

Social-Structure Learning

Samuel J. Gershman  and Mina Cikara 

Department of Psychology, Harvard University

Current Directions in Psychological
 Science
 2020, Vol. 29(5) 460–466
 © The Author(s) 2020
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/0963721420924481
www.psychologicalscience.org/CDPS



Abstract

Social-structure learning is the process by which social groups are identified on the basis of experience. Building on models of structure learning in other domains, we formalize this problem within a Bayesian framework. According to this framework, the probabilistic assignment of individuals to groups is computed by combining information about individuals with prior beliefs about group structure. Experiments with adults and children provide support for this framework, ruling out alternative accounts based on dyadic similarity. More broadly, we highlight the implications of social-structure learning for intergroup cognition, stereotype updating, and coalition formation.

Keywords

Bayesian models, social psychology, structure learning

Social categories are woven tightly into the fabric of our social lives, shaping how we perceive, punish, cooperate with, and learn from other individuals. But where do these categories come from? This is the question addressed by the emerging area of research on *social-structure learning*, which leverages ideas from computational-cognitive science that have been applied to nonsocial domains (see Austerweil, Gershman, Tenenbaum, & Griffiths, 2015). The key idea is that the brain uses statistical-learning algorithms to sort individuals into *latent groups* on the basis of their behavioral patterns, such as choices (and possibly other features). Intuitively, individuals who behave similarly will tend to be grouped together. This simple principle has broad implications, which we will explore in this review.

We begin by reviewing a computational framework that formalizes social-structure learning and then turn to experimental tests of the framework, including recent developmental and neuroimaging experiments. We conclude with a discussion of how social-structure learning may provide insight into aspects of intergroup cognition, stereotype updating, and coalition formation.

Computational Framework

As schematized in Figure 1, social-structure learning addresses the following problem: Given observed behavioral patterns (e.g., choices between movies) for a set of individuals, the observer must infer the latent group assignment for each individual. The normative

solution to this inference problem (for more details, see Gershman, Pouncy, & Gweon, 2017) is given by Bayes's rule, which stipulates that the *posterior probability* over groupings given choices— $P(\text{grouping} | \text{choices})$ —is proportional to the product of the *likelihood*— $P(\text{choices} | \text{grouping})$ —and the *prior probability*— $P(\text{grouping})$. The posterior represents the observer's subjective confidence in each hypothetical grouping. The likelihood represents the match between a hypothetical grouping and the choices, and the prior represents a preference for particular groupings before the data are observed.

To define the likelihood, we need to specify how a grouping gives rise to choices. A basic assumption of this framework is that individuals assigned to the same group will tend to behave similarly (i.e., make similar choices). Thus, groupings with greater within-group homogeneity will have higher likelihood. This can, however, produce many small but homogenous groups, a tendency that can be tempered by enforcing a preference for a small number of groups via the prior. In particular, we have used a prior from the literature on nonparametric statistics known as the *Chinese-restaurant process* (CRP; for an introduction, see Gershman & Blei,

Corresponding Author:

Samuel J. Gershman, Harvard University, Department of Psychology, Northwest Building, 295.05 Oxford St., Room 295.05, Cambridge, MA 02138

