triggered by perception with no inference and in being less committal than belief. And they are direct in being based in

the perception with no person-level causal intermediary (cf. 47). Block sees such judgments as being the hardest cognitive states to distinguish from corresponding perceptions. In such cases, there is a transition from perception to cognition without ampliation. It would be a further claim that, more generally, normally functioning transition from perception to cognition does not involve ampliation—that we are *always* presented with the hardest case when things go well. Would Block endorse some version of Burge's thesis?

Language poses a *prima facie* challenge to Burge's thesis. The transition from case-sensitive to case-insensitive visual word forms is ampliative. And, if that transition is not a transition from the perceptual to the cognitive, the transition from case-insensitive visual word forms to words (with their syntactic and semantic features) is ampliative as well, as is the transition from phonological representations to words.

One might try preserving Burge's thesis by denying that it precludes ampliative transitions from perception to cognition. Burge's claim that "all empirically supported beliefs either are or are supported by [basic] perceptual beliefs" is consistent with some empirical beliefs being supported by basic perceptual beliefs *as well* as perceptual beliefs arrived at by an ampliative transition from perception—perhaps the same perception that gives rise to the supporting basic perceptual belief. But though this reply is consistent with Burge's wording, I am not sure it is consistent with the spirit of Burge's position. Moreover, it still requires a non-ampliative transition for each empirically supported belief, and this remains subject to question. We can see that this is so, perhaps especially by Block's lights, by considering a different response to the challenge.

This reply grants the no-ampliation thesis but claims that it is not violated in utterance comprehension. On this view, comprehension involves a non-ampliative transition from a perceptual word-form representation (whether phonological, orthographic, or some other) to a conceptual word-form representation.[9] The transition to representations of words then occurs post-perceptually. This is in line with Burge's (2020, pp. 57-9) remarks on post-perceptual retrieval of associations in long-term memory. His example is the retrieval, given perception as of a body of a certain size and shape, of the conceptual attributive X-RAY MACHINE via its association in long-term memory with the concept of bodies of that size and shape. In our case, an example would be the retrieval, given perception as of phonemes /ball/, of the conceptual representation of the word *ball*—hopefully, the relevant one—via the association in long-term memory of the conceptual representation of the phonemes with that word (or with the conceptual representation of its syntactic and semantic features, though we have already questioned whether those representations are conceptual).

The question is whether utterance comprehension indeed requires conceptualization of word-forms. Must the perception of word-forms transit to a conceptualization thereof in

---

[9] If we distinguish the cognitive and the conceptual, as above, we might allow a transition to a cognitive, but not conceptual word-form representation. But that doesn't suffice, for Burge, to yield basic perceptual belief. And if there were ampliative processing of the cognitive representation (e.g., in parsing) prior to a transition to the conceptual, then the resulting belief would not be basic.

order to gain access to words? A remark of Block's might suggest a commitment to denying that it must: he claims that most humans lack concepts of phonemes (227).[10] This claim is open to challenge. Normal literate adults do fine with a variety of phonological awareness tasks that plausibly require concepts of phonemes—similarly with kids exposed to literacy instruction, 50% of whom achieve phonological awareness skills around ages 4.5-5 (Dodd & Gillon 2001). But one can interpret Block's comment as not limited to current humans, literacy being a very recent development in our history. A different reply, in light of Block's remark, is that, even if one did not possess concepts of phonemes, one could possess concepts of larger phonological groupings—for example, concepts of syllables. In fact, phonological awareness of syllables developmentally precedes phonological awareness of phonemes (Caravolas & Bruck 1993). Conceptualization of larger phonological groupings would meet the needs of Burge's thesis. It remains the case, however, that young kids' linguistic comprehension precedes their success with phonological awareness tasks and that illiterate adults display poor phonological awareness (Morais & Kolinksy 2019). Absent an alternative explanation of failure to succeed on these tasks, this might indicate utterance comprehension without word-form conceptualization. The transition from perception to cognition might not require this conceptualization (or any other) and might be ampliative.

In conclusion: Language may not lead us to question the joint between perceiving and thinking. But reflection upon it yields fruitful considerations for developing further a conception of the joint. I am very interested to hear Block's take.

