On the speed of conscious perception: How soon is now?

William Turner[1], Hinze Hogendoorn[2], Laura Gwilliams[1,3,4]

[1]Department of Psychology, Stanford University
[2]School of Psychology and Counselling, Queensland University of Technology
[3]Wu Tsai Neurosciences Institute, Stanford University
[4]Stanford Data Science, Stanford University

Correspondence: wft@stanford.edu

Abstract

Fleming and Michel propose that conscious perception is 'slow', with a delay of 350-450ms. But this claim is premature. Here, we will show that the speed of conscious perception remains unresolved. Examining evidence from vision and language research, we will explore how this fundamental question may ultimately be answered, to test the validity of this foundational claim.

Main Text

To understand the function of conscious perception, Fleming and Michel consider its speed. They start by asking: how much time does it take for conscious perception to emerge? If an apple is unexpectedly thrown at you, how long would it take to have some conscious experience of it?

This question has been richly debated, with proposals put forward for 'fast' models in which conscious percepts arise after little to no delay (e.g., Grush 2005; Hogendoorn, 2022), 'slow' models in which conscious percepts emerge after sustained unconscious processing (e.g., Herzog, Drissi-Daoudi, & Doerig, 2020), and models questioning this distinction (e.g., the 'multiple drafts' model, Dennett & Kinsbourne, 1992). Central to this debate are 'postdictive' phenomena, in which one's conscious experience of an event is affected by subsequent events. These indicate either that conscious perception is delayed (the 'slow' perspective), or that conscious perception can be updated in some way (the 'fast' or 'multiple drafts' perspectives).

Fleming and Michel endorse the slow perspective, proposing a delay of 350-450ms. However, in doing so they gloss over the extreme difficulty of constraining theoretical positions with empirical findings in this space. Indeed, Dennett and Kinsbourne (1992) famously argued that 'fast' and 'slow' perspectives (in their words, 'Orwellian' and 'Stalinesque' accounts) are observationally equivalent. That is, both are sufficiently flexible as to account for all possible empirical observations. In the face of this potential impasse, Fleming and Michel take a potentially fruitful approach. Put simply, they weigh up the 'awkwardness' of the theoretical positions each perspective must adopt, given current data. However, we argue that they are premature in declaring a winner.

Fleming and Michel critique 'fast' models on two fronts. First, they state that these models cannot explain why participants respond equally quickly to the midpoint or endpoint of an apparent motion stimulus (Cowan & Greenspahn, 1995). But this is incorrect. In 'timeline' models (e.g., Grush, 2005; Hogendoorn, 2022; Jiang & Rao,

2024) perceptual estimates are updated synchronously with incoming sensory information, predicting equal response times. We therefore remain where Dennett and Kinsbourne left us, with both perspectives accounting for current observations.

Second, they claim that the 'fast' perspective is ad-hoc. If one accepts the 'overflow gambit' for long-lasting postdictive effects (i.e., events are fleetingly experienced but later unreportable), they argue, one must also accept it for short-term effects like masking (which proponents of overflow typically reject). However, given neural processing delays, conscious percepts are undeniably delayed to some extent (in unpredictable contexts). One may argue for a delay of ~150ms (see below), such that brief events are not experienced. Or, one may maintain that brief stimuli are indeed fleetingly experienced but later unreportable. Fleming and Michel write off this possibility, asking for conclusive behavioural proof. But, as we'll show next, finding a behavioural indicator that conclusively favours *either* perspective is an impossible task (Dennett & Kinsbourne, 1992).

Why so slow? The slow perspective also faces awkward questions. For example, consider some of the evidence for rapidly emerging complex perceptual representations: animacy judgements are possible in ~150ms (Kirchner & Thorpe, 2006), neural representations corresponding to conscious perception in binocular rivalry form in ~150ms (Krisst & Luck, 2025), and random vowel-consonant-vowel strings can be verbally repeated in ~200ms (Porter & Castellanos, 1980). So why wait 2-3x times longer for conscious perception to emerge? Fleming and Michel argue that waiting facilitates a trade-off between accuracy and immediacy. A stable snapshot of the world is gradually built and then 'fixed', and, in doing so, is somehow consciously rendered. But is such a trade-off necessary? Timeline models provide an ever-available, dynamically updating 'best estimate' of current and recent events. Of course, proponents of the slow perspective might argue that an analogous process occurs, but pre-consciously. Evolving representations may be behaviourally accessed, but only consciously rendered later on. And therein lies the problem. It is possible to conceive of two almost identical systems, which make identical behavioural predictions, only differing in an as-yet-unconstrained assumption about which, or when, certain representations are *consciously experienced*. With respect to current data, these accounts are observationally equivalent. In treating long-lasting postdictive effects as diagnostic here, Fleming and Michel are conflating evidence about the temporal window of *integration* with evidence about the timing of conscious perception. Thus, despite what they may wish, the speed of conscious perception remains unresolved.