E-mail: gershman@fas.harvard.edu

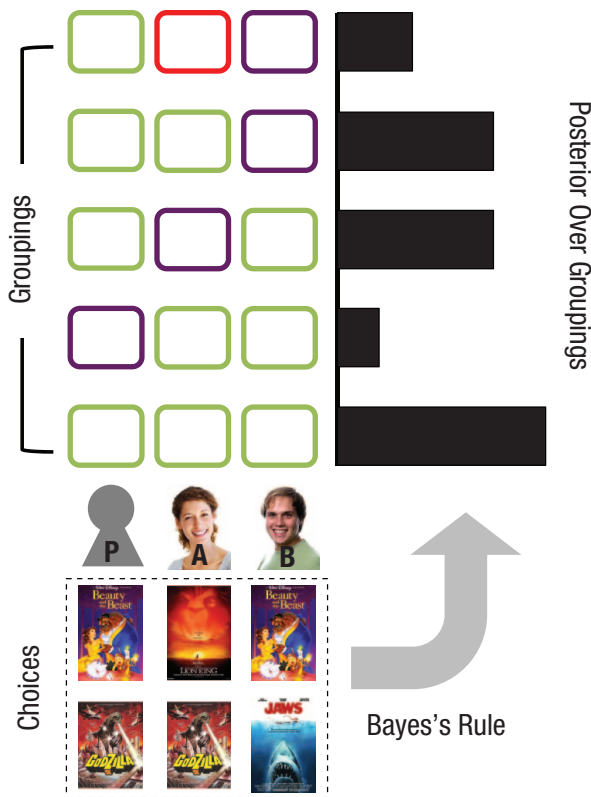


Fig. 1. Model schematic. The input to the model is the set of choices for each individual (including the experimental participant, labeled “P,” and the two other individuals, labeled “A” and “B”). In this example, individuals are choosing between different movies. Each row represents a different choice set. For example, the first row shows a choice between *Beauty and the Beast* and *The Lion King*. Individuals P and B chose *Beauty and the Beast*, whereas individual A chose *The Lion King*. Bayes’s rule is used to combine prior beliefs about groupings with the choice data to form a posterior probability distribution over groupings, indicated by the color-coded rectangles. Colors indicate group membership. Bar length indicates posterior probability.

2012). The analogy refers to a Chinese restaurant with effectively infinite seating capacity in which each individual (a single choice in our setup) is assigned to a table (group) with probability proportional to the number of individuals already seated at the table or to an unoccupied table with some probability determined by a *concentration parameter*. This prior has the property that it favors a small number of latent groups but allows for a possibly unbounded number of groups so that new groups can be added as new individuals are observed. The degree to which a small number of groups is preferred by the prior is controlled by the concentration parameter. Below, we discuss how variation in this parameter may offer insight into the development of social-structure learning.

This model is essentially an adaptation to social domains of Bayesian structure-learning models developed for nonsocial domains, notably categorization and classical conditioning (e.g., Anderson, 1991; Gershman,

Blei, & Niv, 2010; Sanborn, Griffiths, & Navarro, 2010). It is also conceptually related to some non-Bayesian ideas about category discovery, such as the *supervised and unsupervised stratified adaptive incremental network* (SUSTAIN) model (Love, Medin, & Gureckis, 2004), which uses a neural network to adaptively construct clusters of examples. One advantage of the Bayesian framework is that it makes explicit an individual’s assumptions about the environment, which can sometimes be used to ecologically constrain the prior. For example, data about real-world social networks could be used to build more ecologically realistic priors. A second advantage of the Bayesian framework is that it formalizes subjective uncertainty about groups, which provides a principled way of modeling confidence judgments, adaptive learning rates, and decisions under uncertainty.

Although we focus here on structure *learning*, the model can be applied to structure *inference* for well-learned groups (e.g., those based on race, age, or gender). These groups will tend to be frequently encountered and hence have high prior probability under the CRP. This potentially explains why we rely on these groups even when more fine-grained groupings might be warranted by the data—as in the case of the “multiple-category problem.” For example, people who have more overlapping stereotypes for *Black* and *male* categories take longer to accurately categorize Black female faces (Freeman & Johnson, 2016).

Experimental Tests

To test the predictions of the structure-learning model, we have used several variations of a canonical experimental design, schematized in Figure 2. In the first part of each block, a participant (labeled “P”) learns about the preferences of other individuals (labeled “A” and “B”) and also report their own preferences. Our original studies used movie choices (Gershman et al., 2017), but we have found similar results using political-issue positions (Lau, Pouncy, Gershman, & Cikara, 2018). At the end of each block, participants are given a “mystery-choice” trial in which they observe that two individuals (A and B) have made divergent choices. Critically, the movie identity or political issue is concealed, which makes this a pure test of social influence or ally choice. Participants are selecting with whom they want to align.

