

A Dispositional Account of Aversive Racism

Carole J. Lee (c3@uw.edu)

Department of Philosophy, Box 353350
Seattle, WA 98195 USA

Abstract

I motivate and articulate a dispositional account of aversive racism. By conceptualizing and measuring attitudes in terms of their full distribution, rather than in terms of their mode or mean preference, my account of dispositional attitudes gives ambivalent attitudes (*qua* attitude) the ability to predict aggregate behavior. This account can be distinguished from other dispositional accounts of attitude by its ability to characterize ambivalent attitudes such as aversive racism at the attitudinal rather than the sub-attitudinal level and its deeper appreciation of the analogy between traits and attitudes.

Keywords: implicit attitudes; implicit bias; dispositional attitudes; attitudes

Introduction

Aversive racism is characteristic of individuals who consciously endorse egalitarian ideals but harbor less favorable feelings towards the relevant racial group – feelings which may or may not “be admitted inwardly” (Kovel, 1988, pp. 54-5) and cause subtle patterns of discriminatory behavior against members of that group (Dovidio & Gaertner, 2004). There are many reasons, both intellectual and practical, to make our account of what it is to have an attitude answerable to this phenomenon. Socially, aversive racism is thought to be partly responsible for persisting racial gaps on health, education, employment, and economic outcomes despite Americans’ increasing tendency over the last half-century to endorse ideals of equality (Bobo, 2001; Dovidio & Gaertner, 2004). Psychologically, the case of aversive racism against blacks serves as the archetypical case for measuring and explaining how explicit and implicit attitudes can come apart (Greenwald et al., 2009). And, philosophically, norm-discordant conditions like aversive racism seem “to mandate such a radical reconceptualization of the relation between cognition and behavior that traditional notions like belief seem quaint and inadequate” (Gendler 2008b, p. 642).

What characterization of “attitude” best accounts for the attitudes held by aversive racists? In progressive debates about how to draw our “periodic table of attitudes” (Gendler, 2008b, p. 560), philosophers have pursued two strategies. One strategy preserves some core notion of belief (as propositional, norm-sensitive, and consciously-accessible) while adding a new, contrasting category – such as aliefs (Gendler, 2008a, 2008b), patchy endorsements (Levy, 2015), structured beliefs (Mandelbaum, 2015), or co-activated representational contents (Holroyd, 2016) – to the

taxonomy.¹ Although these theorists disagree about how best to distinguish their newly posited attitude of art from beliefs, they adopt a common strategy for characterizing aversive racists: aversive racists hold egalitarian beliefs that are discordant with the new attitude type. An important advantage of this style of approach is that it allows us to explicate aversive racism in terms of scientifically informed but folk-ready categories.² However, a recurring concern about this approach is skepticism about “whether the phenomena that we see are sufficient to motivate wheeling in the big gun of a new fundamental taxonomical category” for attitudes (Egan, 2011, pp. 67-8) rather than accommodating them with conceptual resources already at our disposal (Mandelbaum, 2013; Kwong, 2012).

The second strategy for explicating aversive racism hews to a single attitude category. It adverts to the lower-level basis of the attitude – such as belief-fragments (Egan, 2008, 2011), mental states/processes (Machery, 2016), and finer-grained dispositions (Schwitzgebel, 2010, 2013) – to identify aversive racism as discordance among those sub-attitudinal elements. An advantage of this view is that it sidesteps positing a new taxonomical category. However, it does so by silencing our ability to ascribe aversive racism at the attitudinal rather than sub-attitudinal level: attitudes are not ascribable to those with the fragmented (Egan, 2011), ambivalent (Machery, 2016), or in-between (Schwitzgebel, 2002, 2010, 2013) dispositions of aversive racists.

I will introduce an account of attitude that enjoys the advantages of and avoids the disadvantages of these taxonomical strategies. I will explain how we can describe aversive racism at the attitudinal (rather than the sub-attitudinal) level while positing only one kind of attitude type. However, rather than cast aversive racism into a traditional mold (Kwong, 2012), my solution looks forward by leveraging innovative methods for measuring psychological constructs to better characterize aversive racism and other norm-discordant attitudes.

