

Review

A 3D framework of implicit attitude change

Benedek Kurdi ^{1,4,*,@} and Tessa E.S. Charlesworth^{2,3}

According to early theories, implicit (automatic) social attitudes are difficult if not impossible to change. Although this view has recently been challenged by research relying on experimental, developmental, and cultural approaches, relevant work remains siloed across research communities. As such, the time is ripe to systematize and integrate disparate (and seemingly contradictory) findings and to identify gaps in existing knowledge. To this end, we introduce a 3D framework classifying research on implicit attitude change by levels of analysis (individual vs. collective), sources of change (experimental, ontogenetic, and cultural), and timescales (short term vs. long term). This 3D framework highlights where evidence for implicit attitude change is more versus less well established and pinpoints directions for future research, including at the intersection of fields.

From impossible to potentially tractable: a brief history of implicit attitude change

Since the 1980s, a simple but consequential idea has defined much **social cognition** (see [Glossary](#)) research [1–4]: When humans encounter social targets (whether individuals or groups of people), information stored in long-term memory about these social targets can be activated relatively **automatically**, and specifically in the absence of (or even contrary to) the person's intention to do so [5] (for alternative conceptualizations, see [6,7]). Customarily measured using indirect procedures, such as sequential priming [2,8], the Implicit Association Test [9], the Affect Misattribution Procedure [10], and their variations, **implicit attitudes** have been shown to play a unique role in how we evaluate and relate to our social environments. In particular, implicit attitudes affect how we treat and make decisions about others belonging to social categories such as age, gender, race, and sexual orientation [11], although to what degree and under what conditions remains a topic of inquiry [12,13].

Inspired by contemporaneous models of semantic memory [14], early theories of implicit social cognition posited that, once acquired, implicit attitudes are fundamentally unchangeable [15], or at least that change in implicit attitudes would require exposure to massive amounts of counter-attitudinal information unlikely to be experienced even over a lifetime, let alone over short periods of time [15–18]. One consequence of this pessimistic view about implicit attitude change was that few, if any, studies directly addressed the possibility of genuine change in implicit attitudes. However, even at this early stage of research, it was understood that implicit attitudes can be temporarily modulated by environmental cues, such as the presence of an experimenter with a stigmatized social identity or the prospect of interacting with someone from a stigmatized group [19]. These early studies are distinct from more recent research reviewed below in that these temporary contextual modulations do not involve providing direct attitude-relevant evaluative information to participants and, as such, are unlikely to be creating new evaluative representations — a hallmark of genuine attitude change.

Over the past decade, the view about the (near) impossibility of change in implicit social attitudes has shifted considerably in response to a wealth of experimental evidence obtained from adult participants [20–24]. Even more important, although early theories were

Highlights

We introduce a 3D framework — consisting of levels of analysis (individual vs. collective); sources of change (experimental, ontogenetic, and cultural); and timescales (short vs. long term) — to systematize and integrate complementary but previously siloed lines of research on implicit attitude change.

Contrary to early theories, the evidence is clear that implicit attitudes can change, both at the level of individuals and collectives, and both in the short and long term.

Using the 3D framework, we identify critical gaps in existing knowledge, especially in areas that cross traditional field boundaries and methodological approaches, such as the study of long-term change at the individual level and short-term change at the collective level.

A central priority will be to understand the role of person-by-situation interactions in producing long-term change within individuals.

We highlight methods that may help bridge across dimensions of the 3D framework, such as agent-based modeling and network science.

¹Department of Psychology, Yale University, New Haven, CT 06511, USA

²Department of Psychology, Harvard University, Cambridge, MA 02138, USA

³Department of Psychology, University of Toronto, Toronto, ON, Canada, M5S 3G3

⁴<http://www.benedekkurdi.com/>

*Correspondence: benedek.kurdi@yale.edu (B. Kurdi).
 @Twitter: @BenedekKurdi (B. Kurdi).

formulated exclusively on the basis of experimental studies measuring short-term malleability in adults, the relevant evidence base has since grown rapidly and in new directions: research on implicit attitudes now encompasses investigations of change over **ontogeny** [25], at the level of collectives rather than only individuals [26–28], and over cultural timeframes of decades and even centuries [29,30]. These developments have been jointly enabled by conceptual advances, such as the introduction of the bias of crowds framework [27], and methodological advances, such as the availability of implicit attitude tests from millions of participants [31–33].