References
Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. Journal of experimental psychology. General, 144(4), 744-763.
Block, Ned. Forthcoming. *The Border Between Seeing and Thinking*.
Burge, Tyler. 2018. Iconic representation: Maps, pictures, and perception. In Wuppuluri Shyam & Francisco Antonio Dorio (eds.), *The Map and the Territory: Exploring the Foundations of Science, Thought and Reality*. Springer. pp. 79-100.
Burge, Tyler 2020. Entitlement: The basis for epistemic warrant. In Peter Graham and Nikolaj Pedersen (eds.), *Epistemic Entitlement*. Oxford University Press, pp. 37-142.

---

[10] As context makes clear, the claim is not just that we lack the concept PHONEME. Block is arguing that color perception is non-conceptual in young kids because they perceive colors though they lack color concepts. In the case of phonemes, he makes a parallel claim even for most adults. Similarly, the parallel would not hold if he meant rather something like that we lack non-phenomenal or non-iconic/non-perception-based concepts of phonemes.

Butler, Andrea, Ipek Oruc, Christopher Fox, & Jason Barton. 2008. Factors contributing to the adaptation aftereffects of facial expression. *Brain Research* 1191: 116-126.

Brogaard, Berit. 2017. The Publicity of Meaning and the Perceptual Approach to Speech Comprehension. Protosociology 34, pp. 144-62.

Brogaard, Berit. 2020. Seeing and Hearing Meanings: A Non-Inferential Approach to Speech Comprehension. In Timothy Chan and Anders Nes (eds.), Inference and Consciousness, Routledge.

Caplan, David, & Gloria Waters 2013. Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review* 20: 243-68.

Caravolas, Marketa, & Maggie Bruck. 1993. The Effect of Oral and Written Language Input on Children′s Phonological Awareness: A Cross-Linguistic Study. *Journal of Experimental Psychology* 55: 1-30.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., Riviere, D., 2001. Cerebral mechanisms of word masking and unconscious repetition priming. Nat. Neurosci. 4, 752–758.

Dodd, Barbara, & Gail Gillon. 2001. Exploring the Relationship Between Phonological Awareness, Speech Impairment, and Literacy. *Advances in Speech Language Pathology* 3: 139-147.

Eimas P, Corbit J. 1973. Selective adaptation of linguistic feature detectors. Cogn. Psychol. 4: 99–109.

Evans, Gareth. 1982. *Varieties of Reference*. Oxford: Oxford University Press.

Ferreira, F. & Nye, J. (2017). The modularity of sentence processing reconsidered. In R. G. De Almeida & L. R. Gleitman (Eds.), *On Concepts, Modules, and Language: Cognitive Science At Its Core* (pp. 64–85). Oxford: Oxford University Press.

Fodor, Jerry. 1983. *Modularity of Mind*. MIT Press.

Green, E. J. and Quilty-Dunn, J. 2021. What is an object file? *British Journal for the Philosophy of Science* 72, 665-699.

Gross, Steven. Forthcoming. Is there an empirical case for semantic perception? *Inquiry.*

Gross S, Flombaum J. 2017. Does Perceptual Consciousness Overflow Cognitive Access? The Challenge from Probabilistic, Hierarchical Processes. *Mind & Language* 32: 358–91.

Gwilliams, Laura, David Poepel, Alec Marantz, and Tal Linzen. 2018. Phonological (un)certainty) weights lexical activation. *Cognitive Modeling and Computational Linguistics*.

Hanif, Hashim, Brielle Perler, & Jason Barton. 2013. The visual representations of words and style in text: An adaptation study. *Brian Research* 1518: 61-70.

Morais, José & Régine Kolinksy 2019. Phonological abilities in fully illiterate adults. In Delores Perin (ed.), *The Wiley Handbook of Adult Literacy.* Hoboken: Wiley, pp. 41-61.

Nes, Anders. 2006. "On what we experience when we hear people speak." Phenomenology and Mind 10, 58-85.

Phillips, Ian & Chaz Firestone. This issue. Visual adaptation and the purpose of perception.

Pietroski, Paul. 2018. *Conjoining Meanings: Semantics Without Truth-Values*. Oxford University Press.

