Running head: NEUROIMAGING OF SAME AND DIFFERENT

Using model-based neuroimaging to adjudicate structured and continuous representational accounts in same-different categorization and beyond

Tyler Davis[1], Micah Goldwater[2]

[1] Department of Psychological Sciences, Texas Tech University, Lubbock, TX, 79403

[2] School of Psychology, University of Sydney, Sydney, New South Wales, Australia

Corresponding Author:

Tyler Davis
Department of Psychological Sciences
Texas Tech University
Box 42051 Lubbock, TX 79409
tyler.h.davis@ttu.edu

Abstract

The capacity to categorize using the concepts same and different plays a central role in cognition. However, in any given circumstance, it can be difficult to tell whether a person or animal is performing same and different categorization using structured relational rules from propositional logic or perceptual change detection processes. Discrete, logical behavior can often be produced from continuous perceptual spaces, and continuous behavior can arise from systems relying on structured logical rules. Model-based neuroimaging, when used in conjunction with advances in task development, can aid in understanding how people accomplish same and different categorization. We review recent publications employing the model-based neuroimaging paradigm to isolate latent brain processes associated with use of structured, logical rules and continuous perceptual processes.

Highlights

- Relational or continuous representations can be used to judge same and different
- Representational format is difficult to assess using behavior or neuroimaging alone
- Model-based neuroimaging can be used to adjudicate between representational formats
- The rostrolateral prefrontal cortex supports use of structured representations

Introduction

The concepts same and different provide a basis for higher-level human reasoning in logic, philosophy, and mathematics. Sameness allows us to substitute instantiations of concepts for one another to enable deduction and generalization through inductive reasoning. Difference gives us an idea of why such generalizations may fail. Sameness and difference also provide a key building block for cognition. Basic faculties like object permanence depend on some representation of sameness and difference to allow us to see objects as persisting through identity perserving transformations [1]. However, what representations of sameness and difference are, in a cognitive or neurobiological sense, remains a matter of debate with perhaps no single answer. In this review, we discuss cognitive and neurobiological representations of sameness and difference and how neuroimaging methods can inform research on how humans achieve this key feat.

Abstractly, sameness is a relational concept whereby individuals, objects, or events are the same if they are alike in all of their properties (or at least ones critical to maintaining identity); difference refers to cases in which some properties mismatch. Such a relational process may describe at least some of peoples' same-different categorization behavior [2]; people can, and sometimes likely do, represent sameness and difference by using a structured rule-like representation to check if all relevant properties align across two examples. However, in many cases, people likely do not use such rules. The alternatives are many variations on the idea that categorical behaviors, like identifying sameness and difference, can arise from basic properties of the continuous multidimensional spaces that underlie how our perceptual and conceptual representational systems are instantiated in the brain [3-5]; when noticing whether something is the same or different, we may not represent the abstract relation of sameness per se, but rather derive behavior consistent with representing these constructs from perceptions of change in more basic perceptual systems [6-8].

Main Text of Review

Perhaps surprisingly, there are often not foolproof ways of ensuring whether a person is representing sameness, relationally or continuously, using behavior alone. Continuous representations of sameness are sufficient to guide behavior in many "match-to-sample" tasks where a target stimulus is chosen based on its sharing of the sameness relation with another cue stimulus [9,10]. This is because behavior (choosing correctly) can appear rule-based simply by adopting very strict perceptual criteria on which stimuli to choose. Another common paradigm for studying representations of same and different from the comparative cognition literature involves learning to categorize visual arrays based on whether all elements in the array are the same or different (e.g., arrays with all matching/same stimuli are in category 1; arrays with all different stimuli are in category 2)[11-13]. After learning this categorization rule, participants then classify new examples with intermediate numbers of same and different elements. In humans, who unequivocally have relational concepts of same and different (whether or not they use them in a given circumstance), the modal behavior is to only respond with the same category (e.g., category 1) for instances of all same and respond different for any instances where there is any difference [14]. Many other species, for whom true relational concepts of same and difference are more in question, tend to respond more continuously [12,13]; for each increase in the number of mismatching elements (i.e., the entropy or disorderliness of the array), continuous responders become more likely to choose the different category. Although such comparative differences have been used to argue for a more rule or relational representation of sameness in humans (and some primates [14]) and only perceptual representations of orderliness or entropy in other animals, both behavioral profiles could theoretically arise from the same basic perception of entropy or orderliness. Criterion shifts in humans (relative to animals) toward only the most orderly arrays could produce a similar pattern from a single entropy dimension. Moreover, given some proportion of human participants behave more continuously, like animals, in array learning tasks [14], there remains a possibility