The structure-learning model is usefully contrasted with a simpler alternative based on dyadic similarity, according to which social influence depends only on the choice overlap between pairs of individuals. In the canonical experimental design, the stimuli were constructed so that individuals A and B had equal choice overlap with P. This means that if participants were

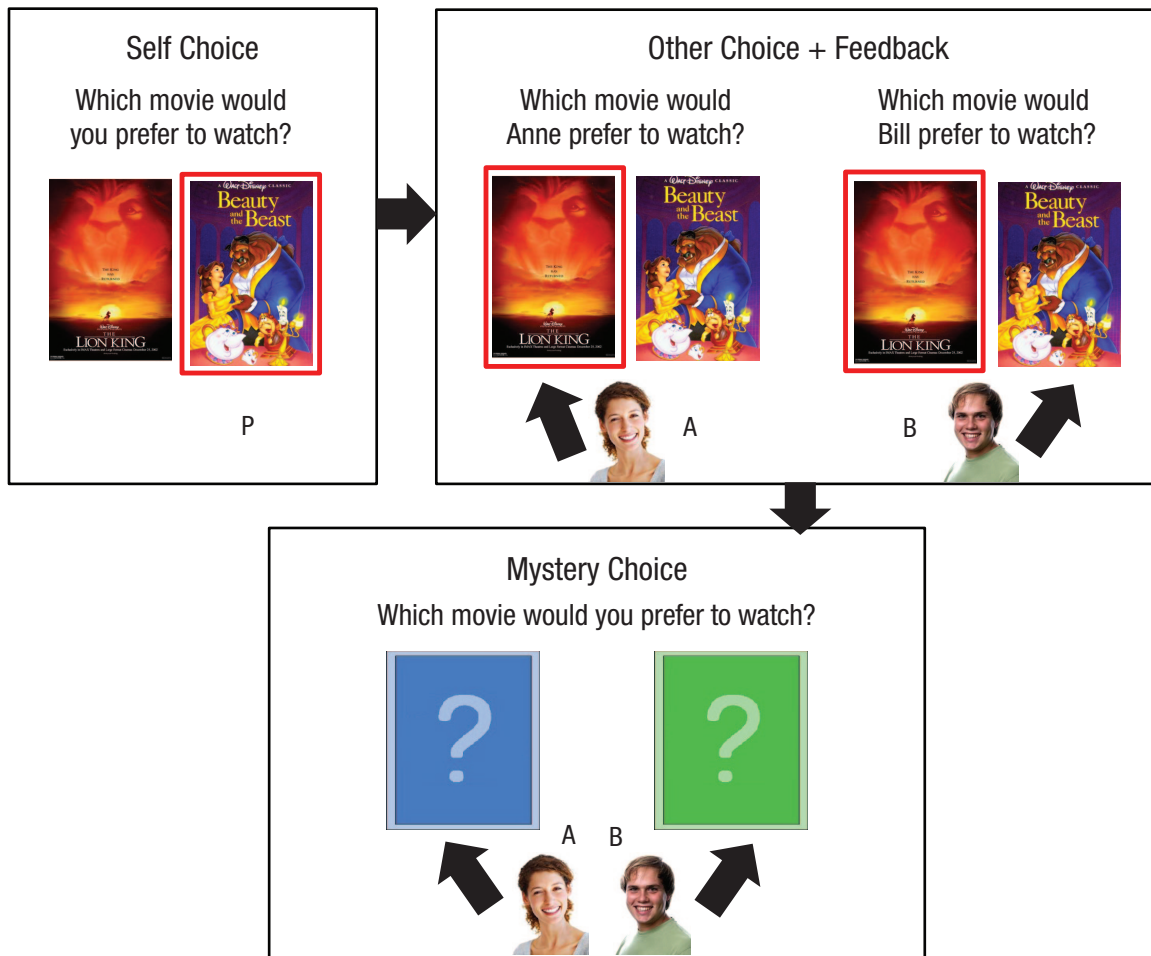


Fig. 2. Canonical experimental design. On each trial, a participant (labeled "P") first makes a choice (e.g., between two movies or political issues) and then predicts the choices of other individuals (labeled "A" and "B"). After each prediction, the participant receives feedback about those individuals' actual choices. Finally, in the mystery-choice trial, participants choose between two options using information only about the choices of other individuals.

guided by dyadic similarity on the mystery-choice trial, they would be indifferent between the two options.

Unlike the dyadic-similarity model, the structure-learning model predicts that A and B can exert differential social influence depending on the choice patterns of a third individual, C (Fig. 3). When C agrees more with P and B than with A, then P and B will tend to be grouped together, which results in B exerting greater social influence on A. Results were consistent with this prediction: Participants were more inclined to follow B's recommendation when agreement was high. Lau et al. (2018) found that high agreement about political-issue positions affected not only mystery choices but also trait attributions: Individual B was rated as more competent, moral, and likable compared with individual A on high-agreement blocks.