¹ Unlike Gendler, Mandelbaum and Levy believe that implicit attitudes are, like explicit beliefs, propositional and norm-sensitive. Despite this, they do not take implicit attitudes to “belong to the same natural kind” as beliefs (Mandelbaum, 2015, p. 636) or to be sufficiently sensitive to other mental representations “to properly be considered beliefs” (Levy 2015, p. 800).

² Social psychologists have preferred to explain aversive racism by adding a new type of attitude, namely implicit attitudes (for a dissenting view, see Fazio, 2007). This allows them to describe aversive racists as those who simultaneously hold egalitarian explicit attitudes and racist implicit attitudes towards the same social group (Son Hing, Li, & Zanna, 2002).

The particular view I will introduce adopts a dispositional account of attitude. Like other dispositional accounts proposed by philosophers (Machery, 2016; Schwitzgebel, 2010, 2013), I will explicate my view by drawing an analogy between attitudes and personality traits. However, I will deepen this analogy in a way that reflects, not just how the structure of traits and attitudes mirror each other, but how meta-theoretical debates about and methods for characterizing traits and attitudes should mirror each other. Moreover, unlike other dispositional accounts – which deny the very possibility of attributing ambivalent attitudes (Machery, 2016, p. 124; Schwitzgebel, 2010, p. 544) – this view has the capacity to describe the contradictory tendencies of ambivalent attitudes such as aversive racism *qua* attitude.

Dispositional Accounts of Attitudes

According to a dispositional approach to conceptualizing attitudes, attitudes are tendencies to cognize and behave towards an object (Cronbach & Meehl, 1955; Krosnick, Judd, & Wittenbrink, 2005; Eagly & Chaiken, 2007; Greenwald & Nosek, 2008). Dispositions are kept conceptually separate from and remain agnostic about claims about the representations and processes underwriting them (Fazio, 2007; Borsboom, Mellenbergh, & van Heerden, 2004). Dispositional constructs posited in psychology include personality traits such as the Big Five (openness, conscientiousness, extraversion, and neuroticism).³

The Analogy between Traits and Attitudes

Personality traits and attitudes – when conceived as dispositional constructs – have a number of structural features in common. Philosophers who advocate for a dispositional approach to attitudes have leveraged some of these commonalities to explicate their views in illuminating ways (Schwitzgebel, 2002, 2013; Machery, 2016). What these views have in common with each other and with the dispositional approach adopted by social psychologists (Krosnick, Judd, & Wittenbrink, 2005; Eagly & Chaiken, 2007; Greenwald & Nosek, 2008) are the following basic ideas. Traits/attitudes are broad track dispositions to cognize and behave in certain ways. These dispositions are modal generalizations (concerning not just how people do in fact cognize and behave but how they would tend to across some relevant range of conditions). These dispositions/generalizations hold *ceteris paribus*. Individuals can be said to have more or less of a trait/attitude or not, where individual differences in the degree to which someone has a trait/attitude can be used to describe and predict behavior and cognition. Asking individuals to self-report their own trait/attitude is just one of many methods

for measuring the presence/degree of the dispositional construct of interest. Individuals can sincerely but falsely report the presence/degree to which they hold a trait/attitude as in cases of self-ignorance and self-deception; and, self reports can vary by context (Breckler & Wiggins, 1989). More generally, “manifestations of attitudes, as assessed by *any* measurement procedure” are “manifested imperfectly both by our measurement procedures and by other observable behaviors that it in part motivates” (Krosnick, Judd, & Wittenbrink, 2005, p. 23). As such, any method used to measure traits/attitudes are understood to be imperfect.

