Cognitive Malleability and the Wisdom of Independent Aggregation

Leonel Garcia-Marques, Ana Sofia Santos, Diane M. Mackie, Sara Hagá & Tomás A. Palma

Published online: 30 Nov 2017.

Submit your article to this journal ⏎

Article views: 323

View related articles ⏎

View Crossmark data ⏎

Citing articles: 2 View citing articles ⏎

Routledge
Taylor & Francis Group

Check for updates

Cognitive Malleability and the Wisdom of Independent Aggregation

Leonel Garcia-Marques [ID][a], Ana Sofia Santos[a], Diane M. Mackie[b], Sara Hagá[a], and Tomás A. Palma [ID][a]

[a]CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal; [b]Department of Communication, University of California at Santa Barbara, Santa Barbara, California

Prejudices and stereotypes are stubbornly hard to change. Negative images of some social groups seem to persist for hundreds, and sometimes even thousands, of years. For instance, early in the third-century BCE, an Egyptian priest called Manetho wrote a revisionist Exodus narrative in which the Pharaoh expelled the Jews from Egypt because of their supposedly misanthropic, conniving, and even cannibalistic (!) nature (van der Horst, 2006). Unfortunately, such accusations still find favor in alternative-right blogs today.

A similar case can be made for the longevity of stereotypes about and prejudices against Romani people across centuries (Hancock, 2002). And yet we have shown experimentally that the average likelihood of a person choosing the same five traits to best describe the Romani in two sessions 14 days apart is below .50 (Garcia-Marques, Santos, & Mackie, 2006). How can prejudices and stereotypes be stable over a period of thousands of years and yet unstable over only a few weeks?

Payne, Vuletich, and Lundberg (this issue) address similar puzzles. In light of the fact that the stereotypic biases revealed by individuals on implicit tests vary dramatically over the course of a few weeks, these authors ask how it is possible that responses to these same implicit tests, when examined at aggregate levels such as experimental groups, cities, states, and nations, reveal (a) large stereotypic biases, (b) developmentally consistent stereotypic biases, and (c) stereotypic biases that are highly predictive of discriminatory behavior. The explanation Payne and colleagues propose is derived from the Wisdom of Crowds effect, which is the finding that in some conditions, averaged judgments, estimates, or decisions are much more accurate or valid than the vast majority of individual judgments, estimates, or decisions that comprise them. The conditions that make this possible are when aggregation allows for error cancellation. The proposed parallel is that individual beliefs about groups are both idiosyncratically variable and idiosyncratically context dependent. But social stereotypes deriving from systemic or structural inequalities and situations in which these inequality-reflecting associations are activated are common to the members of a given society and stable across time. When the results of individual implicit tests are aggregated, individuals' idiosyncratic beliefs and context dependencies cancel one another out, and the systematic social stereotypes inherent in cultures and situations emerge clearly from these data.

At heart then, Payne and colleagues argue that the answer to such puzzles is the situational dependency of implicit biases about groups. There are two parts to the argument. First, as embodied in a "new" prediction, the authors argue that temporal instability, even at the individual level, should disappear if situational stability is held constant. Second, idiosyncratic content might be activated for any given individual across particular situations, but the content activated by societal contingencies across those same situations is highly similar, allowing social or situational stability to merge as idiosyncratic individual content is cancelled out.

We applaud Payne and colleagues' important and parsimonious emphasis on concept accessibility as a possible solution to the discontinuities in conclusions that can be drawn from individual and aggregate level responding on implicit tests of stereotyping and prejudice. Such an emphasis has been central to our explanation of the same discontinuities in our own work on the stability and malleability of social stereotypes. In fact, as we discuss next, we can lend our evidence in support of their "new prediction": Temporal instability disappears when context (we prefer that term for a reason) stability is held constant. Our work leads us to urge Payne and colleagues to go further: The impact of activated information is not unique to the use of implicit tests or even to the abstract nature of stereotype structures but instead characterizes human conceptual knowledge in general. Because of this, these apparent puzzles are particularly telling about the nature of human cognition. Finally, we raise a few questions and challenges that we hope will contribute to the development of the approach proposed by Payne and colleagues.