that such tasks index individual differences in receiver operating characteristics (ROCs) as opposed to differences in representation.

On the other hand, continuous/non-binary behaviors may also arise in cases where people are using structured rule representations, and not just when representations are truly continuous. For example, relational concepts like “predator” and “prey,” have structured rule-like representations wherein an animal either eats other animals or only plants. However, through experience with a variety of “predators” and “prey,” we might associate features like large sharp teeth with “predators” even though teeth are not strictly part of the relational rule. Further, even though such relational categories are not represented as binary-rules, their representations describing the systematic relations among objects (such as “hunt” and “eat”) are very different from those of categories typically modeled with continuous feature-space representations [15,16]. Likewise, when considering “sameness” in the context of substitutability and generalizability in deductive and inductive reasoning (respectively), this inherently involves seeing a functional “sameness” as also revealing continuous levels of difference. For example, analogical transfer during problem-solving involves both recognizing a structural equivalence, but also recognizing key differences [17,18]. Thus, neither continuous nor categorical/binary behavioral response patterns alone are necessary or sufficient for determining the representational format (structured or continuous) people or animals are using in a task.

Given the difficulty of determining representational format of sameness and difference concepts from behavior alone, we argue it is useful to consider how neuroimaging methods may be leveraged to complement behavioral data. A variety of neuroimaging analysis techniques can be used to study the nature of brain representations [19,20], including classic univariate “BOLD” activation techniques, multivoxel/machine learning techniques, and adaptation techniques. Like in behavioral studies, match-to-sample tasks, including n-back tasks, continue to be a mainstay of representational research in many areas of neuroimaging, but are often not interpreted as studies of how people make same and different judgments. In fact, despite the importance of

same and difference to many neuroimaging paradigms, there has been very little neuroimaging research on the question of how same and different are represented per se.

Even though the topic of same-difference categorization has received little attention in neuroimaging research, sameness and difference play central roles in the design and interpretation of many representational analysis techniques. However, as with behavioral measures, it is often difficult to assess whether representational neuroimaging techniques uncover binary, relational representations of sameness, or whether their results refect by-products of more continuous neuronal or regional representations. For example, adaptation techniques were developed from the observation in the single-cell recording literature [21] that repetition of a stimulus tends to lead to decreases in neural firing for the second presentation relative to the first (i.e., repetition suppression [22]). Although such findings show that neurons (or voxels or regions) are sensitive to the sameness relation, like behavior, they are perfectly consistent with more continuous perceptual representational systems. Indeed, while the largest magnitude adaptation effects typically happen for same stimuli (exact repetitions), neurons/voxels can exhibit continuous changes in adaptation as a function of the distance between stimuli in a perceptual space [23]. Further, even in cases where adaptation is observed only for exact stimulus matches, these are more likely to reflect greater specificity of tuning functions, or differences in neural topography of a perceptual feature space (e.g., local vs distributed) as opposed to differences in representational format (structured vs continuous)[24]. Beyond adaptation, other representational neuroimaging techniques focused primarily on the specificity of the brain's reponse to a stimulus would be met with similar interpretational pitfalls because specificity can nearly always be accomodated in continuous feature spaces with changes in ROCs.