A companion functional MRI study (Lau, Gershman, & Cikara, 2020) investigated whether social-structure learning on the basis of political-issue positions relies

partly on a domain-general structure-learning capacity. A previous study (Tomov, Dorfman, & Gershman, 2018) found that right anterior insula was one among several regions that tracked beliefs about nonsocial causal structures in a predictive-learning task. Consistent with this finding, Lau et al.'s (2020) data showed that right anterior insula tracked beliefs about latent groups. Moreover, variability in the insula signal predicted trial-to-trial variability in social influence (i.e., the degree to which A vs. B influenced participants' ally-choice behavior).

The Development of Social-Structure Learning

The perceptual and conceptual foundations of social categorization emerge during the first few years of life (Rhodes & Baron, 2019). By 3 months, infants are attentive to the race and gender of faces (e.g., Quinn et al.,

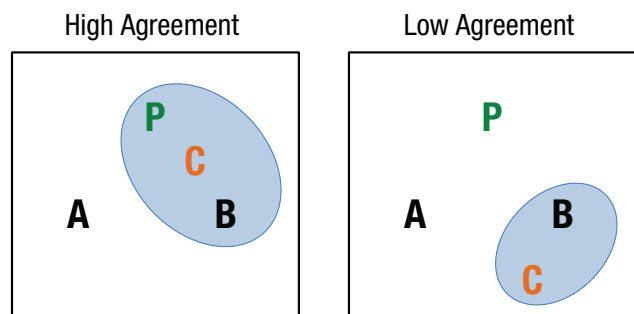


Fig. 3. Model predictions. The pattern of preferences for each individual (the participant, labeled “P,” and three other individuals, labeled “A,” “B,” and “C”) is schematized in a two-dimensional space in which proximity is proportional to preference agreement. The ellipse shows the hypothetical latent group inferred by the model. When individual C tends to agree with both P and B, then all three are grouped together. As a consequence, on the mystery-choice trial, B is expected to exert stronger social influence on P compared with A even though A and B have equal preference agreement with P. When individual C tends to agree with B but not P, then B and C are grouped together, excluding P and reducing the differential social influence of B on P.

2008). These initially perceptually based categories are later augmented by conceptual knowledge, for example that members of the same group should help one another (Jin & Baillargeon, 2017) and that members of the same group will tend to share a common psychological “essence” that licenses inductive inferences about unseen properties (Diesendruck & Eldror, 2011).

The emergence of essentialist reasoning about social groups raises the question of how these essences are determined in the first place. Looking at this question through the lens of social-structure learning, we posit that these essences correspond to beliefs about latent groups, which are endowed with the causal power to produce perceptual and behavioral regularities. If children possess this form of mental model, then observing perceptual and behavioral regularities allows them to draw inferences about underlying group structure.

Some evidence suggests that early beliefs about social categories are relatively coarse. For example, Repacholi and Gopnik (1997) found that 14-month-olds extend their own preferences (e.g., for Goldfish crackers over broccoli) to adults despite observing a demonstration that the adults have a different preference (broccoli over Goldfish crackers). By 18 months, children understand that adults may have different preferences from themselves. Later work by Doan, Denison, Lucas, and Gopnik (2015) showed that 14-month-olds could be trained to differentiate preferences by exposing them to multiple adults with different preferences. We have shown that this “diverse-desires” training effect follows naturally from the principles of social-structure

learning under the assumption that children are simultaneously learning about latent groups and the concentration parameter that controls their expectations about group complexity (Gershman et al., 2017). Specifically, diverse-desires training provides evidence that an initially strong preference for coarse grouping (everybody has the same preference) should give way to more complex grouping in order to accommodate the heterogeneity of behavioral patterns.

We formalized this idea in terms of Bayesian inference over the concentration parameter, which allowed the model to recapitulate the developmental trajectory (Gershman et al., 2017). Specifically, children were modeled as having uncertainty about both the latent groups and the parameter governing the distribution of groups. By computing a posterior over both variables, the model could capture how children learn not only who shares their preferences but also the degree to which other people tend to share their preferences. Lucas and colleagues (2014) developed a similar approach based on the idea that children build econometric models of choice and weigh the evidence for joint versus independent preferences (a simplified version of the general structure-learning problem).