The Malleability of Conceptual Knowledge

In the last several decades, a growing body of empirical evidence supports the view that category representations are highly flexible and unstable knowledge structures, varying widely across different contexts and individuals (e.g., Barsalou, 1983; Barsalou & Sewel, 1984; Roth & Shoben, 1983; for reviews, see Barsalou, 1987, 1993; Yeh & Barsalou, 2006). The most decisive evidence for representational malleability, however, comes from research looking at the reliability of category representations over time, within the same individual. Using test–retest procedures, such research reveals only moderate agreement between participants' responses solicited on two separate occasions. For example, McCloskey and Glucksberg (1978) asked participants to judge the category membership of

the same exemplars in two sessions 1 month apart. The results showed that participants often changed their minds about whether an exemplar was or was not a member of a category. Barsalou, Sewell, and Ballato (1986) observed similar within individual variability across time for judgments of exemplar typicality. Thus judgments of both category membership and exemplar typicality vary within individuals over time.

Such within-individual malleability is also typical of memory retrieval. Bellezza (1984a) asked participants to generate as many exemplars from the same common taxonomic categories as they could in two sessions separated by 1 week. He found that only 69% of the exemplars retrieved in the first session were retrieved again 1 week later. Retrieval reliability is even lower for noun meanings, facts about people, and scripts (Bellezza, 1984b, 1984c, 1988). Perhaps most surprisingly, retrieval reliability of information about the self, thought to be better organized and more salient than information about other knowledge domains (e.g., Epstein, 1973; Greenwald, 1981; Kuiper & Rogers, 1979), is in fact even lower (51%) than the retrieval reliability for common taxonomic categories such as *fruit* and *fish* (76%; Bellezza, 1987; see also Greenwald, Bellezza, & Banaji, 1988).

Such findings suggest that what is retrieved from memory on any specific occasion reflects a constructive process that is highly constrained by the surrounding context, as well as by people's ongoing goals and recent experiences (Barsalou, 1987, 1993; Yeh & Barsalou, 2006; for recent empirical and theoretical works, see Connell & Lynott, 2014; Lebois, Wilson-Mendenhall, & Barsalou, 2015; Yee & Thompson-Schill, 2016). These findings suggest that within-individual conceptual malleability is rather general and not due to the cognitive peculiarities of stereotypes and prejudices or even of implicit measures. Our own research with social stereotypes tells a similar story.

The Malleability of Social Stereotypes and Other Social Knowledge Structures

Our own work has dealt specifically with the issue of the intraindividual malleability of social stereotypes. In a series of studies, we asked participants to choose the five traits (from a list of 43) that best described several different social groups and to rate those groups on several bipolar trait scales in two separate sessions two to four weeks apart (Garcia-Marques et al., 2006). Our results showed considerable between-individual (group level or aggregate) stability in both trait choices and ratings such that, for instance, the rank order in which traits were chosen was basically identical across sessions. When we looked only at such across-participants within-item agreement, as previous research has done (Devine & Elliot, 1995), the evidence was firmly consistent with the idea that stereotypes are stable across time.

In stark contrast, however, the within-participants measures revealed considerable malleability in people's representation of social categories (similar to the degree of malleability found in representations of common nonsocial categories; Barsalou et al., 1986; Barsalou, Spindler, Sewell, Ballato, & Gendel, 1987; Bellezza, 1984a, 1984b, 1984c). In fact, regardless of whether we looked at the choice of the traits thought to be most stereotypic of a social category and trait ratings (Study 1), typicality ratings of various category members (Study 2), or retrieval of

category exemplars (Study 3), the results all converged to provide evidence of considerable malleability. The level of within-participant malleability we obtained could not be explained by feature or exemplar centrality or importance. Although there was significantly less malleability in more central or important trait attributes (Study 1) and in more typical group members (Studies 2 and 3), even central, important, and typical traits and exemplars showed considerable intraindividual malleability. Thus we found the same discontinuity between interindividual and intraindividual stability that Payne and colleagues set out to explain with our explicit measures of stereotypes: These effects are not limited to implicit measures.

