Causal inference and the evolution of opposite neurons

Stephanie Badde[a] , Fangfang Hong[b] , and Michael S. Landy[b,c,1]

A pesky mosquito continues to annoy you and you are poised to swat it. You see it hovering above your arm, and feel a gentle tickle, but in a slightly different spot (Fig. 1A). Where should you strike? The mathematically optimal solution is to average the locations indicated by vision and touch, with greater weight given to the more reliable signal, the one that typically leads to smaller errors. A substantial literature indicates that for most modality pairings and perceptual tasks humans behave in accordance with this optimal prescription for sensory integration (1–4). However, if vision and touch indicate very different locations, the tickle might be due to another cause such as an old mosquito bite (Fig. 1B). In this case, it makes sense to segregate the sensory signals, ignore touch, and swat at the location indicated by vision. This decision requires making a "causal inference," that is, an inference as to whether two sensory signals derive from a common source or separate sources. Humans (5, 6) and monkeys (7, 8) behave as if they perform causal inference; they do not integrate signals unlikely to come from the same source. The challenging question is, how are sensory cue integration and causal inference implemented in the brain?

In PNAS, Rideaux et al. (9) demonstrate how the interplay between different types of neurons could accomplish both optimal integration and causal-inference judgments. They simulate a particularly puzzling but also well-researched case of multisensory perception, visual and vestibular signals of self-motion. These signals converge in brain areas that include the dorsal medial superior temporal area (MSTd) and the ventral intraparietal area (VIP). Neurons in these areas are often tuned for heading direction, that is, these neurons fire the most when sensory cues indicate a particular direction, and fire less and less the more the signaled direction differs from their preferred one. Many of the neurons that receive input from both modalities are congruent neurons: They have similar tuning for the two modalities (Fig. 1C). Thus, congruent neurons seem predestined to perform multisensory integration (10,

11). Curiously, many other neurons in MSTd and VIP are opposite neurons (Fig. 1D): They are tuned to visual and vestibular information indicating opposite heading directions, e.g., rightward motion signaled by visual stimuli and leftward motion signaled by vestibular stimuli (10, 12). Opposite neurons appear to be ideally equipped to detect when sensory signals arise from different sources. In turn, the interplay of congruent and opposite neurons might enable the brain to perform causal inference (10, 13). A direct test of this hypothesis would require simultaneous recordings of congruent and opposite neurons in MSTd and VIP as well as the neurons they project to, which is a near-impossible task. However, artificial neural networks make it easy to inspect the interconnected behavior of neurons across different brain areas.

Rideaux et al. (9) used a particularly clever approach to this problem. Rather than constructing an artificial neural network with a layer of hand-tuned congruent and opposite neurons, they trained an unconstrained artificial neural network to perform causal-inference judgments as well as self- and world-motion estimation and afterward inspected the tuning and connectivity of multisensory neurons. This multilayer, feedforward network had two sets of inputs, visual and vestibular. The visual inputs were short sequences of natural images that translated at various velocities in four directions (left–right, up–down, toward–away, and rotations about the line of sight). The vestibular inputs were from units tuned to velocities along each of these four axes and slightly corrupted by noise. Separately for each motion direction, the output neurons of the neural network determined motion velocity (trained to match the average of the vestibular and visual input velocities; Fig. 1E), world-motion velocity (trained to match the difference between the two input velocities; Fig. 1F), and a common-source judgment (trained to match a binary categorization of whether the difference between the two input velocities was large or small; Fig. 1G). Notably, the world-motion estimation task differs from multisensory integration in other