What is needed to move forward in studying the difference between structured representational accounts and continuous perceptual accounts of same/different -- or any cases where structured "rule" representations are pitted against more continuous similarity or

perceptual representations -- is to move toward computational model-based neuroimaging frameworks, in conjunction with behavioral paradigm development. In model-based imaging, quantitative predictions are generated from a formal mathematical model of a candidate cognitive process or a set of competing hypotheses [25]. These quantitative predictions can be anything from the choice probabilities (e.g., decisional uncertainty) underlying a decision, or can include intermediate steps between perceptual inputs and behavior that constitute more “latent” psychological processes leading up to a decision [26]. To the degree that models using structured representations predict differences in such latent processes from those relying on more continuous similarity-based feature spaces, brain data can potentially be used to adjudicate between different representational formats.

One recent study on same and difference representation used a model-based imaging approach to test whether participants learned structured or continuous representations in an array categorization like those described above [27] (Figure 1). During training, participants learned to sort arrays of sine wave gratings (‘Gabor patches’) into novel categories (A or B) using trial and error. The category rule was based on whether all gratings in an array had the same or different orientations. Later, during a generalization phase, participants were tested with new arrays with different numbers of same and different gratings. Model-based fMRI analyses tested how two latent processes related to same-different categorization mapped onto the brain (Figure 1). One measure was based on the entropy of the arrays, a continuous representation of same-different. A second measure was based on a relational matching algorithm, instantiated in a hybrid exemplar model (relational generalized context model; rel-GCM) that allowed for categorization to be based on a structured, relational processes. The entropy measure was associated with activation in a number of prefrontal regions that tend to track uncertainty in decision making [28-32] including the lateral PFC and vmPFC. In contrast, the relational matching measure uniquely captured activation in the rostrolateral PFC (rlPFC).

The results from this study are important for several reasons. First, by showing that a signal related to relational matching correlates with brain activation during same-different learning, this modeling work provides some of the first non-behavioral evidence that relations are being used in array categorization judgments. Relational matching was a latent, model-based measure, and was not itself directly correlated with behavior. Thus, these results are potentially more immune to being re-explained by differences in ROCs. Indeed, the rlPFC region that was associated with our relational matching measure is known to be involved in relational reasoning [33] and other cognitive processes, like analogy, that depend upon more abstract or episodic cognitive control [34-36]. Interestingly, other PFC regions tracked the continuous entropy-based representation of same-different. The co-existence of relational and continuous representations is consistent with behavioral and computational research on relational thinking [37-41].

However, a further complexity is that entropy, in this task, was highly correlated with a decisional uncertainty measure from the rel-GCM based on relational matching. Thus, there are at least two possibilities to explain the apparent co-existence of relational and continuous representations of same-different in the PFC. First, it is possible that some of the brain regions associated with the entropy measure may reflect the decisional outputs of a more discrete, relational decision process as opposed to a continuous representation of same-different per se. Second, it is possible this apparent continuous representation is truly driven by a bottom-up perceptual signal, consistent with how entropy is typically conceived. Future research may develop model-based connectivity approaches to test whether the observed continuous same-different signals in the brain are being driven by relational matching processes in rlPFC or by more perceptual processing regions.

