UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Iannetti, G. D., & Cimpian, A. (2026). The Neuroscience of Empathy: Why Methods Matter. *Collabra: Psychology*, *12*(1). https://doi.org/10.1525/collabra.159918

Methodology and Research Practice

The Neuroscience of Empathy: Why Methods Matter

Gian Domenico Iannetti[1][a], Andrei Cimpian[2][b]

[1] Neuroscience and Behaviour Laboratory, Italian Institute of Technology, Rome, Italy, [2] Department of Psychology, New York University, New York, USA

Keywords: Empathy, Prosocial behavior, Social neuroscience, Neural overlap, Pain, Brain decoding, Anterior cingulate cortex (ACC)

https://doi.org/10.1525/collabra.159918

Collabra: Psychology

Vol. 12, Issue 1, 2026

Empathy and prosocial behavior are fundamental to social life, and understanding their neural mechanisms is a central goal of social neuroscience. Animal models offer unique opportunities to investigate these mechanisms at the cellular level, but realizing this potential requires rigorous methods. In this commentary, we raise a set of methodological concerns about an article published in *Nature* by Zhang et al. (2024), who claim to have identified a cortical mechanism in mice for helping conspecifics in pain. We identify several issues that undermine the authors' conclusions: inconsistent criteria for interpreting neural overlap as evidence for distinct versus shared representations; decoder results that are at odds with the claimed representational structure; potential biases arising from fixed experimental order; and a mismatch between region-wide experimental manipulations and population-specific mechanistic claims. We situate these concerns within a broader methodological challenge in neuroscience: the difficulty of inferring functional specificity from limited experimental contrasts. Wherever feasible, we offer concrete suggestions for future research. By calling attention to these methodological challenges, this commentary aims to strengthen standards for mechanistic inference in social and systems neuroscience.

The capacity to perceive and respond to others' distress is a cornerstone of social life. Across many species, individuals can detect when conspecifics are injured or distressed and often engage in behaviors that reduce harm or promote recovery (e.g., de Waal & Preston, 2017; Decety & Jackson, 2004). Understanding the neural mechanisms underlying these behaviors is a central goal of social neuroscience. In recent years, rodent models have emerged as powerful tools for investigating these mechanisms, allowing cellular-resolution recordings and causal manipulations that are not feasible in humans and non-human primates (e.g., Bartal et al., 2011; Burkett et al., 2016). However, the inferential challenges inherent in linking neural activity to complex social-cognitive constructs are substantial, and the promise of these models depends critically on the rigor with which neural data are collected, analyzed, and interpreted.

Zhang et al. (2024) describe a mouse model of prosocial behavior in which an observer engages in targeted "allolicking" of a cagemate's injury site. In this model, one mouse (the "demonstrator") received an injection of melittin (a pain-inducing component of bee venom) into its hind paw; when reunited with a cagemate (the "observer"), the observer licked the demonstrator's injured paw. Zhang et al. consider this allolicking a helping behavior, distinct from general allogrooming, because it partially substitutes for the demonstrator's own self-licking, which is an instinctive response that promotes wound care and pain relief. Using calcium imaging in freely behaving observers, the authors recorded both single-neuron and ensemble activity in the anterior cingulate cortex (ACC) while observers interacted with demonstrators that were in different states (naive, stressed, or in pain), or while observers themselves experienced first-hand pain. Based on these recordings, Zhang et al. conclude that the ACC contains distinct neural representations for conspecifics' pain versus non-pain states (which, they argue, would allow the observer to detect when helping is needed) as well as shared neural representations for pain in the self and in others (which, they argue, generates the motivation to help by evoking a similar negative state in the observer). Further, they used optogenetic and chemogenetic manipulations to test the causal role of ACC activity in modulating allolicking behavior, finding that ACC activation increased allolicking, and ACC inhibition decreased it.