How are attitudes, understood as dispositional constructs, related to the mental states and processes posited by mental state theorists who prefer to posit cognitive theories of attitude? In my view, the way to answer this question is to, again, think analogically to the study of personality traits. Like the attitudes literature, the personality trait literature underwent meta-theoretical, conceptual, and methodological debates about whether to characterize their psychological posit of interest at the level of dispositional construct or representation/process (Fazio, 2007; Mischel & Shoda, 1998). In what I’ll call the “standard approach” to solving this dilemma, personality theorists characterize traits as dispositions that are underwritten by representations and processes that serve as the psychological basis for those dispositions (Epstein, 1994). Likewise, social psychologists (Greenwald & Nosek, 2008; Eagly & Chaiken, 2007) and some philosophers (Machery, 2016) who advocate for a dispositional approach to characterizing attitudes suggest that attitudes are dispositions that are underwritten by representations and processes – representations and processes that mental state theorists aim to uncover.

Sub-Attitudinal Accounts of Aversive Racism

Social psychologists who posit two types of dispositional attitude (one implicit and one explicit) can characterize aversive racism as a conflict at the attitudinal level: e.g., between positive explicit attitudes versus racist implicit attitudes towards the same group). However, philosophers who posit dispositional accounts of attitude have been taxonomically more conservative, preferring to posit one rather than two attitude types. In what follows, I will explain why, according to these philosophical views, aversive racism is describable only at the sub-attitudinal rather than the attitudinal level. I will then explicate my own view, which draws on new methods for characterizing personality traits, describes aversive racism at the attitudinal level, and, in so doing, articulates my solution to the taxonomical puzzle.

Aversive Racism as an In-Between Attitude

The heart of Eric Schwitzgebel’s dispositional account is the idea that “[t]o have an attitude is, primarily, to have a dispositional profile that matches, to an appropriate degree and in appropriate respects, a stereotype for that attitude, typically grounded in folk psychology” (Schwitzgebel,

³ Constructs are simply theoretical posits that figure in psychological generalizations and explanations (Shadish Jr., Cook, & Campbell, 2002). For a classic discussion on the validation of dispositional constructs, see Campbell and Fiske (1959).

2013, p. 78). Each dispositional, folk-psychological stereotype can be broken down into sub-attitudinal dispositions, including behavioral, cognitive, and phenomenal dispositions (Schwitzgebel, 2002). For example, being extraverted is just to have further stereotypical dispositional tendencies like enjoying meeting new people, enjoying parties, “to be talkative, and to take the lead in social situations” (Schwitzgebel, 2013, p. 81). Likewise, regarding one’s colleagues as a talented group is to be disposed to feel proud to be among them, be unsurprised when they win awards, and seek them out for insight, among other things (Schwitzgebel, 2013).

Schwitzgebel rightly observes that “[f]ew of us are 100% extravert or 100% introvert, 100% high-strung or 100% mellow” – that we tend to match such stereotypical dispositions imperfectly (Schwitzgebel, 2013). Analogously, he observes that we can fail to match stereotypical attitude dispositions imperfectly (Schwitzgebel, 2013). In imperfect cases, Schwitzgebel suggests that “[r]oughly speaking, *the greater the proportion* of stereotypical dispositions a person possesses, and the *more central* these are to the stereotype, the more appropriate it is to describe him as having the belief in question” (Schwitzgebel, 2001, p. 81, italics mine): methodologically, the injunction is to attribute an attitude when a subject has the *majority* of the stereotypical sub-attitudinal dispositions or just the *central* ones, whatever those turn out to be for that particular context.

How does Schwitzgebel’s account of attitude describe aversive racism? On his view, aversive racism can’t be characterized as a contradictory attitude since contradictory attitudes are “in general impossible” (Schwitzgebel, 2010, p. 544): we can’t describe a subject as simultaneously having the majority/central stereotypical sub-attitudinal dispositions *for* and *against* *p*. Instead, Schwitzgebel suggests that we treat aversive racism as an in-between attitude (Schwitzgebel, 2010, 2013).⁴ For in-between attitudes, “it’s not quite right, as a general matter, either to ascribe or to deny” attitudes/traits “*simpliciter*” (Schwitzgebel, 2010, pp. 535-7). Instead of trying to describe such cases at the attitudinal level, he suggests that we move instead to “more complicated appeals to specific dispositions or sets of dispositions” (Schwitzgebel, 2002, p. 266): in the case of aversive racism, we should describe individuals in terms of sub-attitudinal behavioral, phenomenal, and cognitive dispositions (e.g., the tendency to report holding egalitarian views while also tending to differentially attribute “brilliance” to white rather than black students). Accordingly, Schwitzgebel’s view accounts for and describe aversive racism at the sub-attitudinal rather than the attitudinal level.⁵

⁴ Although Schwitzgebel originally characterized his analysis of in-between cases as cases of in-between *belief* (Schwitzgebel, 2001), he has since extended the generality of his account to cover cases of in-between *attitude* as well (Schwitzgebel, 2013).