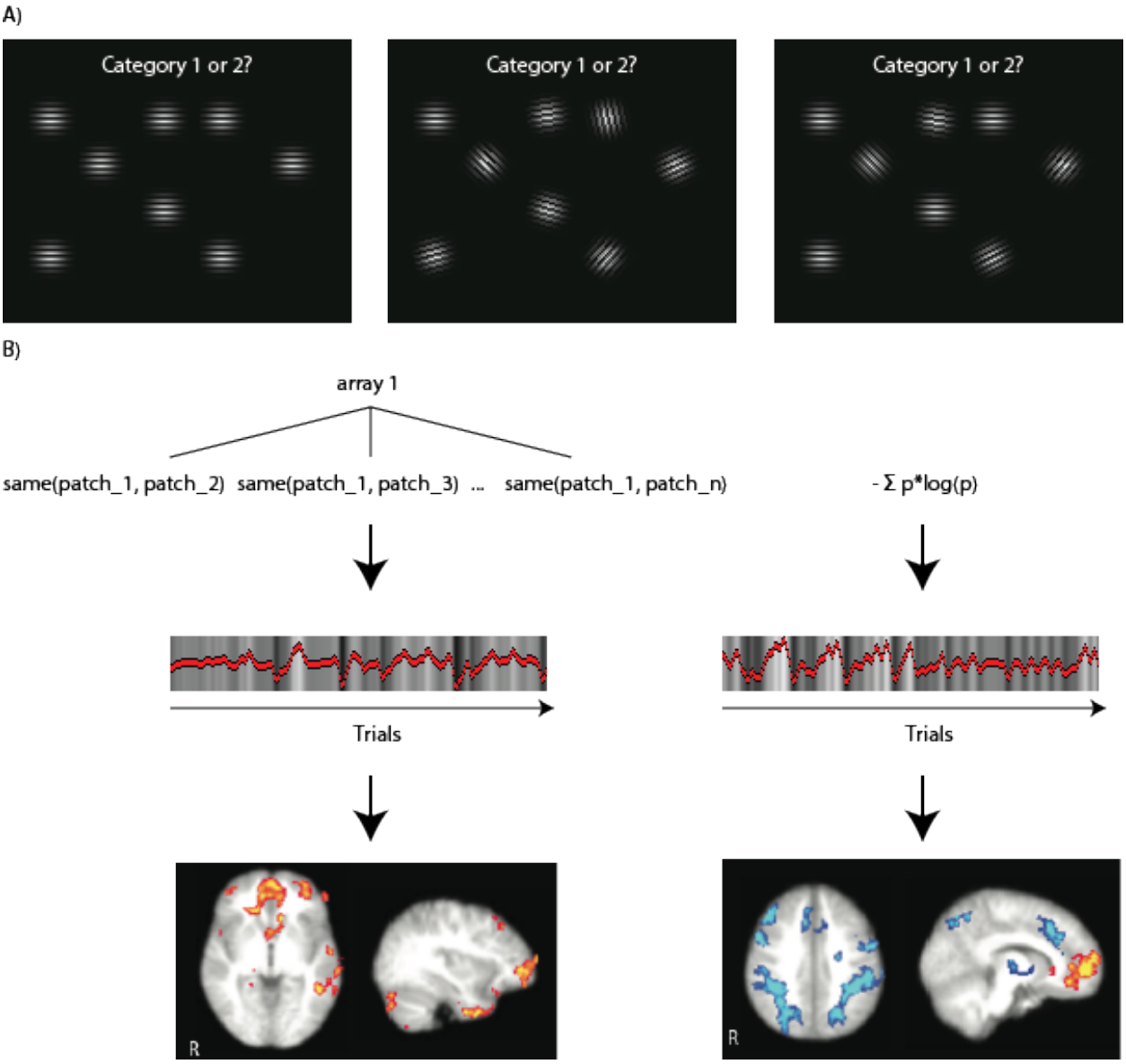

Figure 1. Depiction of the study design and results from Davis, Goldwater, and Giron [27]. A) Examples of the category learning task with a same array (left), different array (middle) and an array with intermediate number of same and different elements (right). B) A depiction of the model-based imaging framework and results. Predicted hemodynamic response functions are generated from models assuming structured relations (left) or continuous representation of entropy (right). These are fitted to brain activation using linear regression, yielding brain regions that differentially track relational processes or more continuous decision processes related to display entropy.