In this commentary, we raise five concerns about Zhang et al.'s (2024) conclusions. First, Zhang et al. apply inconsistent—and largely unarticulated—criteria when interpreting neural overlap (i.e., the degree to which the same

a Correspondence: giandomenico.iannetti@iit.it

b Correspondence: andrei.cimpian@nyu.edu

neurons respond to different stimuli): They treat 35% overlap as evidence of distinct representations while treating 8–13% overlap as evidence of shared representations, without explicating the rationale for these different thresholds. Second, Zhang et al.'s decoder analyses (i.e., machine-learning classifiers trained to distinguish neural activity across stimuli) yield a pattern opposite to what their representational claims would predict: The "shared" self/other comparison was decoded more accurately than the "distinct" pain/non-pain comparison. Third, Zhang et al.'s use of a fixed experimental order complicates the interpretation of their results: Because stimuli were always presented in the same sequence rather than counterbalanced, neural habituation effects may have artificially inflated the apparent distinctness of neural populations across comparisons. Fourth, we identify a mismatch between Zhang et al.'s region-wide optogenetic and chemogenetic manipulations and their population-specific mechanistic claims. Finally, we situate these issues within a broader methodological challenge in neuroscience: the difficulty of inferring functional specificity from limited experimental contrasts. Wherever feasible, we offer concrete suggestions for future research.

1. Inconsistent Criteria for Interpreting Neural Overlap

A central goal of systems neuroscience is to understand how neural activity represents the world. A common strategy is to compare which neurons are engaged by different experimental stimuli and to treat the degree of neural population overlap as informative about representational similarity: All else equal, substantial overlap is often taken to suggest that the brain treats the stimuli similarly (i.e., that they share representational components), whereas limited overlap is taken to suggest more distinct representations (e.g., Kriegeskorte et al., 2008). This inferential logic, linking the degree of overlap to shared versus distinct representations, underlies many claims about neural coding across domains.

When researchers observe that two experimental stimuli recruit partially overlapping neural populations, how should the degree of overlap be interpreted? The answer to this question is not self-evident: The same value can look large or small depending on the relevant baseline (e.g., how many neurons respond at all, how broadly tuned they are, and how much overlap would be expected by chance). Moreover, the appropriate benchmark may legitimately differ across comparisons (for example, when stimuli differ in perceptual similarity; Op de Beeck et al., 2008), so any claim about shared versus distinct representations demands an explicit rationale for the benchmark being applied. Without such a principled framework, ideally specified in advance and applied consistently, conclusions become difficult to evaluate and vulnerable to post hoc interpretation.

The results of Zhang et al. (2024) face precisely this interpretive challenge. We focus our critique on two components of the neural mechanisms they investigated. The first component consists of distinct neural representations of conspecifics' pain versus non-pain states: To guide helping behavior, the underlying neural mechanisms must be able to detect when another individual is in pain. The second component consists of shared neural representations of self-pain and other-pain: Neural mechanisms for helping behavior must generate the motivation to help, and Zhang et al. argue that this motivation arises from shared neural responses to self and others in pain: "the perception of others' pain can evoke a similar negative state in the observer" (p. 141), which the observer then seeks to alleviate through helping. We note that Zhang et al. investigated other components of these neural mechanisms as well, such as distinct neural representations of self- and other-pain: Without a robust self/other distinction, an animal would be unable to direct help toward the correct target. However, we focus on the first two components mentioned above (distinct representations of pain and non-pain in others and shared self-/other-pain representations) because they most clearly reveal the interpretive issues intrinsic to Zhang et al.'s argument.

As evidence for the first component of the putative neural mechanisms (namely, that the neural representations of conspecifics' pain and non-pain states are distinct), Zhang et al. (2024) described two populations of neurons in the ACC: some responding to others in pain and some responding to others in a naive state. These populations overlapped by 35.2% (174 shared out of 494 pain-responsive neurons; Fig. 3e). Zhang et al. interpreted this amount of overlap as evidence of distinct representations: "A substantial fraction of neurons in these two populations did not overlap [...], suggesting that the ACC responds differently at the single-cell level to animals in naive and pain states" (p. 139). As evidence for the second component of the putative neural mechanisms (namely, that the neural representations of pain in the self and in others are shared), the authors described two other populations of ACC neurons: some responding to pain in the self and some responding to pain in others. These populations overlapped by 8.7% (19 shared out of 219 neurons that responded to pain in others; Fig. 3o) or 13.7% (30 out of 219; Fig. 3q), depending on the type of noxious stimulus causing pain in others. Zhang et al. interpreted this amount of overlap as evidence of "shared encoding of other's pain versus self-pain" (Fig. 3, p. 140).