⁵ In passing, Schwitzgebel suggests that an attitude like aversive racism could be ascribable at the attitudinal level as soon as the

Aversive Racism as an Ambivalent Attitude

On Edouard Machery’s view, to have an attitude is to have a broad-track disposition “to behave and cognize (have thoughts, attend, emote, etc.) toward an object (its formal object) in a way that reflects some preference” (Machery, 2016, p. 112). Machery takes what I call the standard approach for solving meta-theoretical questions about how dispositional constructs and mental states/processes relate to one another: he suggests that attitudes (*qua* dispositions) are underwritten by mental states and processes which serve as the psychological basis of the attitude. For example, the degree to which a person can be described as courageous depends on mental states and processes – including “her moral beliefs (e.g., whether fear is shameful), on the nature of her fear reactions, on the strength of her pride, on her capacity for self-control, etc.” (Machery, 2016, p. 112). Likewise, the degree to which someone can be characterized as being racist depends on mental states and processes – including “moral beliefs (e.g., for most of us the belief that racism is wrong or, for some racists, the belief that racism is right), on non-propositional associations between concepts (e.g., an association between the concept of a black man and the concept of danger), on emotions (e.g., fear when confronted with black men), and on a weak self-control” (Machery, 2016, p. 112).

How does Machery deal with the problem of aversive racism? Aversive racism is a kind of ambivalent attitude, characterized by cognition and behavior that reflects both favorable and unfavorable evaluations of the relevant racial/ethnic group. However, this account denies the possibility of ambivalent attitudes “except perhaps in pathological cases” (Machery, 2016, p. 124). To better motivate and contextualize why Machery might hold this view, I’ll expand the working analogy between the attitudes and traits literatures to talk at the methodological level. As I mentioned before, in personality psychology, the standard approach conceptualizes traits as dispositional constructs underwritten by the mental states/processes that form their psychological basis. Early on, personality theorists characterized these dispositions as *central tendencies* – operationalized as a person’s mean or average tendency to cognize/behave – so as to describe that person as an individual and to describe how she differs from others. When conceptualized in this way, a dispositional construct is more useful and informative the more closely that person’s cognitions and behaviors track some average tendency: it would not be very useful to describe an individual as “extraverted on average” if she fluctuated

relevant folk stereotype has been established (Schwitzgebel, 2013, pp. 94-5). So long as our folk psychological repertoire has not been so changed, Schwitzgebel’s view does not account for and describe aversive racism at the attitudinal level. Even if/when our folk psychological concepts make such a shift, I think there are good reasons to prefer my dispositional view, which adopts a more standard psychological approach for characterizing what lies at the sub-attitudinal level and grounds attitude dispositions in scientifically more satisfying ways.

wildly between extraversion and introversion on different occasions (Fleeson, 2004, pp. 83-4).

I think this is why Machery says of ambivalent attitudes that “[i]f the hypothesized co-referential, differently valenced mental states” lead people to “act and cognize in a way that expresses a positive preference in some contexts” and “a negative preference in other contexts,” then “their *aggregate behavior* cannot be predicted (even imperfectly) by postulating a trait” (Machery, 2016, p. 124, italics mine). The more that aversive racist cognitions and actions vary, the less traction we have, conceptually and methodologically, to attribute attitudes to them. Machery is not alone in adopting this *method of means*: social psychologists who advocate for the existence of one type of attitude disposition also advert to an individual’s mean preference for an object when measuring the strength and direction of their attitude about that object (Eagly & Chaiken, 2007).⁶

Sub-attitudinal versus attitudinal accounts

As we have seen, competing dispositional accounts of attitude can describe aversive racism, but do so by adverting to sub-attitudinal elements rather than appeal to an attitude proper.