Although not on representation of sameness and difference per se, recent category learning research has extended the model-based imaging framework to adjudicate between

structured and continuous representations in other tasks [42]. O'Bryan, et al., [43] examined an 'inverse base-rate' task where participants learned to categorize fictitious rare and common diseases based on visual or semantic cues. The past findings on this task are, when asked to categorize an ambiguous stimulus that has features of both rare and common diseases, people tend to choose the rare disease more than one would expect based on their base rates [44]. Like same-different categorization, hypotheses for why people do this have included structured representational accounts, such as the use of eliminative rules [45], and continuous perceptual or associative accounts, such as greater attention to rare features [46]. Using model-based fMRI measures that included predictors for dissimilarity- (eliminative) and similarity-based processes, we found that rlPFC uniquely tracked dissimilarity-based processing, consistent with a structured representation account. Together with the previous study [27], these results suggest rlPFC supports the use of structured representations in category learning and generalization.

Importantly, however, rlPFC activation alone should not be interpreted as a brain signature of structured rule use. Although rlPFC tends to track cognitive control demands in abstract or episodic control tasks where structured rules are used, it is also often activated in tasks less clearly related to rule use, such as for exploratory decision making [36](but see [47]). Further, there are cases where rule use can be inferred based on behavior, such as in the Shanks-Darby patterning task [48], but where strong neuroimaging dissociations between people putatitively using such structured rules and those using more continuous or associative strategies have not been forthcoming [49].

One of the reasons why rlPFC may not show activation differences during apparent rule use is that people tend to automate rule use and use associations stored in memory as a task becomes more familiar [50,51]. For example, rlPFC tends to be activated when a problem solving procedure needs to be generalized to a novel situation, but not when they can be applied in a familiar manner [52]. Likewise, in category learning, for simple match-to-sample and

classification rules, rlPFC activation may be high during initial rule discovery and use, but then disipate as uncertainty decreases [53].

A key question for future research concerns the behavioral consequences of moving from the use of cognitive control processes supported by the lateral PFC to the use of more automated stategies. In animal models of habit learning [54], related shifts can lead to behavioral rigidity. However, in humans, evidence suggests that increases in expertise can have the opposite effects. For example, in jazz musicians, compared to novices, the shift away from lateral PFC control mechanisms may lead to greater flexibility and creativity [55]. In relational tasks, experts tend to recognize common relational structures across different examples more readily than novices [56,57], which may reflect use of episodic memory strategies. Even in brief learning tasks, participants can develop strategies that allow them to accurately classify according to relational rules without fully engaging relational matching processes [58].

Conclusions

To solve the puzzle of how humans represent structured relations, such as same and different, we expect increased use of model-based neuroimaging frameworks will be needed. Where behavioral studies and representational neuroimaging techniques have frequently had difficulty in firmly dissociating continuous and structured representational accounts, model-based imaging can leverage differences in how such algorithms are instantiated computationally to uncover “latent” brain states that are more uniquely associated with a specific type of representation. This approach has had success in both same/different and other categorization tasks where structured representational and continuous feature-based accounts have previously reached an impasse.

Funding: This work was supported by National Science Foundation grant #1923267 to T.D.

References

1. Bremner JG, Slater AM, Johnson SP: **Perception of object persistence: The origins of object permanence in infancy.** *Child Dev Perspect* 2015, **9**:7-13.

2. Wasserman E, Castro L, Fagot J: **Relational thinking in animals and humans: From percepts to concepts.** In *APA handbooks in psychology®. APA handbook of comparative psychology: Perception, learning, and cognition.* Edited by Call J, Burghardt GM, Pepperberg IM, Snowdon CT, Zentall T. APA Press; 2017:359–384.

3. Gärdenfors P: *Conceptual spaces: The geometry of thought*. MIT press; 2004.

4. Guntupalli JS, Hanke M, Halchenko YO, Connolly AC, Ramadge PJ, Haxby JV: **A model of representational spaces in human cortex.** *Cereb Cortex* 2016, **26**:2919-2934

5. Kriegeskorte N, Diedrichsen J: **Peeling the Onion of Brain Representations.** *Annu Rev Neurosci* 2019, 42:407-432.

6. Farell B: **" Same"–"different" judgments: A review of current controversies in perceptual comparisons.** *Psychol Bull* 1985, **98**:419-456.