Where to now? How might we ultimately determine the speed of consciousness? First, Dennett and Kinsbourne's (1992) argument for observational equivalence falters when neural observations come into play (see Block 1992). Investigation of the neural circuit level computations underlying perceptual processing may therefore prove crucial. As an example of nascent work in this vein, Gwilliams and colleagues (2018) found evidence that the brain waits ~450ms to commit to a categorical interpretation of incoming phonemes (e.g., /b/ vs. /p/), allowing word identity to postdictively update interpretations of earlier speech sounds. This evidence, of slowly emerging categorical representations corresponding to perception, could sway the balance of plausibility in favour of 'slow' models of (speech) perception (but see Dahan, 2010).

Second, the field should expand its focus beyond vision, with spoken language a useful added domain. Since information in speech obligatorily unfolds over time, perceptual mechanisms need to retain and integrate information over extended periods. This extended timing may enable experimenters to better disentangle conscious and unconscious representations, isolating transitions between states. Moreover, psycholinguistics has its own rich vein of postdictive phenomena, termed 'right context effects'. Careful work has already been done to isolate perceptual effects here (Samuel, 1981), laying the foundation for productive neuroimaging work.

Finally, metacognitive measures should be integrated into existing paradigms. For example, in the sequential metacontrast paradigm, postdictive and non-postdictive stimuli are assumed to be subjectively equivalent, but true equivalence requires establishing that they are metacognitively indiscriminable ('metacognitive metamers'). To do so, future studies may employ metacognitive measures (e.g., confidence ratings or ABX tasks), following research that overturned the assumption of perceptual equivalence between illusory and veridical flashes in the sound-induced flash illusion (e.g., Maynes et al., 2023; van Erp, Philippi, & Werkhoven, 2013).

Acknowledgements: We thank Joe Melling for insightful discussions.

Competing Interests Statement: The authors declare none.

Funding Statement: WT acknowledges funding from the American Australian Association. HH acknowledges funding from the Australian Research Council (DP180102268 and FT200100246). LG is funded by the BRAIN Foundation A-0741551370; Esther A. & Joseph Klingenstein Fund; Whitehall Foundation 2024-08-043.

References:

Block, N. (1992). Begging the question against phenomenal consciousness. *Behavioral and Brain Sciences*, 15(2), 205-206. doi:10.1017/S0140525X00068266

Cowan, N., & Greenspahn, E. (1995). Timed reactions to an object in apparent motion: Evidence on Cartesian and non-Cartesian perceptual hypotheses. *Perception & Psychophysics*, 57, 546-554. doi:10.3758/BF03213079

Dahan, D. (2010). The time course of interpretation in speech comprehension. *Current Directions in Psychological Science*, 19(2), 121-126. doi:10.1177/09637214103647

Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain sciences*, 15(2), 183-201. doi:10.1017/S0140525X00068229

Grush, R. (2005). Internal models and the construction of time: Generalizing from state estimation to trajectory estimation to address temporal features of perception, including temporal illusions. *Journal of Neural Engineering*, 2(3), S209. doi: 10.1088/1741-2560/2/3/S05

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, 38(35), 7585-7599. doi: 10.1523/JNEUROSCI.0065-18.2018

Herzog, M. H., Drissi-Daoudi, L., & Doerig, A. (2020). All in good time: long-lasting postdictive effects reveal discrete perception. *Trends in Cognitive Sciences*, 24(10), 826-837. doi: 10.1016/j.tics.2020.07.001

Hogendoorn, H. (2022). Perception in real-time: predicting the present, reconstructing the past. *Trends in Cognitive Sciences*, 26(2), 128-141. doi: 10.1016/j.tics.2021.11.003

Jiang, L. P., & Rao, R. P. (2024). Dynamic predictive coding: A model of hierarchical sequence learning and prediction in the neocortex. *PLOS Computational Biology*, 20(2), e1011801. doi: 10.1371/journal.pcbi.1011801

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11), 1762-1776. doi: 10.1016/j.visres.2005.10.002

Krisst, L. C., & Luck, S. J. (2025). Electroencephalographic Decoding of Conscious versus Unconscious Representations during Binocular Rivalry. *Journal of Cognitive Neuroscience*, 37(8), 1381-1390. doi: 10.1162/jocn_a_02308

Maynes, R., Faulkner, R., Callahan, G., Mims, C. E., Ranjan, S., Stalzer, J., & Odegaard, B. (2023). Metacognitive awareness in the sound-induced flash illusion. *Philosophical Transactions of the Royal Society B*, 378(1886), 20220347. doi: 10.1098/rstb.2022.0347

Porter R. J., & Castellanos, F. X. (1980). Speech-production measures of speech perception: Rapid shadowing of VCV syllables. *The Journal of the Acoustical Society of America*, 67(4), 1349-1356. doi: 10.1121/1.384187

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4), 474. doi: 10.1037/0096-3445.110.4.474

van Erp, J. B., Philippi, T. G., & Werkhoven, P. (2013). Observers can reliably identify illusory flashes in the illusory flash paradigm. *Experimental Brain Research*, 226(1), 73-79. doi: 10.1007/s00221-013-3413-8