Broader Implications for Intergroup Science

We would be remiss if we did not note that this work is inspired by and fits within a broader research tradition (e.g., agent-based and game-theoretic models of coalition formation); however, a thorough review of this work is beyond the scope of this article. Instead, we close by reviewing current gaps in the intergroup literature and the potential that social-structure learning has for changing how we think about categorization, identification, and stereotype updating.

Categorization and identification

Driven in part by the prominence of social-identity theory (Tajfel & Turner, 1979) and the minimal-groups paradigm, contemporary intergroup literature has emphasized the role of category membership (e.g., Black vs. White; Rattlers vs. Eagles) over coalitional structure (i.e., friends vs. foes). However, studies based on pairs of social categories make it difficult to infer from them anything about generalized group processes. For example, some but not all social groups are intrinsically confounded with differences in the visual appearance of targets, many groups carry with them particular stereotypes and associated prejudices, perceivers’ familiarity with the groups in question will vary, and so on.

The category-based approach is further limited because it is context insensitive; it breaks down as agents' goals shift or agents' other intersecting identities become salient. Context sensitivity of this kind is central to several theories of categorization and identification, including self-categorization (Turner, Oakes, Haslam, & McGarty, 1994) and optimal distinctiveness (Brewer, 1991). Both theories predict that identity salience and self-categorization are determined in large part by who else is around—for example, our identities as Americans are not very salient when we are in the United States, but that changes when we are traveling abroad. The challenge is that neither self-categorization nor optimal distinctiveness provides a means of making quantitative predictions about which identity or attribute will be made most salient or motivationally potent in a given context or precisely how identity salience shifts in response to changes in the environment (the metacontrast principle comes closest, although it is limited in that it assumes only two groups: the in-group and a single out-group). In alternative frameworks, categorizing people by specific social categories is a by-product of adaptations that evolved for detecting more general coalitions (Cosmides, Tooby, & Kurzban, 2003; Pietraszewski, Cosmides, & Tooby, 2014; Sidanius & Pratto, 1999). If this were the case, humans would have to have a flexible, common neural code for learning about and representing “friends,” “not friends,” and “foes” invariant to the particular social category or features along which group boundaries are drawn (Cikara, Van Bavel, Ingbretsen, & Lau, 2017; for a review, see Cikara & Van Bavel, 2014).

The social-structure-learning approach can accommodate these challenges—context sensitivity and common code—because, as in self-categorization theory, agents rely on current information to infer group structures. New environments and agents will prompt new inferences that include more complex structures than just an in-group and single out-group. Furthermore, social-structure learning can integrate existing categories as priors. In the absence of context-specific information, people may rely on visual cues to group membership (e.g., skin tone) but quickly redraw group boundaries on the basis of how agents interact with perceivers and each other. We found evidence of exactly this pattern in Lau et al.'s (2018) work.

Subgrouping and subtyping

One of our most remarkable capacities is the ability to create models of the world, to update those models given new information, and then to generalize knowledge from those models to new contexts, agents, and experiences. However, these remarkable capacities do

not operate in equal measure across all domains (Barsalou, 1983; Gershman, 2019). Stereotyping in particular constitutes one context in which individuals exhibit stickiness in their representations.

Allport (1954) explained one likely source of stereotypes' intransigence:

There is a common mental device that permits people to hold prejudgments even in the face of much contradictory evidence. It is the device of admitting exceptions. . . . By excluding a few favored cases, the negative rubric is kept intact for all other cases. (p. 23)

Early investigations of this *subtyping* documented individuals' propensity to do it (Maurer, Park, & Rothbart, 1995; Taylor, 1981), which was quickly followed by work examining the conditions under which subtyping was most likely to be deployed. For example, if counterstereotypical evidence is restricted to only a few targets or counterstereotypical targets are atypical along many additional dimensions, then group-level stereotypes remain intact (Hewstone, 1994; Johnston & Hewstone, 1992; Kunda & Oleson, 1995). A third alternative to updating and subtyping is subgrouping, which refers to the reclassification of stereotype-inconsistent individuals into a subordinate group (Maurer et al., 1995; Park, Ryan, & Judd, 1992). In this case, the superordinate group is only partially insulated from stereotype updating.