Research on implicit attitude change has now proliferated to such a degree and takes approaches of such methodological diversity (from developmental psychology to computer science to sociology) that little sharing of methods and findings across different research traditions exists. This state of affairs is unfortunate because, at their core, these diverse research perspectives seek to answer the same underlying question: are implicit attitudes capable of changing, and if so, in response to what type(s) of input and via what type(s) of process?

Although the motivating question is the same, so far there have been few if any attempts at integration across approaches, despite their potential to mutually inform and constrain one another. The main goal of this review is to provide such integrative insights into implicit attitude change. To this end, we introduce a 3D framework, which offers conceptual scaffolding to systematize existing findings; identifies commonalities and points of convergence across different approaches; and, most important, highlights key areas that have remained unaddressed or under-addressed in empirical work.

Introducing the 3D framework

The diversity of approaches taken across currently siloed subfields necessitates a conceptual framework from which to understand, organize, and productively integrate the wealth of research on implicit attitude change. We propose three intersecting axes along which to systematize existing findings, uncover gaps in currently available evidence, and identify potential points of synergy across perspectives and methodologies.

The first axis in the 3D framework (Figure 1) represents level of analysis, which can take on two values: individual or collective level. The former considers implicit attitude change within a single person, whereas the latter considers implicit attitude change in groups or aggregates of people (such as organizations, geographic regions, or countries).

The second axis represents sources of attitude change, which can take on three values: experimental (defined as implicit attitude change taking place in response to a controlled intervention); ontogenetic (defined as taking place over the course of an individual's development); and cultural (defined as taking place in response to processes unfolding or events occurring at the macrolevel of societies). Experimental change is controlled and **exogenously** administered to allow for causal inference but can lack generalizability; by contrast, ontogenetic and cultural change unfold naturally and, therefore, have the benefit of ecological validity, paired with the drawback of less clear-cut causal inferences.

Finally, the third axis represents timescale of change, which can take on two values: short and long term. We intentionally do not identify a cutoff that separates short-term from long-term change because the specific length of time denoted by these terms depends, to a large extent, on the level of analysis and source of change. Instead, we define short-term as encompassing instances of transient malleability, such as those occurring over a single experimental session or as a temporary disruption in response to a short-term cultural shock. Long-term change, by contrast, encompasses instances of enduring transformation, such as those identified through repeated measurement of individuals or

Glossary

Aging effects: processes of attitude change reflecting age-related shifts within individuals.

Approach–avoidance training: a procedure designed to change attitudes by having participants (symbolically) move closer to or away from social targets.

Associative processes: learning processes reflecting only the fact that two stimuli co-occur and the number of co-occurrences (but not how they are related to each other).

Attitudes: evaluations of social targets along a positive–negative continuum.

Automaticity: the extent to which a cognitive process is characterized by lack of awareness, lack of intention, efficiency, and lack of controllability.

Cohort effects: differences in mean-level attitudes across populations born at particular times, reflecting unique cultural inputs characterizing those cohorts.

Cohort replacement: processes of mean-level attitude change reflecting the replacement of older generations with younger generations.

Cross-sectional design: research designs relying on comparing multiple groups at one timepoint.

Diagnosticity: the extent to which information reveals a social target's deep underlying character.

Difference in differences: a statistical technique used for causal inference in observational data, capitalizing on differences across both time and a second factor (e.g., geography) that determines exposure to a cultural source of change.

Evaluative conditioning: a procedure to create attitude change by repeated pairings of social targets with intrinsically positive and negative stimuli.

Exogenous: a variable whose value does not depend on other variables in the model.

Implicit: automatically revealed (as opposed to explicit or controlled).

Longitudinal design: research designs relying on measuring the same individuals repeatedly across time.

Natural language processing (NLP): a set of machine learning methods used to analyze large amounts of text data.

Network science: a discipline in which social phenomena are modeled as nodes of a network that are connected to each other via edges.