[a]Department of Psychology, Tufts University, Medford, MA 02155; [b]Department of Psychology, New York University, New York, NY 10003; and [c]Center for Neural Science, New York University, New York, NY 10003
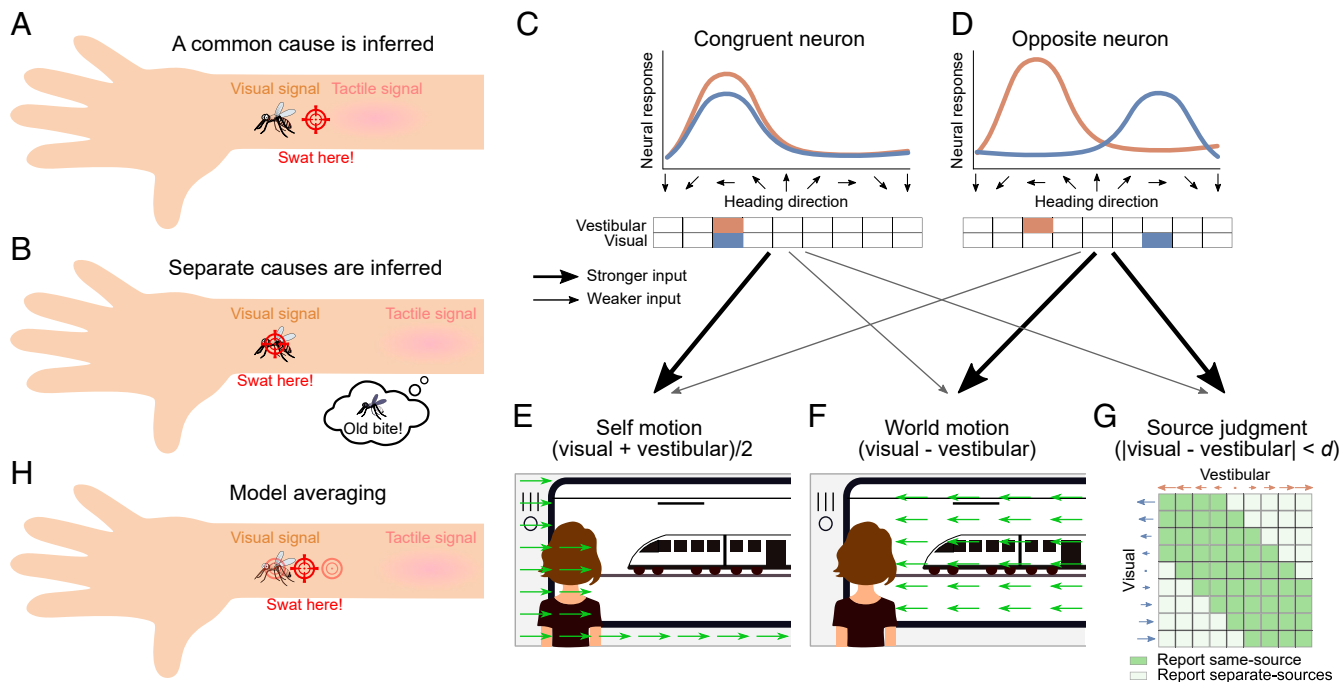Author contributions: S.B., F.H., and M.S.L. wrote the paper.

Fig. 1. Multisensory integration and causal inference. (*A*) When a common cause is inferred, sensory signals are integrated; (*B*) when separate sources are inferred, the segregated visual signal is used. (*C*) Congruent neurons have similar tuning for heading direction across modalities; (*D*) opposite neurons' preferred directions differ across modalities. Both types of neurons contribute to (*E*) self- and (*F*) world-motion estimation as well as (*G*) causal-inference judgments, but to different degrees. (*H*) In Bayesian estimation, the integrated and segregated estimates are combined with weights equal to the probabilities of each causal scenario.

domains, where integration is typically contrasted against segregation, i.e., reliance on one modality alone (compare Fig. 1 *A* and *B*). Thus, it is interesting how the network would generalize to multisensory perception of spatial, temporal, or other features.

The key contribution of the paper is that after successful training, the network developed neurons with the same characteristics as congruent and opposite neurons in macaque MSTd and VIP. More specifically, in the "MSTd" layer of the network, neurons had clear tuning for heading direction (where direction was computed from velocity along the left–right and forward–backward axes), and most neurons had either congruent visual and vestibular tuning or showed opposite motion direction tuning for the two modalities. Both types of neurons provided significant input for causal-inference judgments, confirming the initial hypothesis that the balance between congruent and opposite neurons is crucial for inferring whether two signals originate from the same source. Regarding the network's motion velocity percepts, congruent cells provided stronger input for self-motion estimates and opposite cells provided stronger input for world-motion estimates. Both types of neurons were also able to contribute to the other perceptual estimate, but to a lesser degree. Previous computational models with hand-tuned congruent and opposite neurons had already demonstrated that such networks are able to perform causal inference (13, 14). However, so are artificial neural networks without these properties (15). In contrast to these top–down approaches, Rideaux et al. (9) show that the requirement to make both perceptual and causal-inference judgments leads to the development of congruent and opposite neurons, suggesting that this neural substrate is the optimal solution for the computation.

The network's performance in the perceptual tasks qualitatively mirrored human and monkey behavior in heading-discrimination tasks (11). After successful training, the network was presented with visual and vestibular motion inputs with a small cross-modal discrepancy. For these tests, the visual stimulus was changed to a collection of moving dots; its reliability was manipulated by varying the proportion of dots moving in the same direction. The network integrated visual and vestibular inputs according to their reliabilities: Self-motion estimates agreed more with the vestibular input if the visual input was less reliable and toward the visual input, if it was more reliable. Notably, during training, self-motion estimates were reinforced to match the 50–50 average of the visual and vestibular signals, raising the question whether the ability to perform reliability-weighted integration emerged from the combined training of several tasks or was due to the architecture of the network. The influence of visual input on self-motion estimates was lower when the network inferred separate causes than when it inferred a common source of the two signals. Such a difference in cross-modal biases automatically emerges if one input modality is noisy, that is, if the same stimulus results in slightly different internal measurements across trials (6). However, in its current form, the network will not be able to reproduce the behavioral hallmark of causal inference: the reduction of cross-modal biases when the signals are more discrepant and thus less likely to have emerged from the same source. Given that the model is strictly feedforward, with separate outputs for causal inference and self- and world-motion estimates, the model is essentially unable to have the causal-inference judgment affect self-motion estimates. In other words, the model is unable to ignore the tickling sensation from the old mosquito bite in our introductory example. In contrast, Bayesian causal-inference models (6) replicate the nonlinear dependency of cross-modal biases on the discrepancy between the two signals (6, 16). They achieve this by summing the integrated and segregated estimates, weighted by the inferred probability of a common and