7. Thompson RK, Oden DL: **A profound disparity revisited: Perception and judgment of abstract identity relations by chimpanzees, human infants, and monkeys.** *Behav Processes* 1995, **35**:149-161.

8. Addyman C, Mareschal D: **The perceptual origins of the abstract same/different concept in human infants.** *Anim Cogn* 2010, **13**:817-833.

9. Hochmann JR, Tuerk AS, Sanborn S, Zhu R, Long R, Dempster M, Carey S: **Children's representation of abstract relations in relational/array match-to-sample tasks.** *Cogn Psychol* 2017, **99**:17-43.

10. Walker CM, Gopnik A: **Discriminating relational and perceptual judgments: Evidence from human toddlers.** *Cognition* 2017, **166**:23-27.

11. Wasserman EA, Hugart JA, Kirkpatrick-Steger K: **Pigeons show same-different conceptualization after training with complex visual stimuli.** *J Exp Psychol Anim Behav Process* 1995, **21**:248-252.

12. Young ME, Wasserman EA: **Entropy detection by pigeons: Response to mixed visual displays after same–different discrimination training.** *J Exp Psychol Anim Behav Process* 1997, 23:157-170.

l3. Wasserman EA, Young ME: **Same–different discrimination: The keel and backbone of thought and reasoning.** *J Exp Psychol Anim Behav Process* 2010, **36**:3-22.

14. Fagot J, Wasserman EA, Young ME: **Discriminating the relation between relations: the role of entropy in abstract conceptualization by baboons (Papio papio) and humans (Homo sapiens).** *J Exp Psychol Anim Behav Process* 2001, **27**:316-328.

15. Goldwater MB, Schalk L: **Relational categories as a bridge between cognitive and educational research.** *Psychol Bull* 2016, **142**:729–757.

16. Gentner D, Asmuth J. **Metaphoric extension, relational categories, and abstraction.** *Lang Cogn Neurosci* 2019, **34**:1298-1307.

17. Cushen PJ, Wiley J: **Both attentional control and the ability to make remote associations aid spontaneous analogical transfer.**
Mem Cognit 2018, **46**:1398-1412.

18. Kurtz KJ, Honke G. **Sorting out the problem of inert knowledge: Category construction to promote spontaneous transfer. *J Exp Psychol Learn Mem Cogn*, 2019. https://doi.org/10.1037/xlm0000750
This paper argues that having learners construct categories of natural phenomena explained by shared scientific principles helps them recognise their underlying sameness, and then transfer this understanding to further disparate contexts.

19. Davis T, Poldrack RA: **Measuring neural representations with fMRI: practices and pitfalls.** *Ann N Y Acad Sci* 2013, **1296**:108-134.

20. Kriegeskorte N, Diedrichsen J:**Peeling the Onion of Brain Representations**, 2019. *Annu Rev Neurosci*, **42**:407-432.

21. Miller EK, Li L, Desimone R. **A neural mechanism for working and recognition memory in inferior temporal cortex.** *Science* 1991, **254**:1377–1379.

22. Barron HC, Garvert MM, Behrens TE: **Repetition suppression: a means to index neural representations using BOLD?** *Philos Trans R Soc Lond B Biol Sci* 2016, **371**:20150355.

23. Aguirre GK. **Continuous carry-over designs for fMRI.** *Neuroimage* 2007, **35**:1480–1494.

24. Drucker DM, Aguirre GK: **Different spatial scales of shape similarity representation in lateral and ventral LOC.** *Cereb Cortex* 2009, **19**:2269-2280.