Pace these accounts, I will explain how it is conceptually possible for dispositional accounts to characterize ambivalent attitudes such as aversive racism. By conceptualizing aversive racism as an attitude, we provide a perspicuous way to speak about how individuals belonging to that category will cognize/ behave in the future, how they are similar to each other, and how they differ from those who hold different attitudes towards the same racial/ethnic group.

A New Dispositional Account

I think that we can address limitations to Machery’s view by drawing from advances in psychology. Like Machery, I will adopt the “standard view” for resolving the dispositional construct versus mental state/process debate. However, I will advocate for a finer-grained method for characterizing dispositional attitudes.⁷

⁶ Some social psychologists who advocate for mental state views also take the means approach, whereby “attitudes are defined as *summary evaluations*” (Fazio, 2007, p. 608, italics mine).

⁷ Because of hard limits on space, I will not be able to provide an argument for dispositional accounts of attitude over mental state accounts here. Schwitzgebel favors his dispositional account over more traditional mental state approaches because his view, unlike all-or-nothing mental state approaches, could characterize in-between attitudes at the sub-attitudinal level (Schwitzgebel, 2002, 2001, 2010). However, mental state theorists have since adopted more complex views to accommodate in-between cases (Gendler 2008b; Mandelbaum, 2015; Levy, 2015). Machery uses an inference to best explanation to argue for a dispositional approach over a mental state approach (Machery, 2016). However, he and I disagree about what inferences should be drawn from the mixed psychometric evidence (Lee, forthcoming). I think that there are other reasons for preferring a dispositional account, including its

To get us started with the basic idea, let’s consider a stylized example from the domain of moods. Let’s imagine we are trying to characterize the emotional lives of two individuals. We ask them to report their mood along a single scale that ranges from the negative range (sad) to the positive (happy) many times over a multi-week period. Imagine that the two individuals turn out to have the same mean/average mood (which lies mid-way on our scale). However, one individual is almost always in a neutral emotional state while the other rapidly cycles between being extremely happy and extremely sad in equal amounts. If we were to characterize their emotional lives simply in terms of their mean mood, we would lose crucial information that could be used to capture what distinguishes these distinctive individuals.

This is what happens when we characterize dispositions in terms of mean scores while overlooking information about their full distribution of scores. When we discard information about the distribution of scores, we forgo characterizing and drawing distinctions between finer-grained psychological categories. My proposal is to amend the method of means by characterizing dispositions not simply in terms of their mean tendency but also in terms of their distribution.

William Fleeson proposed and fruitfully applied this methodological injunction to the study of personality traits: personality traits should be distinguished not simply by their means, but also by *distributions* over time and contexts: personalities should be said to differ not simply when their means differ but when their distributions differ (Fleeson, 2001). Characterizing traits in terms of means *and* distributions allows personality psychologists to accommodate within-person variability while still characterizing traits as *stable distributions*, so long as an individual’s mean *and* distribution are stable attributes of that individual. Here, variability in cognition and behavior is its own kind of stable individual-differences characteristic (Fiske, 1961; Larsen, 1989; Murray, 1938). So, if an individual veers wildly between low and high extraversion, but that fluctuation is a stable property of that individual across a suitably large range of times and contexts, we can characterize and distinguish her as a different kind of extravert than the consistently moderate extravert who has the same mean extraversion score.

I propose that we carry over this methodological injunction to the case of attitudes. Conceptualizing attitudes in terms of an individual’s mean and distribution of reflected likings and dislikings gives us an intuitive and powerful way of characterizing attitudes – especially, ambivalent attitudes. To see this, let’s consider another stylized example: how should we characterize the aversive racist versus an individual who holds relatively neutral views about the relevant racial/ethnic group?⁸ Here, the

epistemic modesty, which can have pragmatic benefits from a stakeholder/institutional perspective (Lee, forthcoming).