In the decades since this work began, dozens of articles have accumulated documenting the conditions under which updating, subgrouping, and subtyping occur; however, there is still no unified theory of hierarchical social-structure learning. Another exciting application of the social-structure approach is the ease with which it can be adapted to make principled, quantitative predictions specifying these conditions. Taking into account the heterogeneity of a group, the model makes predictions about just how stereotype inconsistent a new agent has to be to get lumped in with the superordinate category, subtyped out, or seeded as a member of a subgroup. When an individual is subtyped out, he or she can have no impact on the superordinate-category stereotypes (one can think of that person as being assigned to an “exception” group on the same level as the superordinate group), whereas individuals who come to form subgroups will have a small updating effect, and individuals lumped in with the superordinate category will have the strongest effect.

Being able to specify just how “atypical” agents need to be in order to shift stereotypes would be incredibly helpful in efforts to correct overly negative stereotypes. For example, this approach could be used to rehabilitate

perceptions of immigrants, who are often characterized as criminal (Stephan, Ybarra, & Bachman, 1999) despite data indicating either no relationship or a small negative relationship between immigration inflows and local crime rates (Ousey & Kubrin, 2018). In line with this idea, findings have shown that participants who read stories about counterstereotypic (i.e., high-achieving Syrian and Mexican) immigrants, along with high-achieving German and Russian immigrants, exhibited more positive and similar (across nationalities) evaluations of those exemplars' nationality groups relative to prestory evaluations (Martinez, Feldman, Feldman, & Cikara, 2020).

Multiple-group membership

Hierarchical social-structure learning deals with cases in which an individual can belong to multiple nested subgroups simultaneously. It is also possible for individuals to belong to multiple parallel groups simultaneously (see work on intersectionality). For example, you might sort your friends in different ways on the basis of their music preferences, movie preferences, political attitudes, and so on. This is an example of “cross-cutting” categorization, which has been studied extensively in nonsocial domains (e.g., Heit & Rubinstein, 1994), in which people use multiple notions of similarity between examples to make different inductive inferences. Shafto, Kemp, Mansinghka, and Tenenbaum (2011) developed a Bayesian structure-learning model for cross-cutting categorization in which examples can be sorted into multiple categories simultaneously, and these categories are discovered from patterns of feature covariation. This approach could be extended to social-structure learning from choice data.

Conclusion

We have argued that recent evidence and computational modeling support a structure-learning account of social-category formation. This account draws heavily from prior work on nonsocial-structure learning and advances existing social-psychological theory on category formation and hierarchical-structure updating. For researchers interested in domain-general structure learning, the social dimension adds new complexity and uncharted territory because of the fact that it invokes a notion of *self* in relation to *others*. For researchers interested in intergroup phenomena, this approach supports more specific predictions about (a) the mechanisms by which people infer social-group boundaries and (b) the temporal dynamics of this process. Thus, integrating insights from these models into intergroup cognition allows for greater predictive precision and potentially stimulates innovative strategies for reducing group-based bias. We hope this integration will

generate a groundswell of research for several areas across the psychological sciences.

Recommended Reading

- Cikara, M., & Van Bavel, J. J. (2014). (See References). An overview of intergroup cognition from a cognitive-neuroscience perspective.
- Gershman, S. J., Pouncy, H. T., & Gweon, H. (2017). (See References). The first study of social influence to present direct evidence for social-structure learning.
- Lau, T., Pouncy, H. T., Gershman, S. J., & Cikara, M. (2018). (See References). An extension of the Gershman et al. (2017) study to grouping on the basis of political-issue positions and ally-choice behavior that also indicated that social-group inferences drive trait attribution (e.g., how competent or likeable a target is).

Transparency

Action Editor: Robert L. Goldstone

Editor: Robert L. Goldstone

Declaration of Conflicting Interests


The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by a faculty research grant from the Harvard Mind Brain Behavior Initiative.

ORCID iDs

Samuel J. Gershman  <https://orcid.org/0000-0002-6546-3298>

Mina Cikara  <https://orcid.org/0000-0002-6612-4474>

Acknowledgments

We are grateful to Tatiana Lau, Thomas Pouncy, and Hyo Gweon for indispensable collaboration on this work.