Dimensions of implicit attitude change

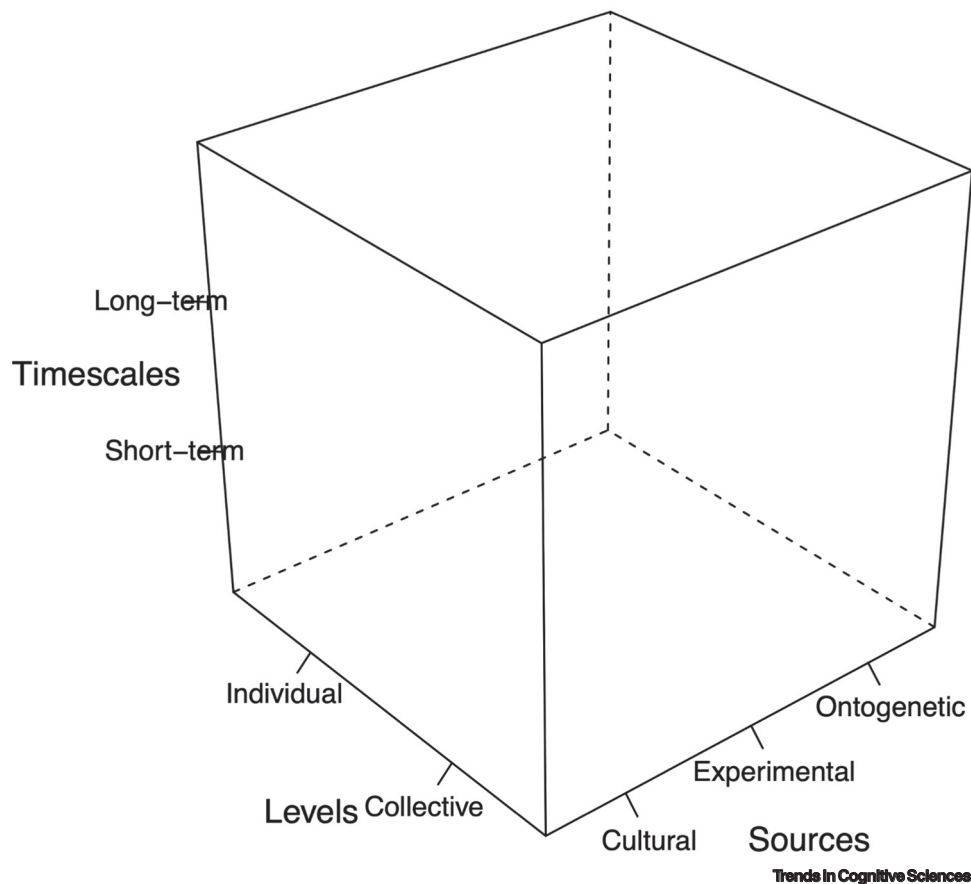


Figure 1. Dimensions of implicit attitude change in the 3D framework. The dimensions include levels of analysis (individual vs. collective); sources of change (experimental, ontogenetic, and cultural); and timescales (short vs. long term).

collectives across extended periods. We note that this axis of the framework refers to the duration of attitude change rather than the duration of the intervention giving rise to it; whether the duration of change is necessarily commensurate with the duration of interventions remains an important open question.

Importantly, the three axes corresponding to levels of analysis, sources of change, and time-scales are represented orthogonally, in line with our belief that the three features should be treated as independent of each other (Figure 1). As such, the 3D framework gives rise to a taxonomy of 12 clusters of studies (Figure 2, Key figure); some of which have been widely explored, and others of which are virtually unpopulated. We next review each of these individual clusters, defined by their placement along the three axes of the 3D framework.

Evidence on implicit attitude change: twelve clusters of research

Within each cluster emerging from the 3D framework, we provide a brief review of existing evidence and identify opportunities for future work. We start with the individual level of analysis, moving from experimental to ontogenetic to cultural sources of change, first discussing short-term and then

Ontogeny: natural development of individuals over their lifespan.

Period effects: processes of change reflecting shared cultural effects, consistent across all age groups and cohorts.

Propositional processes: learning processes reflecting how stimuli are related to each other (e.g., a drug may cause or prevent symptoms, or a social group may oppress or be oppressed by another group).

Quadruple process model: a formal model in which responses on implicit attitude measures are posited to emerge from four cognitive processes.

Reinforcement learning: a set of algorithms designed to maximize cumulative reward over time.

Reinterpretation: attitude change emerging from reconsidering the evaluative implications of already known information.

Social cognition: the study of how humans think about, evaluate, and relate to other humans, both as individuals and as members of social groups.

Test-retest reliability: an index of a measure's consistency over time, obtained by administering it to the same individuals on multiple occasions.