It should also be noted that these results were not unique to individuals' idiosyncratic responses. In the first of the aforementioned studies, we chose to manipulate whether people responded about cultural or individual stereotypes—an important distinction in social stereotyping introduced by Devine and Elliot (1995). According to these authors, the cultural stereotype content—what "people in general" think—seems to be more stable than individual or personal stereotypes are. However, because their data were not longitudinal, Devine and Elliot (1995) were not able to assess whether cultural stereotypes have greater stability than personal stereotypes in a within-participant sense. Our design allowed us to do so. When participants chose the five traits (that best described a group from a list of 43 traits, not only was the probability of including the same trait at both Session 1 and Session 2 approximately equal to the probability of including it at only one time, but this result also held for both personal and cultural stereotypes. This finding is perhaps at odds with the Bias of Crowds proposal, as it seems to suggest no differences in malleability between cultural and individual stereotypes.

Our focus then turned to what drove stability or malleability at the individual level. In the last study in this series, we were able to show that the level of malleability in an individual's stereotypes over time depended crucially on whether the context, and thus what other information was activated, in which the stereotype was activated was stable or changed. In this study, we assessed participants' stereotypes over time by asking them to choose five traits that best described a particular group in each of two sessions held two weeks apart. Before reporting their stereotypes in the first session, participants read a description of three group members and rated their typicality. They read about and rated a second set of three members of the group before reporting their stereotypes in the second session. The exemplars described were either stereotype-consistent (two stereotype-consistent and one neutral) or stereotype-inconsistent (two stereotype-inconsistent and one neutral), and we manipulated the context in which the stereotype was assessed by crossing the stereotype consistency of the exemplars rated during the first session and the stereotype consistency of the exemplars rated during the second session. Results revealed that within-individual stereotype stability was largely a function of context stability. When the context was stable (the different exemplars activated in both sessions were either both stereotype-consistent or both stereotype-inconsistent), the degree of within-subject stereotype stability was considerable. When the context differed (the exemplars differed in stereotype-consistency), stereotype stability greatly declined.

Thus, our results revealed that malleability or stability in stereotype representation depended crucially on the apparent incorporation of currently activated information into the stereotype across times when a response is required. When the context dictated that the same content was activated across response trials, stereotypes looked stable. When the context caused inconsistent information to be activated across responses, stereotypes looked variable.

More recent studies (Santos, Garcia-Marques, Mackie, Ferreira, et al., 2012; Santos, Garcia-Marques, Mackie, Palma, et al., 2017) have shown the extent to which such incorporation can influence stereotypes. Even stereotype-unrelated traits may become "one of the best descriptors of the group" when they have been subtly primed in an unrelated task. When participants were asked to list the traits that make up the stereotype of computer programmers, almost no one described them as friendly. But when "friendly" was primed, a large majority now reported that friendly was part of the computer programmer stereotype.

These results are also not unique to the study of social stereotypes. In another line of work, we explored the stability of people's implicit personality theories (Hagá, Garcia-Marques, Santos, Barros, & Nunes, in preparation). We found that the stability at the item level—namely, traits—aggregating across participants is higher than the stability at the participant level aggregating across items. In other words, we looked at how many participants, in a first session, thought that a given pair of traits (e.g., friendly and tolerant) tend to co-occur in people's personality. We looked again in a second session 3 weeks apart. These values result from the aggregation of individuals, and in order to examine stability we calculated the overlap between sessions. Simultaneously, we looked at how many trait pairs a given participant formed in the first session and looked again in the second session. These values result from the aggregation of items, and again we calculated the overlap between sessions. As mentioned, the overlap using the first method was higher than the overlap using the second method, suggesting higher interindividual than intraindividual stability. Notably, once again, contextual factors seem to play an important role. If different groups of participants, while completing an ostensibly unrelated task, are exposed to exemplars who vary in their combination of traits (e.g., an exemplar is warm and motivated or warm and unmotivated), then the perceived relationships between a different set of traits will change at the group level. Although implicit personality theories might also be thought of as abstract, recall also the context sensitivity of retrieval for exemplars of natural and object categories, and even of information about the self. Thus although the abstractness of stereotypes might increase their context dependency, it is not a boundary condition for incorporation of activated information into a response.