Our concern is that Zhang et al. (2024) interpret their overlap analyses using different—and largely unarticulated—standards. In one instance, a 35.2% overlap between the neuronal populations responsive to demonstrators in pain versus naive demonstrators is presented as evidence that these states have *distinct* representations. In another, markedly smaller overlaps (8.7% or 13.7%) between self-pain and other-pain conditions are presented as evidence of *shared* neural representation of these states. To be clear, neural representations can certainly exhibit both shared and distinct components. The issue is that Zhang et al. apply inconsistent interpretive standards to the observed neural overlaps. Without consistent criteria for interpreting percentages of neural overlap, any pattern of results could be marshaled to support predetermined conclusions.

One might argue, and rightly so, that the same percentage of neural overlap need not be interpreted identically across all comparisons. For instance, one of Zhang et al.'s (2024) comparisons involves two perceptually similar conditions (observing a conspecific in pain vs. in a naive state), while the other involves two perceptually dissimilar conditions (experiencing pain in oneself vs. observing a conspecific in pain). Due to this difference in perceptual similarity, more overlap might be expected in the former comparison simply because the stimuli are more alike (e.g., Op de Beeck et al., 2008). On this view, the threshold for interpreting overlap as evidence of shared pain representations might reasonably vary: Perhaps the 35.2% overlap in response to perceptually similar states (observing a conspecific in pain vs. in a naive state) might count as evidence of distinct representations, while the smaller overlap of 8.7% or 13.7% in response to perceptually dissimilar states (experiencing pain in oneself vs. observing a conspecific in pain) might suffice as evidence of shared representations. This interpretation, while in principle possible, is highly speculative and, crucially, goes well beyond the reasoning presented in the original paper; it reflects a line of argument that Zhang et al. themselves never articulate. Nowhere do the authors offer justification for applying different thresholds or state any general principles for interpreting neural overlap. Without such justification, it remains unclear what pattern of results would have led the authors to draw different conclusions.

How might future research avoid these interpretive ambiguities? We offer several concrete suggestions. First, researchers should specify—before examining overlap values—their inferential logic linking overlap metrics to representational claims, and preregister this framework in the same way they would preregister hypotheses or analysis plans. This need not require committing to a single rigid percentage threshold; rather, it entails stating what pattern(s) of evidence would justify describing two stimuli as having more shared versus more distinct representations, and under what assumptions (e.g., differences in baseline responsivity or perceptual similarity that might warrant different benchmarks across comparisons). Second, overlap should be interpreted relative to an explicit baseline rather than based on raw percentages alone. For example, researchers could quantify expected overlap under an appropriate null model—such as by shuffling stimulus labels within animal or session—and then ask whether observed overlap is meaningfully higher or lower than expected given the sizes of the responsive populations. Such baselines do not eliminate judgment, but they constrain it, making it harder to characterize overlap as "substantial" or "minimal" without a quantitative anchor. Together, these practices would strengthen the evidential value of overlap analyses and reduce the risk that claims about shared versus distinct representations are shaped by flexible, post hoc interpretation.

2. Decoder Performance and Representational Claims

A second source of evidence for Zhang et al.'s (2024) representational claims comes from decoder analyses. Whereas overlap analyses ask which neurons are engaged by different stimuli, decoder analyses ask whether patterns of population activity carry information that distinguishes those stimuli. The basic inferential logic is that if a classifier trained on neural activity can predict above chance (on held-out trials) which stimulus the neurons were responding to, then some aspect of the population response reliably differentiates the stimuli. Conversely, chance-level decoding suggests that there is insufficient reliable information to discriminate the stimuli. (Notably, decoding failures can also reflect limited power, noise, or modeling choices rather than true representational equivalence.)

Importantly, decoder performance is often used to support representational claims: If two stimuli are argued to have highly distinct representations, one would typically expect them to be more separable in population activity than stimuli argued to share a common representational structure, all else equal (e.g., Norman et al., 2006). This expectation is not a logical necessity, but it is part of the evidential logic by which decoder results are often interpreted in systems neuroscience (for a cogent critique of the physiological conclusions based on brain decoding results, see Vigotsky et al., 2024). Conversely, to the extent that "shared representation" is taken to mean that two stimuli evoke similar patterns in representational space, one would expect reduced separability, and thus more difficult decoding, relative to contrasts hypothesized to rely on distinct representations.