⁸ In making this comparison, I adopt the working assumption in psychology that neutral attitudes can be distinguished from

neutral individual demonstrates a more consistent indifference, reflecting neither a liking nor a disliking for the racial/ethnic group in their cognition and behavior, including results from direct tests for explicit attitudes and indirect tests for implicit attitudes. In contrast, the aversive racist's preferences fluctuate between a strong liking and a strong disliking of the racial/ethnic group on direct and indirect tests. My approach allows us to distinguish the attitudes held by the aversive racist and the neutral individual in terms of differences in the distributions of their reflected preferences.

By conceptualizing and measuring attitudes in terms of their full distributions – rather than rely solely on their mode or means as Schwitzgebel and Machery do – my account of dispositional attitudes makes it possible for ambivalent attitudes *qua* attitude to predict aggregate behavior. Here, ambivalent attitudes can describe and predict *trends* in cognition/behavior so long as there is stability in an individual's *distribution* of reflected preferences across a suitably large range of times and contexts. (Ambivalent attitudes do not describe and predict cognition/behavior strongly enough to predict *individual* events – there is too much intra-individual variation for that.) By bestowing descriptive and predictive power to ambivalent attitudes *qua* attitude, my account can thereby describe and ascribe aversive racism at the attitudinal level, and do so while positing only one attitude type.

Conclusion

In this paper, I have motivated and articulated a dispositional account of aversive racism. My account can be distinguished from other dispositional accounts by its ability to characterize aversive racism at the attitudinal level and its deeper appreciation of the analogy between traits and attitudes.

To evaluate the feasibility and fruitfulness of this account, future research will need to address further questions about how to implement this proposed account. First, how are different attitudes towards an object – characterized by means and by distributions – distinguished from each other? For example, on what grounds would we distinguish a category of strong racist from a weak racist from a neutral individual, especially if there is intra-individual variability in scores for each category? Second, how should available direct and indirect measures/tests for attitudes be selected/combined to provide an appropriate sample (and scaling) of attitude scores for characterizing an attitude

ambivalent attitudes by their relative indifference to the object and the stability of that indifference across contexts (Kaplan, 1972; Bell, Esses, & Maio, 1996; Jonas, Broember, & Diehl, 2000). As such, neutral and ambivalent attitudes have different functional characteristics. For example, ambivalent attitudes can be socially flexible in ways that neutral attitudes are not, where “the co-existence of positive and negative components allows people to express their position by putting forward the component that best fit the specific normative context” without being forced to change their “general attitude” (Cavazza & Butera, 2008, p. 2).

toward an object?⁹ These questions should be explored psychometrically, though such evidence may not point to unique solutions, since some approaches may be more suitable for some purposes than others.

Acknowledgments

I am grateful for Royalty Research Fund Grant #A79071 (University of Washington) which afforded research time to complete this article. For illuminating comments and/or conversations, many thanks to Liam Kofi Bright, Michael Brownstein, Anthony Greenwald, Edouard Machery, Jennifer Nagel, Eric Schwitzgebel, Yuchi Shoda, Samuel Wang, the reviewers, and the audience at the 2015 Central American Philosophical Association meeting where I first presented some of these ideas. I am not writing in my capacity as a contractor for the U.S. National Institutes of Health.

References

- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A Handbook of Social Psychology*. Worcester, MA: Clark University Press.
- Bell, D. W., Esses, V. M., & Maio, G. R. (1996). The utility of open-ended measures to assess intergroup ambivalence. *Canadian Journal of Behavioural Science*, 28, 12-18.
- Bobo, L. D. (2001). Racial attitudes and relations at the close of the twentieth century. In N. J. Smelser, W. J. Wilson & F. Mitchell (Eds.), *America Becoming: Racial Trends and Their Consequences*. Washington D.C.: National Academy Press.
- Borsboom, D., Mellenbergh, J. G., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061-1071.
- Breckler, S. J., & Wiggins, E. C. (1989). On defining attitude and attitude theory: *Once more with feeling*. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude Structure and Function*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cavazza, N. & Butera, F. (2008). Bending without breaking: Examining the role of attitudinal ambivalence in resisting persuasive communication. *European Journal of Social Psychology*, 38, 1-15.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology*, 36, 1-52.