We have argued that both our own and others' findings on concept malleability and context sensitivity can be most parsimoniously explained by appealing to constructionist models of stereotypes (see Garcia-Marques et al., 2006; for a more general reference, see the source of activation confusion model: Ayers & Reder, 1998; Reder & Schunn, 1996). According to this perspective, because the information associated with or encoded in knowledge structures like stereotypes is vast and diverse, only part of that information will be available at a given moment depending on the perceivers' current context, goals, and tasks. Thus, instead of all the information associated with a stereotype being activated in an all-or-nothing fashion, a stereotype must be reassembled whenever it is needed, creating at that given moment a "working stereotype" (Santos et al., 2012) that then guides further processing. In our view, a working stereotype is a dynamic online structure that is assembled from or constructed of what is activated—not only the activated part of the information associated with the stereotype but also other contextually accessible information. Our proposal assumes that the working stereotype varies across time and context because, although information closely associated with the stereotype is more likely to be both activated and incorporated, the reassembling process is vulnerable to influence at a number of levels. First, given the vast amount of information usually associated with a stereotype, stereotype-assembly is not expected to be an exhaustive process, and different activated subsets of that information will contribute different content as the stereotype is reassembled. Second, other concepts activated in the immediate context become potentially available for incorporation as the stereotype is reassembled. Such contamination typically happens because people often lack direct awareness of the source of concept activation, or more simply why information is activated. In this sense, one way of interpreting stereotype malleability is as a side effect of source confusion errors. Indeed, a very similar and convergent view can be found in the source monitoring framework (Johnson, 2006), according to which thoughts/images/feelings that come to mind do not include abstract tags or labels that name their sources but rather have qualitative and quantitative characteristics that are more or less diagnostic of the source.

According to our proposal, intraindividual malleability is not the result of the use of flawed individual measures; instead it is the result that should be expected from adaptive and flexible cognitive processes. When context dependency is controlled for—for instance, in situations in which the basic characteristics of the immediate context across sessions are held constant—intraindividual stability is very high.

Similarities and Differences Between the Bias of Crowds Approach and Ours

Despite the fact that Payne and colleagues (this issue) focus on implicitly measured stereotypic biases and our own work focuses on explicit indices of stereotyping, the similarities between the two approaches should be apparent. Both approaches have simultaneous responses at the individual level that are characterized by malleability but that produce at the aggregate level responses that are characterized by consistency and both approaches assume that interindividual aggregation of choices, judgments, or responses allow for reinforcement of the overlapping knowledge (i.e., social stereotypes) and mutual cancellation of idiosyncratic knowledge.

Both approaches produce similar outcomes. Therefore, at least some of our differences must be secondary to the obtained pattern of results. Thus our convergence means that these dissociations between intraindividual malleability and interindividual stability must not occur because of something special

about implicit measures, nor because implicit measures are more "abstract" than explicit measures. Our convergence of results should instead result from the context-dependent nature of human cognition. In fact and as aforementioned, when the stability of the context is manipulated, intraindividual measures became very stable, too.

According to Payne and colleagues (this issue), the basic lesson to be learned from these data is that implicit tests are not valid measures of individual representations and may be conceived, instead, as valid measures of social situations. In contrast, we take the temporal unreliability of implicit (and explicit) measures to be but a reflection of adaptive, flexible, and context-dependent cognitive processes. Our answer to the malleability of intraindividual implicit measures is that implicit measures (and many other individual measures) tap on socially shared knowledge between individuals. As a consequence, and although shared knowledge is very important in accounting for both intra and interindividual representations, it is useless in a purely individual-differences approach. That is, because implicit (and other) measures tap on shared knowledge and beliefs, they are reliable at assessing common knowledge and beliefs but unable to discriminate between individuals.

In the context of the overall compatibility of our approaches, we raise some questions that we believe can usefully push these fascinating ideas forward.