Applied to Zhang et al.'s (2024) argument, this logic suggests that decoding should be easier for distinguishing pain from non-pain in others (which they argue to have largely distinct representations based on neural overlap) than for distinguishing self- from other-pain (which they argue to have partly shared representations). However, Zhang et al. report successful decoding in both cases (Figs. 3h and 3p)—and, strikingly, the pattern is reversed. The decoder separated self- from other-pain (i.e., the comparison Zhang et al. describe as involving shared encoding; average AUROC = .921) *more accurately* than it separated pain from non-pain in others (i.e., the comparison they describe as involving distinct representations; average AUROC = .713). Rather than showing a dissociation in a direction consistent with the authors' conclusions from the analyses of neural overlap (namely, stronger decoding for the putatively distinct contrast), the decoder results contradicted the paper's claims. Moreover, the unusually high performance of the self–other decoder (AUROC = .921) warrants scrutiny; unfortunately, Zhang et al. provided only summary statistics (e.g., per-animal AUROC values), not the trial-level inputs or cross-validation specifications that would allow independent verification of the decoding pipeline.

What lessons for future research can be drawn from this concern? First, when decoder analyses are used alongside overlap analyses to support representational claims, the

two lines of evidence should be interpreted jointly: They should cohere, or, if they diverge, the tension should be explicitly acknowledged and addressed. More generally, researchers should state in advance what pattern of overlap and decoding results would count as supportive versus problematic for their representational interpretation. When results do not align with these expectations (for example, when a contrast previously interpreted as "shared" is at least as separable as a contrast previously interpreted as "distinct"), this warrants discussion and clarification of what notion of "sharing" is intended. Second, when decoder performance is very high (e.g., AUROC > .90), it becomes especially important to document and, ideally, enable independent verification of the decoding pipeline. This includes sharing trial-level inputs, specifying how cross-validation folds were constructed, and clarifying how independence was ensured between training and test sets.

3. Order Effects and Representational Claims

Valid inferences about representational overlap and distinctness require experimental designs in which the presentation of stimuli from different categories is not confounded with time in session. In Zhang et al.'s (2024) experiments, stimuli from different categories were presented in fixed sequences rather than counterbalanced orders. For example, in the pain versus non-pain comparison, the observers (i.e., the subjects) interacted with demonstrators first in a naive state and subsequently after melittin injection, which means that naive and pain exposures were presented in a consistent naive → pain order (Extended Data Fig. 8a,b). In the same experimental sequence, the observers were then subjected to pain themselves (Extended Data Fig. 8b). Thus, both the pain/non-pain contrast for others and the self/other contrast are potentially entangled with session progression.

This matters because neural population activity can change over time for reasons unrelated to the experimental manipulations of interest. One possibility is habituation or adaptation (e.g., Grill-Spector et al., 2006): Neurons that are genuinely responsive to two stimulus categories could still appear selective to one of the two categories if their responsivity attenuates across time and the two categories are presented in a fixed sequence. In this case, habituation would reduce neural responses to the stimuli of the second category presented, thereby creating the illusion of "distinct" populations. More generally, changes in arousal or fatigue can produce session-wide drift that alters overlap estimates and may also affect decoder performance. The core concern is not that any particular conclusion is necessarily wrong, but that fixed order makes it difficult to disentangle representational organization from order-dependent dynamics (e.g., Driscoll et al., 2022). Notably, this confound biases the interpretation in the same direction for both key comparisons: It would tend to make neural populations look more distinct than they truly are, complicating Zhang et al.'s (2024) claim to have identified shared self/other encoding alongside distinct pain/non-pain encoding.

Future work can mitigate this concern by decoupling stimuli from time. The most direct solution is to counterbalance order across animals or sessions (e.g., naive → pain vs. pain → naive; other → self vs. self → other), or to measure responses to stimuli in separate sessions under comparable baseline states. When counterbalancing is infeasible, researchers can still assess the influence of order effects by testing whether key results are robust to time-related controls—for example, by modeling time in session as a covariate, by comparing responses from early versus late portions within each stimulus exposure to assess within-stimulus drift, or by testing whether the key results hold when analyses are restricted to subsets of trials matched on other time-varying factors (e.g., arousal, movement). More generally, using manipulations that permit interleaving or randomization of stimuli would reduce serial dependencies and strengthen inferences about shared versus distinct representations.