Quilty-Dunn, Jake. 2020. Perceptual pluralism. Nous 54, pp. 807-38.

Rey, Georges. 2020. *Representation of Language: Philosophical Issues in Chomskyan Linguistics*. Oxford: Oxford University Press.

Samuel, Arthur. 1997. Lexical activation produces potent phonemic percepts. Cognitive Psychology 32, pp. 97-127.

Samuel, Arthur. 2001. Knowing a word affects the fundamental perception of the sounds within it. Psychol. Sci. 12: 348–51.

Shea, Nicholas. 2015. Distinguishing top-down from bottom-up effects. In D. Stokes, M. Matthen, & S. Biggs (eds.), *Perception and Its Modalities*. Oxford: Oxford University Press, pp. 73-91.

Tian, X., Huber, D. E. (2010). Testing an associative account of semantic satiation. Cognitive Psychology, 60, 267–290.

Webster, Michael. 2015. Visual Adaptation. *Annual Review of Vision Science* 1: 547–67.
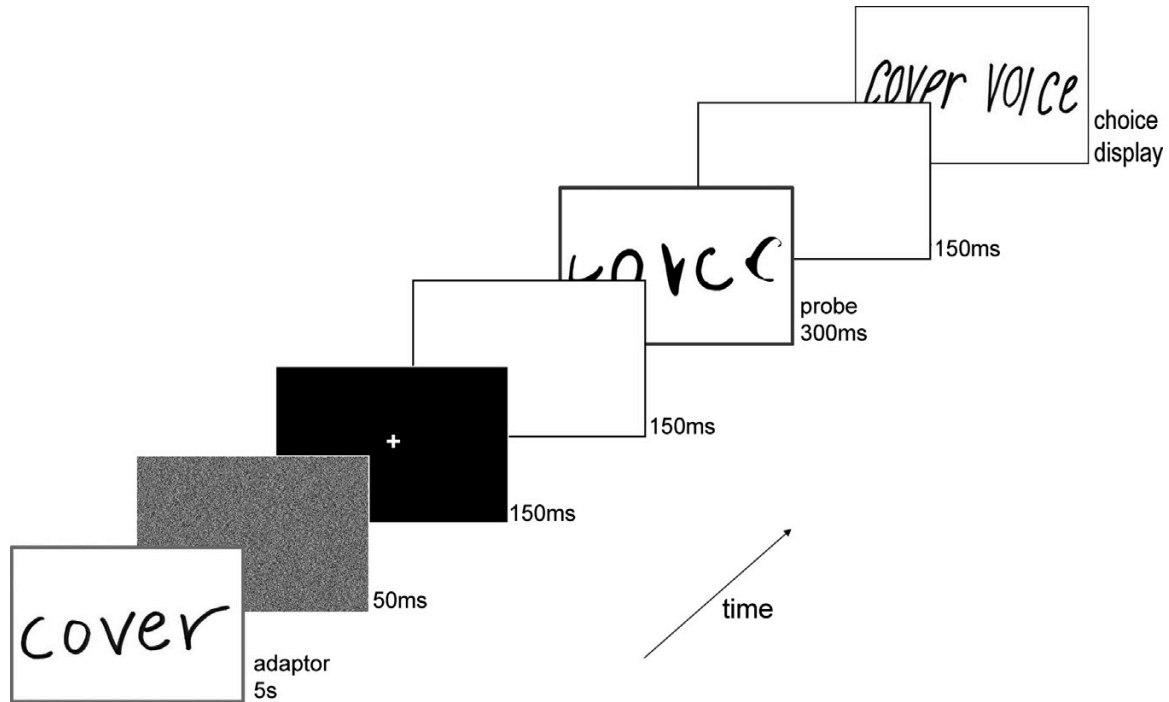
Figure 1

Fig. 1 – Trial example. Each trial begins with a 5s view of an adapting stimulus. This is followed by a 50ms mask, then a fixation cross at screen center for 150ms, followed by a 150ms blank screen, and then a 300ms glimpse of the ambiguous probe stimulus. After another blank screen of 150ms, a choice display appears and subjects must choose which of the two items the probe resembled more. [Fig. 4 of Hanif et al. 2013 reproduced with their permission.]