⁹ There is some precedent for managing such difficulties by combining existing measures/traits into an overall measure/test. For example, the Luria-Nebraska Neuropsychological Test Battery for measuring brain damage combines together new, adapted, and previously existing scales such as the Wechsler Adult Intelligence Scale (Rust & Golombok, 2009)

- Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition, 25*, 582-602.
- Egan, A. (2008). Seeing and believing: Perception, belief formation and the divided mind. *Philosophical Studies, 140*, 47-63.
- Egan, A. (2011). Comments on Gendler's "The epistemic costs of implicit bias." *Philosophical Studies, 156*, 65-79.
- Epstein, S. (1994). Trait theory as personality theory: Can a part be as great as the whole? *Psychological Inquiry, 5*, 120-122.
- Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition, 25*, 603-637.
- Fiske, D. W. (1961). The inherent variability of behavior. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of Varied Experience*. Homewood, IL: Dorsey Press.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*, 1011-1027.
- Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science, 13*, 83-87.
- Gendler, T. S. (2008a). Alief and belief. *Journal of Philosophy, 105*, 634-663.
- Gendler, T. S. (2008b). Alief in action (and reaction). *Mind & Language, 23*, 552-585.
- Greenwald, A. G., & Nosek, B. A. (2008). Attitudinal dissociation: What does it mean?" In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the New Implicit Measures*. Hillsdale, NJ: Lawrence Erlbaum.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology, 97*, 17-41.
- Holroyd, J. (2016). What do we want from a model of implicit cognition? *Proceedings of the Aristotelian Society, 116*, 153-79.
- Jonas, K., Broemer, P., & Diehl, M. (2000). Attitudinal ambivalence. *European Review of Social Psychology, 11*, 35-74.
- Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin, 77*, 361-72.
- Kovel, J. (1988). *White Racism: A Psychohistory*. London: Free Association Books.
- Krosnick, J. A., Judd, C. A., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *Handbook of Attitudes and Attitude Change*. Mahwah, NJ: Erlbaum.
- Kwong, J. M. C. (2012). Resisting aliefs: Gendler on belief-discordant behaviors. *Philosophical Psychology, 25*, 77-91.
- Larsen, R. J. (1989). A process approach to personality psychology: Utilizing time as a facet of data. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions*. New York: Springer-Verlag.
- Lee, C. J. (2007). The Representation of Judgment Heuristics and the Generality Problem. *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1211-1216). Austin, TX: Cognitive Science Society.
- Lee, C. J. (Forthcoming). Collective implicit attitudes: A stakeholder conception of implicit bias. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*. Madison, WI: Cognitive Science Society.
- Levy, N. (2015). Neither fish nor fowl: Implicit attitudes as patchy endorsements. *Noûs, 49*, 800-823.
- Machery, E. (2016). De-Freuding Implicit Attitudes. In J. Saul & M. Brownstein (Eds.), *Implicit Bias and Philosophy Volume 1: Metaphysics and Epistemology*. Oxford: Oxford University Press.
- Mandelbaum, E. (2013). Against alief. *Philosophical Studies, 165*, 197-211.
- Mandelbaum, E. (2015). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs, 49*, 629-658.
- Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology, 49*, 229-258.
- Murray, H. A. (1938). *Explorations in personality: A clinical and experimental study of fifty men of college age*. New York: Oxford University Press.
- Rust, J., & Golombok, S. (2009). *Modern psychometrics: The science of psychological assessment*. 3rd ed. New York: Routledge.
- Schwitzgebel, E. (2001). In-between believing. *The Philosophical Quarterly, 51*, 76-82.
- Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Noûs, 36*, 249-275.
- Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly, 91*, 531-553.
- Schwitzgebel, E. (2013). A dispositional approach to attitudes: Thinking outside of the belief box. In N. Nottelmann (Ed.), *New Essays on Belief: Constitution, Content, and Structure*. New York: Palgrave MacMillan.
- Shadish Jr., W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Son Hing, L. S., Li, W., and Zanna, M. P. (2002). Inducing hypocrisy to reduce prejudicial response among aversive racists. *Journal of Experimental Social Psychology, 38*, 71-77.