4. Mismatch Between Population-Specific Claims and Region-Wide Manipulations

We applaud Zhang et al.'s (2024) use of state-of-the-art optogenetic and chemogenetic tools to modulate helping behavior. Yet, because these manipulations targeted the entire ACC, they do not establish a causal link between the specific neural populations highlighted in their investigation of neural mechanisms (e.g., those showing self/other overlap) and the observed helping behavior. Thus, independent of the methodological issues discussed above, this mismatch between region-wide experimental manipulations and population-specific mechanistic claims represents an additional limitation when evaluating Zhang et al.'s conclusions.

5. Inferring Functional Specificity from Limited Experimental Contrasts

A recurring interpretive challenge in systems and social neuroscience is distinguishing selectivity for a hypothesized psychological dimension (e.g., pain) from responses to other dimensions that covary with it (e.g., salience, arousal, negative valence). This issue has been discussed extensively in the literature—see, for example, Iannetti et al. (2013), Mouraux and Iannetti (2018), Pessoa (2023), Poldrack (2006), Rütgen and Lamm (2024), and Zaki et al. (2016). In the current case, simply observing that a neural population is active when an animal witnesses a conspecific's pain-like behavior and less active in a single control condition (here, witnessing a conspecific in a naive state) does not, by itself, establish that the population encodes pain rather than a correlated property that happens to often covary with pain, such as high salience or negative affect.

An analogy may help illustrate the inferential gap: If someone is wet on a rainy day and dry on a clear day, we cannot conclude that wetness "encodes" rain. Wetness can be caused by many other factors (e.g., swimming). Zhang et al. (2024) take a valuable step by including an additional comparison condition (witnessing stressed conspecifics), which helps rule out one plausible alternative interpretation. However, a single additional contrast cannot ad-

judicate among the many remaining possibilities. Ruling out stress, like ruling out swimming in the example above, still leaves open alternative explanations for the observed neural responses, including nonspecific tracking of salient, arousing, or emotionally intense events.

The claim of pain specificity would be strengthened by control conditions that explicitly disentangle pain from these broader dimensions while remaining feasible in mouse imaging studies of this kind. For example, adding a non-social, high-salience control (e.g., a loud noise or brief flashing light) would test whether the putatively "pain-responsive" neurons are instead tuned to startling events or abrupt sensory changes. Adding a positively valenced control (e.g., sucrose or a highly motivating social stimulus) would probe whether these neurons track general emotional intensity or motivational salience rather than specific nociceptive states. Reducing the demonstrator's pain pharmacologically (e.g., via local analgesia) while holding other aspects of the social interaction constant would provide a stronger dissociation between nociception-related signals and general arousal or salience.

This broader-contrast logic mirrors developments in human neuroimaging: Early interpretations of activity in medial regions such as the ACC and the insula as pain-specific (e.g., Eisenberger et al., 2003; Kross et al., 2011) were later reconsidered when richer stimulus sets showed robust responses to non-painful but surprising or salient events (Mouraux et al., 2011). We therefore view Zhang et al.'s (2024) findings as supporting a more limited conclusion: namely, that ACC activity differentiates among the states they tested in their paradigm. Stronger claims that the ACC selectively encodes pain (rather than tracking broader dimensions such as salience or negative affect) are not warranted, as they would require broader controls of the sort outlined above. To be clear, we view these suggestions as natural next steps in an incremental research program. Establishing phenomena and preliminary insights is valuable; our point is simply that the stronger interpretive claims would be better supported by the additional controls we have described.

Conclusion

Zhang et al. (2024) offer an innovative paradigm for studying prosocial behavior in mice and provide evidence that ACC activity is engaged during interactions with conspecifics in pain. Our goal in this commentary was not to dismiss their contribution, but to clarify methodological issues that complicate the paper's strongest claims—especially claims about distinct versus shared representations and about pain-specific encoding. We hope this commentary contributes to ongoing efforts to strengthen inferential standards in social and systems neuroscience.

Ethics Statement

This research did not involve testing of human or non-human participants.