separate sources, respectively (Fig. 1*H*). In this view, the neural network model by Rideaux et al. (9) encompasses the first stage of a two-stage perceptual process. In fact, human brain activity in a multisensory context is consistent with separate representations for integrated, segregated, and the combined final estimates (17, 18). Another key component of the Bayesian approach to causal inference is the assumed prior probability of a common source. In the model of Rideaux et al. (9), this prior might be reflected in the weights of the connections between the multisensory MSTd layer and the output layer. However, this common-cause prior changes with the experimental context (19, 20), suggesting the need for additional input to the causal-inference process. Thus, a more complete model of multisensory integration and causal inference will require room for representations of both same-

and separate-source perceptual estimates along with flexible common-source priors.

Rideaux et al. (9) provide a convincing solution to the puzzle of the role of congruent and opposite neurons in causal inference. Training an artificial neural network to concurrently derive estimates of self- and world-motion as well as causal-inference judgments resulted in the development of congruent and opposite neurons. Whether the brain implements these inferences in this simple feedforward manner, and how causal inference is involved in perceptual judgments, are important future questions. Similarly, it would be fascinating to explore whether sensory experience is required for the development of opposite neurons during ontogenesis, or whether the processes invoked by training of this artificial neural network have played out during evolution.

1 M. O. Ernst, M. S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
2 M. S. Landy, M. S. Banks, D. C. Knill, "Ideal-observer models of cue integration" in *Sensory Cue Integration*, J. Trommershäuser, K. Körding, M. S. Landy, Eds. (Oxford University Press, New York, 2011), pp. 5–29.
3 M. S. Landy, L. T. Maloney, E. B. Johnston, M. Young, Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Res.* **35**, 389–412 (1995).
4 D. Alais, D. Burr, The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**, 257–262 (2004).
5 S. Gepshtein, J. Burge, M. O. Ernst, M. S. Banks, The combination of vision and touch depends on spatial proximity. *J. Vis.* **5**, 1013–1023 (2005).
6 K. P. Körding *et al.*, Causal inference in multisensory perception. *PLoS One* **2**, e943 (2007).
7 K. Dokka, H. Park, M. Jansen, G. C. DeAngelis, D. E. Angelaki, Causal inference accounts for heading perception in the presence of object motion. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9060–9065 (2019).
8 W. Fang *et al.*, Statistical inference of body representation in the macaque brain. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20151–20157 (2019).
9 R. Rideaux, K. R. Storrs, G. Maiello, A. E. Welchman, How multisensory neurons solve causal inference. *Proc. Natl. Acad. Sci. U.S.A.* **118**, 10.1073/pnas.2106235118 (2021).
10 Y. Gu, D. E. Angelaki, G. C. Deangelis, Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* **11**, 1201–1210 (2008).
11 C. R. Fetsch, A. Pouget, G. C. DeAngelis, D. E. Angelaki, Neural correlates of reliability-based cue weighting during multisensory integration. *Nat. Neurosci* **15**, 146–154 (2011).
12 A. Chen, G. C. Deangelis, D. E. Angelaki, Functional specializations of the ventral intraparietal area for multisensory heading discrimination. *J. Neurosci.* **33**, 3567–3581 (2013).
13 H. R. Kim, X. Pitkow, D. E. Angelaki, G. C. DeAngelis, A simple approach to ignoring irrelevant variables by population decoding based on multisensory neurons. *J. Neurophysiol.* **116**, 1449–1467 (2016).
14 W. H. Zhang *et al.*, Complementary congruent and opposite neurons achieve concurrent multisensory integration and segregation. *eLife* **8**, e43753 (2019).
15 I. Yamashita, K. Katahira, Y. Igarashi, K. Okanoya, M. Okada, Recurrent network for multisensory integration-identification of common sources of audiovisual stimuli. *Front. Comput. Neurosci.* **7**, 101 (2013).
16 M. T. Wallace *et al.*, Unifying multisensory signals across time and space. *Exp. Brain Res.* **158**, 252–258 (2004).
17 T. Rohe, U. Noppeney, Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol.* **13**, e1002073 (2015).
18 Y. Cao, C. Summerfield, H. Park, B. L. Giordano, C. Kayser, Causal inference in the multisensory brain. *Neuron* **102**, 1076–1087.e8 (2019).
19 S. Badde, K. T. Navarro, M. S. Landy, Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. *Cognition* **197**, 104170 (2020).
20 R. Gau, U. Noppeney, How prior expectations shape multisensory perception. *Neuroimage* **124** (Pt A), 876–886 (2016).

Badde et al.
Causal inference and the evolution of opposite neurons

PNAS | 3 of 3
https://doi.org/10.1073/pnas.2112686118