References

- Allport, G. W. (1954). *The nature of prejudice*. Oxford, England: Addison-Wesley.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.
- Austerweil, J. L., Gershman, S. J., Tenenbaum, J. B., & Griffiths, T. L. (2015). Structure and flexibility in Bayesian models of cognition. In J. R. Busemeyer, J. T. Townsend, Z. Wang, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 188–208). Oxford, England: Oxford University Press.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.
- Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, *17*, 475–482.
- Cikara, M., & Van Bavel, J. J. (2014). The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science*, *9*, 245–274.
- Cikara, M., Van Bavel, J. J., Ingbretsen, Z., & Lau, T. (2017). Decoding “us” and “them”: Neural representations of

- generalized group concepts. *Journal of Experimental Psychology: General*, *146*, 621–631.
- Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*, 173–179.
- Diesendruck, G., & Eldror, E. (2011). What children infer from social categories. *Cognitive Development*, *26*, 118–126.
- Doan, T., Denison, S., Lucas, C., & Gopnik, A. (2015). Learning to reason about desires: An infant training study. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 578–583). Austin, TX: Cognitive Science Society.
- Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences*, *20*, 362–374.
- Gershman, S. J. (2019). How to never be wrong. *Psychonomic Bulletin & Review*, *26*, 13–28.
- Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*, 1–12.
- Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*, 197–209.
- Gershman, S. J., Pouncy, H. T., & Gweon, H. (2017). Learning the structure of social influence. *Cognitive Science*, *41*, 545–575.
- Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 411–422.
- Hewstone, M. (1994). Revision and change of stereotypic beliefs: In search of the elusive subtyping model. *European Review of Social Psychology*, *5*, 69–109.
- Jin, K. S., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences, USA*, *114*, 8199–8204.
- Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: 3. Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology*, *28*, 360–360.
- Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology*, *68*, 565–579.
- Lau, T., Gershman, S. J., & Cikara, M. (2020). Social structure learning in human anterior insula. *eLife*, *9*, Article e53162. doi:10.7554/eLife.53162
- Lau, T., Pouncy, H. T., Gershman, S. J., & Cikara, M. (2018). Discovering social groups via latent structure learning. *Journal of Experimental Psychology: General*, *147*, 1881–1891.
- Love, B., Medin, D., & Gureckis, T. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.
- Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., . . . Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PLOS ONE*, *9*(3), Article e92160. doi:10.1371/journal.pone.0092160
- Martinez, J. E., Feldman, L., Feldman, M., & Cikara, M. (2020). Narratives shape cognitive representations of immigrants and policy preferences. *PsyArXiv*. doi:10.31234/osf.io/d9hrj
- Maurer, K. L., Park, B., & Rothbart, M. (1995). Subtyping versus subgrouping processes in stereotype representation. *Journal of Personality and Social Psychology*, *69*, 812–824.
- Ousey, G. C., & Kubrin, C. E. (2018). Immigration and crime: Assessing a contentious issue. *Annual Review of Criminology*, *1*, 63–84.
- Park, B., Ryan, C. S., & Judd, C. M. (1992). Role of meaningful subgroups in explaining differences in perceived variability for in-groups and out-groups. *Journal of Personality and Social Psychology*, *63*, 553–567.
- Pietraszewski, D., Cosmides, L., & Tooby, J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PLOS ONE*, *9*(2), Article e88534. doi:10.1371/journal.pone.0088534
- Quinn, P. C., Uttley, L., Lee, K., Gibson, A., Smith, M., Slater, A. M., & Pascalis, O. (2008). Infant preference for female faces occurs for same- but not other-race faces. *Journal of Neuropsychology*, *2*, 15–26.
- Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*, 12–21.
- Rhodes, M., & Baron, A. (2019). The development of social categorization. *Annual Review of Developmental Psychology*, *1*, 359–386.
- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*, 1144–1167.
- Shafto, P., Kemp, C., Mansinghka, V., & Tenenbaum, J. B. (2011). A probabilistic model of cross-categorization. *Cognition*, *120*, 1–25.
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY: Cambridge University Press.
- Stephan, W. G., Ybarra, O., & Bachman, G. (1999). Prejudice toward immigrants: An integrated threat theory. *Journal of Applied Social Psychology*, *29*, 2221–2237.
- Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worschel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Pacific Grove, CA: Brooks/Cole Publishing.
- Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 88–114). Hillsdale, NJ: Erlbaum.
- Tomov, M. S., Dorfman, H. M., & Gershman, S. J. (2018). Neural computations underlying causal structure learning. *The Journal of Neuroscience*, *38*, 7143–7157.
- Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, *20*, 454–463.