Questions and Challenges

- We prefer the term *context* to *situation* because of its agnosticism when it comes to influences internal and external to the individual. In arguing that activation is a source of malleability in implicit responses, Payne and colleagues end up arguing that implicit measures—at least at the aggregate level—are better gauges of what situations activate than of what individuals activate. But activation works both ways. It is not immediately clear that internal contexts are any less salient or stable than external socially driven contexts. Chronic mood might be just as stable an internal context as societal inequality is a stable external context. Heat in the environment might be just as fleeting an external activator as a passing fever. Further, the two interact: An external cue may trigger self-categorization. An individual thinking about herself as an ingroup member compared to someone not self-categorized will interpret the cues in a situation differently. The activation mechanism (and all the models that build upon it) tells us that it is activation that counts, and not how or why the activation occurred. Thus reifying situations may be as problematic as reifying individual differences in trying to understand when and why intraindividual responses vary even as interindividual responses converge.
- The effects of Wisdom of Crowds presuppose independence between agents—only with independent agents can error cancellation occur. However, the very notion of socialization seems to derive from direct influence, vicarious learning, imitation, and *nonindependence* in general. As Tajfel (1981) pointed out, when learning about different social groups and acquiring the prevalent social

stereotypes along the way, social feedback regarding social biases is likely to confirm these prevalent biases instead of correcting them. Thus, biases about a social group can, therefore, become quite consensual and stable across perceivers, not because of independent assessment of social groups but quite the opposite, as a result of direct social influence and social learning.
- Another set of findings consistent with the current perspective are the recent findings that cross-cultural differences in responses are driven more by the individual's perception of the values endorsed by other members of the culture (producing aggregate-level stability) than by an individual member of the culture's own personally endorsed values (producing idiosyncratic variability). For example, Zou et al. (2009) found that perceptions of consensual collectivism differed between the United States and Poland, even as measures of personal collectivism did not. Consensual (but not personal) collectivism mediated the well-established effect of culture on for example, causal attribution and counterfactual thinking. As Chiu, Gelfand, Yamagishi, Shteynberg, and Wan (2010) put it, "What individuals see inside themselves … does not always channel psychological processes; what the individuals see when looking outward at their social environments can also direct behaviors" (pp. 482–483). Such results are equally prevalent among collectivists and individualists. Although the outcome of this process is the same as what Payne and colleagues predict, it is not clear that the mechanism that produces it is simply the activation of more stable features of *situations* but rather of (perhaps even false) perceptions of associations shared and valued by, or representative of, group or cultural memberships
- The problem of interindividual stability versus intraindividual malleability of cognitive representations seems to cut across many different domains of human knowledge. What is then special about social stereotypes? Should not we instead take interindividual stability versus intraindividual malleability of cognitive representations in general as our starting point?
- Malleability is not a necessary outcome of intraindividual measures. Fishbein and Ajzen (1974), for instance, have already shown that, although attitude scores (measured by attitude scales) can be highly correlated with total behavioral indices (that aggregate across a series of behaviors performed by the same individual), the average correlations of this attitude score with each individual behavior can be low and nonsignificant. One could argue then that stability results from ergodic[1] aggregation. That is, aggregation across individuals in a given moment in time or across moments for a single individual give rise to identical results.
- Although the Bias of Crowds is a very parsimonious approach, there are alternative views equally or even

---

[1]Ergodicity is a feature of dynamical systems in which transactional sampling across its sampling sample at one given moment, produce identical results to longitudinal sampling from a segment of that space. Such systems are highly used, for instance, in Economics (Samuelson, 1976).