Competing Interests

We have no competing interests to disclose.

Acknowledgments

We are grateful to two anonymous reviewers and our action editor, Dr. Helena Hartmann, for their constructive comments throughout the review process. We also thank Andrew Vigotsky and Rory Bufacchi for their insightful feedback on previous drafts of the manuscript.

Contributions

GI: Conceptualization (Equal), Methodology (Supporting), Writing – original draft (Equal), Writing – review & editing (Equal)

AC: Conceptualization (Equal), Methodology (Lead), Writing – original draft (Equal), Writing – review & editing (Equal)

Editors: Helena Hartmann (Associate Editor)

Submitted: December 16, 2025 PDT. Accepted: February 27, 2026 PDT. Published: April 20, 2026 PDT.

References

Bartal, I. B. A., Decety, J., & Mason, P. (2011). Empathy and pro-social behavior in rats. *Science*, *334*(6061), 1427–1430. https://doi.org/10.1126/science.1210789

Burkett, J. P., Andari, E., Johnson, Z. V., Curry, D. C., de Waal, F. B., & Young, L. J. (2016). Oxytocin-dependent consolation behavior in rodents. *Science*, *351*(6271), 375–378. https://doi.org/10.1126/science.aac4785

de Waal, F. B., & Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498–509. https://doi.org/10.1038/nrn.2017.72

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. https://doi.org/10.1177/1534582304267187

Driscoll, L. N., Duncker, L., & Harvey, C. D. (2022). Representational drift: Emerging theories for continual learning and experimental future directions. *Current Opinion in Neurobiology*, *76*, 102609. https://doi.org/10.1016/j.conb.2022.102609

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, *302*, 290–292. https://doi.org/10.1126/science.1089134

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. https://doi.org/10.1016/j.tics.2005.11.006

Iannetti, G. D., Salomons, T. V., Moayedi, M., Mouraux, A., & Davis, K. D. (2013). Beyond metaphor: contrasting mechanisms of social and physical pain. *Trends in Cognitive Sciences*, *17*, 371–378. https://doi.org/10.1016/j.tics.2013.06.002

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kross, E., Berman, M. G., Mischel, W., & Wager, T. D. (2011). Social rejection shares somatosensory representations with physical pain. *Proceedings of the National Academy of Sciences USA*, *108*, 6270–6275. https://doi.org/10.1073/pnas.1102693108

Mouraux, A., Diukova, A., Lee, M. C., Wise, R. G., & Iannetti, G. D. (2011). A multisensory investigation of the functional significance of the "pain matrix." *NeuroImage*, *54*, 2237–2249. https://doi.org/10.1016/j.neuroimage.2010.09.084

Mouraux, A., & Iannetti, G. D. (2018). The search for pain biomarkers in the human brain. *Brain*, *141*(12), 3290–3307. https://doi.org/10.1093/brain/awy281

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Op de Beeck, H. P., Torfs, K., & Wagemans, J. (2008). Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *Journal of Neuroscience*, *28*(40), 10111–10123. https://doi.org/10.1523/JNEUROSCI.2511-08.2008

Pessoa, L. (2023). *The Entangled Brain: How Perception, Cognition, and Emotion Are Woven Together*. MIT Press.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Rütgen, M., & Lamm, C. (2024). Dissecting shared pain representations to understand their behavioral and clinical relevance. *Neuroscience & Biobehavioral Reviews*, *163*, 105769. https://doi.org/10.1016/j.neubiorev.2024.105769

Vigotsky, A. D., Iannetti, G. D., & Apkarian, A. V. (2024). Mental state decoders: game-changers or wishful thinking? *Trends in Cognitive Sciences*, *28*, 884–895. https://doi.org/10.1016/j.tics.2024.06.004

Zaki, J., Wager, T. D., Singer, T., Keysers, C., & Gazzola, V. (2016). The anatomy of suffering: Understanding the relationship between nociceptive and empathic pain. *Trends in Cognitive Sciences*, *20*(4), 249–259. https://doi.org/10.1016/j.tics.2016.02.003

Zhang, M., Wu, Y. E., Jiang, M., & Hong, W. (2024). Cortical regulation of helping behaviour towards others in pain. *Nature*, *626*, 136–144. https://doi.org/10.1038/s41586-023-06973-x