25. Palmeri TJ, Love BC, Turner BM: **Model-based cognitive neuroscience.** *J Math Psychol* 2017, **76**:59-64

26. Mack ML, Preston AR, Love BC: **Ventromedial prefrontal cortex compression during concept learning. *Nat Commun* 2020, **11**:1-11.
This study used model-based predictions from the category learning model SUSTAIN to test the hypothesis that the ventromedial prefrontal cortex (vmPFC) engages dimensionality reduction mechanisms to filter out irrelevant information during learning. Consistent with this hypothesis, the results revealed that the dimensionality of vmPFC activation patterns was associated with categorization problem complexity. Further, individual differences in participants' abilities to filter out irrelevant information (as indexed by attention weights from SUSTAIN) correlated with the amount of dimensionality reduction observed in vmPFC activation patterns.

27. Davis T, Goldwater M, Giron J: **From concrete examples to abstract relations: The rostrolateral prefrontal cortex integrates novel examples into relational categories.** *Cereb Cortex* 2017, **27**:2652-2670.

28. Lebreton M, Abitbol R, Daunizeau J, Pessiglione M: **Automatic integration of confidence in the brain valuation signal.** *Nat Neurosci* 2015, **18**:1159-1167.

29. Gherman S, Philiastides MG: **Human VMPFC encodes early signatures of confidence in perceptual decisions.** *Elife* 2018, **7**:e38293.

30. Bang D, Fleming SM: **Distinct encoding of decision confidence in human medial prefrontal cortex.** *Proc Natl Acad Sci U S A* 2018, **115**:6082-6087.

31. Davis T, LaCour M, Beyer E, Finck JL, Miller,MF: **Neural correlates of attitudes and risk perception for food technology topics.** *Food Qual Prefer* 2020, **80**:103836.

32. Bobadilla-Suarez S, Guest O, Love BC. **The neural link between subjective value and decision entropy.** *bioRxiv,* 2020. https://doi.org/10.1101/2020.02.18.954362

*33. Hartogsveld B, Bramson B, Vijayakumar S, van Campen AD, Marques JP, Roelofs K, Toni I, Bekkering H, Mars RB. **Lateral frontal pole and relational processing: activation patterns and connectivity profile.** *Behav Brain Res* 2018, **355:**2-11.
This study uses multiple methods to show that relational processing, independent of stimulus type (faces or geometric shapes) engages an anatomically distinct rostrolateral prefrontal cortex region. Both anatomical and functional connectivity of this rostrolateral prefrontal cortex region are distinct from neighboring caudal or medial areas of the frontal pole and lateral prefrontal cortex.

34. Westphal AJ, Reggente N, Ito KL, Rissman J: **Shared and distinct contributions of rostrolateral prefrontal cortex to analogical reasoning and episodic memory retrieval.** *Hum Brain Mapp* 2016, **37**:896-912.

35. Wendelken C, Ferrer E, Ghetti S, Bailey SK, Cutting L, Bunge SA: **Frontoparietal structural connectivity in childhood predicts development of functional connectivity and reasoning ability: A large-scale longitudinal investigation.** *J Neurosci* 2017, **37**:8549-8558.

36. Badre D, Nee DE: **Frontal cortex and the hierarchical control of behavior.** *Trends Cogn Sci* 2018, **22**:170-188.

37. Forbus KD, Gentner D, Law K: MAC/FAC: **A model of similarity-based retrieval.** *Cogn Sci* 1995, **19**:141-205.

38. Doumas LA, Hummel JE, Sandhofer CM: **A theory of the discovery and predication of relational concepts.** *Psychol Rev* 2008, 115:1-43.

39. Goldwater MB, Markman AB, Stilwell CH: **The empirical case for role-governed categories.** *Cognition* 2011, **118**:359-376.

40. Petkov G, Petrova Y: **Relation-based categorization and category learning as a result from structural alignment. The RoleMap model.** *Front Psychol* 2019, **10**:563.

*41. Goldwater MB, Don HJ, Krusche MJF, Livesey EJ: **Relational discovery in category learning.** *J Exp Psychol Gen* 2018, **147**: 1–35.
In this series of category learning experiments, participants could either learn to categorize stimuli via perceptual features, or structured spatial relations. The paper examined what individual differences in learners, and which task structure variables pushed people towards a feature-based or relational strategy.