more parsimonious. For instance, according to Sherif, Sherif, and Nebergall (1965), an individual's position "is not represented adequately as a point along a continuum. Different persons espousing the same position may differ considerably in their tolerance around this point" (p. 222). Or as Wyer (1973) put it, "A subject's rating of an object along a category scale is interpreted as the "expected value" of a distribution of subjective probabilities that the object belongs to each of the scale categories" (p. 446). Similar views have been offered, more recently by Mussweiler and Strack (2001). According to such views, intraindividual variability is to be expected (at least, within the acceptable latitude of each individual) and interindividual stability is but a consequence of the overlap of the different latitudes of acceptance. Thus individual responses are both unstable (within each latitude of acceptance) and stable (interlatitudes of acceptance). If we take within-latitude variability as context dependent, the similarity with our views is obvious. The latitudes of acceptance views can also parsimoniously account for Payne and colleagues' developmental conundrum. If we take into account that children's implicit biases seem to be as unstable as adults' (e.g., Rae, 2016), then this conundrum becomes a particular instance of the first one. As long as we conceive of individuals' responses as moving freely within latitudes of acceptance, instead of reflecting fixed points, the apparent developmental stability would be the result of overlapping latitudes between children and adults samples. This perspective is compatible with the view that some stereotypic and evaluative associations may be learned early in life and that bias, at least in part, may become ingrained, as proposed by previous theories cited by Payne and colleagues.

- The constructionist approach also adds to the debate about whether the responses that are assembled in any given response are conscious or unconscious. All semantic wrangling aside, contamination of a reassembled stereotype can often occur because people often lack direct awareness of the source of concept activation, or more simply why information is activated.

Conclusion

Imagine a world in which each person would only experience single and unique events. Each moment would never repeat, and each person would live it in a complete idiosyncratic manner. We could not model new experiences in terms of our previous experiences. Our experiences could not be shared with others. No matter how many times we would experience a sunset, without developing expectancies about what a sunset is and without sharing them, each sunset could be beautiful but always unknowable.

Now imagine a world in which novelty and uniqueness were totally absent. In this world there would be nothing to learn and little reason for humans to acquire differentiated identities.

Both the aggregation of overlapping and the cancelation of nonoverlapping judgments, responses, and/or representations are not mere statistical curiosities. In fact, according to simulation, globalist, and/or connectionist models (Barsalou, 1999;

Hintzman, 1986; Smith, 1996), our knowledge and memories are superimposed or, at least, combined at the output level. As a consequence, learning experiences that vary too widely cancel one another and become forgettable, whereas repeated or overlapping experiences are preserved and allow for abstraction and generalization. Thus, as we have argued before, both lower and higher malleability derive from aggregation and cancelation (be it across individuals in a given moment in time or across moments for a single same individual). But aggregation and cancelation is the stuff cognition is made of. If we are to grasp the workings of social cognition, we have to understand both and to model their combined effects. The Bias of Crowds approach is a superb step in that direction.

ORCID

Leonel Garcia-Marques http://orcid.org/0000-0003-0800-7664
Tomás A. Palma http://orcid.org/0000-0003-2936-4732

References

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, 5, 1–21.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition*, 11, 211–227.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts reconsidered: The ecological and intellectual bases of categories* (pp. 101–140). New York, NY: Cambridge University Press.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. *Theories of Memory*, 1, 29–31.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, 22, 637–660.

Barsalou, L. W., & Sewell, D. R. (1984). *Constructing representations of categories from different points of view*. Technical Report. Atlanta, GA: Emory University. Retrieved from http://eprints.gla.ac.uk/112570/

Barsalou, L. W., Sewell, D. R., & Ballato, S. M. (1986). *The instability of categories as measured by graded structure*. Unpublished manuscript, Emory University, Atlanta, GA.

Barsalou, L. W., Spindler, J. L., Sewell, D. R., Ballato, S. M., & Gendel, E. M. (1987). *Assessing the stability of category representations with property generation*. Unpublished manuscript, Emory University, Atlanta, GA.

Bellezza, F. S. (1984a). Reliability of retrieval from semantic memory: Common categories. *Bulletin of the Psychonomic Society*, 22, 324–326.

Bellezza, F. S. (1984b). Reliability of retrieval from semantic memory: Information about people. *Bulletin of the Psychonomic Society*, 22, 511–513.

Bellezza, F. S. (1984c). Reliability of retrieval from semantic memory: Noun meanings. *Bulletin of the Psychonomic Society*, 22, 377–380.

Bellezza, F. S. (1987). Reliability of retrieving information from knowledge structures in memory: Self information. *Bulletin of the Psychonomic Society*, 25, 407–410.

Bellezza, F. S. (1988). Reliability of retrieving information from knowledge structures in memory: Scripts. *Bulletin of the Psychonomic Society*, 26, 11–14.

Chiu, C.-Y., Gelfand, M. J., Yamagishi, T., Shteynberg, G., & Wan, C. (2010). Intersubjective culture: The role of intersubjective perceptions in cross-cultural research. *Perspectives in Psychological Science*, 5, 482–493.

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, 6, 390–406.

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin*, 21, 1139–1150.

Epstein, S. (1973). The self-concept revisited: Or a theory of a theory. *American Psychologist*, 28, 404–416.

Fishbein, M., & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 81, 59–74.

Garcia-Marques, L., Santos, A. S. C., & Mackie, D. M. (2006). Stereotypes: Static abstractions or dynamic knowledge structures? *Journal of Personality and Social Psychology*, 91, 814–831.

Greenwald, A. G. (1981). Self and memory. *Psychology of Learning and Motivation*, 15, 201–236.

Greenwald, A. G., Bellezza, F. S., & Banaji, M. R. (1988). Is self-esteem a central ingredient of the self-concept? *Personality and Social Psychology Bulletin*, 14, 34–45.

Hagá, S., Garcia-Marques, L., Santos, A. S. C., Barros, C., & Nunes, L. (in preparation). *How stable are our implicit personality theories across time? A group and individual level analysis*.

Hancock, I. F. (2002). *We are the Romani people*. Hatfield, UK: University of Hertfordshire Press.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411–428.

Johnson, M. K. (2006). Memory and reality. *American Psychologist*, 61, 760–771.

Kuiper, N. A., & Rogers, T. B. (1979). Encoding of personal information: Self-other differences. *Journal of Personality and Social Psychology*, 37, 499–514.

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, 39, 1764–1801.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory and Cognition*, 6, 462–472.

Mussweiler, T., & Strack, F. (2001). Considering the impossible: Explaining the effects of implausible anchors. *Social Cognition*, 19, 145–160.

Rae, J. R. (2016). *From kids to counties: New directions in implicit social cognition* (Unpublished doctoral dissertation). University of Washington, Seattle, WA.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45–78). Hillsdale, NJ: Erlbaum.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, 15, 346–378.

Samuelson, P. (1976). Optimality of sluggish predictors under ergodic probabilities. *International Economic Review*, 17, 1–7.

Santos, A. S., Garcia-Marques, L., Mackie, D. M., Ferreira, M. B., Payne, B. K., & Moreira, S. (2012). Implicit open-mindedness: Evidence for and limits on stereotype malleability. *Journal of Experimental Social Psychology*, 48, 1257–1266.

Santos, A. S., Garcia-Marques, L., Mackie, D. M., Palma, T. A., Costa, R. S., & de Almeida, F. (2017). Something in the way you primed me: Belief monitoring when source identification is not possible. *Social Cognition*, 35, 273–298.

Sherif, C. W., Sherif, M., & Nebergall, R. E. (1965). *Attitude and attitude change: The social judgment-involvement approach*. Philadelphia, PA: Saunders.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology*, 70, 893–912.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. New York, NY: Cambridge University Press.

van der Horst, P. W. (2006). *Jews and Christians in their Graeco-Roman context. Selected Essays on Early Judaism, Samaritanism, Hellenism, and Christianity*. Tübingen, Germany: Mohr Siebeck GmbH & Co. KG.

Wyer, R. S. (1973). Category ratings as "subjective expected values": Implications for attitude formation and change. *Psychological Review*, 80(6), 446–467.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, 23, 1015–1027.

Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *American Journal of Psychology*, 119, 349–384.

Zou, X., Tam, K. P., Morris, M. W., Lee, S.-L., Lau, I. Y.-M., & Chiu, C.-Y. (2009). Culture as common sense: Perceived consensus versus personal beliefs as mechanisms of cultural influence. *Journal of Personality and Social Psychology*, 97, 579–597.