42. Zeithamova D, Mack ML, Braunlich K, Davis T, Seger CA, van Kesteren MT, Wutz A: **Brain Mechanisms of Concept Learning.** *J Neurosci* 2019, **39**:8259-8266.

41. O'Bryan SR, Worthy DA, Livesey EJ, Davis T: **Model-based fMRI reveals dissimilarity processes underlying base rate neglect.** *eLife* 2018, 7:e36395.

44. Medin DL, Edelson SM: **Problem structure and the use of base-rate information from experience.** *J Exp Psychol Gen* 1988, **117**:68 –85

45. Juslin P, Wennerholm P, Winman A: **High-level reasoning and base-rate use: Do we need cue-competition to explain the inverse base-rate effect?** *J Exp Psychol Learn Mem Cogn* 2001, **27**:849.

46. Kruschke JK: **Base rates in category learning.** *J Exp Psychol Learn Mem Cogn* 1996, **22**:3-26.

*47. Blanchard TC, Gershman SJ: **Pure correlates of exploration and exploitation in the human brain.** *Cogn Affect Behav Neurosci* 2018, **18**:117-126.

This study is the first reinforcement learning study to truly separate out exploration and exploitation processes into distinct trials by allowing participants to observe (pure exploration) or choose (pure exploitation) without immediately observing the results. Contrary to previous work in reinforcement learning, rostrolateral prefrontal cortex did not track exploration. Instead, the results revealed that vmPFC activation tracked model-based predictions for evidence accumulation mechanisms, and anterior cingulate and insula were more activated for exploratory responses.

48. Shanks DR, Darby RJ: **Feature-and rule-based generalization in human associative learning.** *J Exp Psychol Anim Behav Process* 1998, **24**:405-415.

49. Milton F, Bealing P, Carpenter KL, Bennattayallah A, Wills AJ: **The neural correlates of similarity-and rule-based generalization.** *J Cogn Neurosci* 2017, **29**:150-166.

50. Roeder J., Ashby FG: **What is automatized during perceptual categorization?** *Cognition* 2016, **154**:22-33.

51. Soto FA, Bassett DS, Ashby FG: **Dissociable changes in functional network topology underlie early category learning and development of automaticity.** *NeuroImage* 2016, **141**:220-241.

52. Anderson JR, Fincham JM: **Extending problem-solving procedures through reflection.** *Cogn Psychol* 2014, **74**:1-34.

53. Paniukov D, Davis T: **The evaluative role of rostrolateral prefrontal cortex in rule-based category learning.** *NeuroImage* 2018, **166**:19-31.

54. Balleine BW, Dickinson A: **Goal-directed instrumental action: contingency and incentive learning and their cortical substrates.** *Neuropharmacology* **1998**, 37:407-419.

55. Rosen DS, Erickson B, Kim YE, Mirman D, Hamilton RH, Kounios J. **Anodal tDCS to right dorsolateral prefrontal cortex facilitates performance for novice jazz improvisers but hinders experts.** *Front Hum Neurosci* 2016, **10**:579.

56. Chi MT, Feltovich PJ, Glaser R: **Categorization and representation of physics problems by experts and novices.** *Cogn Sci* 1981, **5**:121-152.

56. Rottman BM, Gentner D, Goldwater MB: **Causal systems categories: Differences in novice and expert categorization of causal phenomena.** *Cogn Sci* 2012, **36**:919-932.

*58. Corral D, Kurtz KJ, Jones M. **Learning relational concepts from within- versus between-category comparisons.** *J Exp Psychol Gen* 2018*,* **147**:1571–1596

This paper argues that often people can learn relational categories without engaging in full structural comparisons between stimuli, and instead adopt representational short-cuts that appear to mimic feature-based learning processes.