Key Figure

Key Figure. Evidence for implicit attitude change across clusters of the 3D framework

		Sources of change					
		Induced		Naturally occurring			
		Experimental		Ontogenetic		Cultural	
		Short-term change	Long-term change	Short-term change	Long-term change	Short-term change	Long-term change
Levels of analysis	Individual	Strong evidence for change	Some evidence for change with novel targets	Evidence for moderate test-retest reliability	No empirical work available	Little empirical work available	No empirical work available
		Inputs and processes not well understood	No evidence for change with real group targets			Available work shows evidence for change	
	Collective	Short-term change	Long-term change	Short-term change	Long-term change	Short-term change	Long-term change
		No empirical work available	No empirical work available	Little empirical work available Available work shows evidence for change	Strong evidence for stability, with few exceptions	Some evidence for change	Strong evidence for change Inputs and processes not well understood

Trends in Cognitive Sciences

Figure 2. The 3D framework of implicit attitude change – consisting of levels of analysis (individual vs. collective); sources of change (experimental, ontogenetic, and cultural); and timescales (short vs. long term) – results in 12 clusters, which can be used to systematize existing empirical work and to identify gaps in current knowledge. Green background denotes evidence for implicit attitude change; red background denotes evidence for lack of implicit attitude change; and yellow background denotes lack of empirical evidence or unresolved questions.

long-term change within each source. The collective level of analysis comes second and mirrors the same organization. Key takeaways from each cluster are summarized in [Figure 2](#).

Individual level of analysis

Short-term experimental change

At the individual level of analysis, the bulk of relevant studies have focused on experimentally induced short-term malleability in implicit attitudes. Early theories [15–18] posited that once acquired, an individual's implicit attitudes should be resistant to the effects of counter-attitudinal information. Contrary to this view, a wealth of evidence now demonstrates that implicit attitudes are, in fact, malleable in response to a wide range of inputs [34], including **approach–avoidance training** [35], **diagnostic** behavioral information [36–38], **evaluative conditioning** [39], changes in goals [40], and the **reinterpretation** of old information [41,42]. Going forward, both translational and theoretically focused work would benefit from more effectively systematizing the types of input that are more or less effective at inducing malleability in implicit attitudes.

Additionally, the processes mediating such malleability remain relatively poorly understood. To date, the dominant theoretical distinction [21,22,43] has been that between associative and propositional processes ([Box 1](#)). Specifically, early theories posited that implicit attitudes are uniquely malleable via **associative processes** (reflecting, e.g., co-occurrences of Black individuals with violence), to the exclusion of **propositional processes** (reflecting, e.g., relational information about Black individuals being perpetrators vs. victims of violence). Although this distinction has produced a sizable body of empirical work [44], it faces major conceptual and empirical challenges, including the fact that relevant theories are not easily falsifiable.

To facilitate falsifiable predictions, a small but growing number of projects have applied computational modeling tools, including the **quadruple process model** [45,46] and **reinforcement**

Box 1. Associative–propositional debate

The dominant memory models of the 1970s represented semantic knowledge as connections between nodes in large-scale conceptual networks [14]. Such connections, in turn, were thought to be strengthened via co-occurrences of the corresponding stimuli. For example, it was assumed that if bread and butter frequently co-occur in the world, then the concepts bread and butter will share a strong connection in long-term memory. Conceptual relatedness was measured using sequential priming procedures, under the assumption that response latencies in the processing of a target (e.g., 'bread') following a prime (e.g., 'butter') are reflective of the strength of semantic connection between the two [107,108].

When implicit social cognition emerged in the 1980s, the same methods and theoretical assumptions were used to study the organization of social information in long-term memory [1,2,109], including attitudes (e.g., Black–good or White–bad) and stereotypes (e.g., Asian–competent or female–warm). Critically, implicit attitudes were assumed to be represented associatively in the mind (e.g., Black–good), and such semantic associations were thought to shift only very slowly, if at all, and only in response to vast amounts of associative experience [15–18].

This view remained virtually unchallenged until propositional accounts were introduced in the mid-2010s [21,22,104]. These accounts posit that, instead of associative representations, implicit attitudes are mediated by automatically retrieved propositions (e.g., 'Black people are good'), which reflect not only long-term co-occurrences of stimuli in the environment but also how stimuli are related to each other. For example, on the associative view, a drug that prevents an illness should be subject to implicit negativity because illnesses are negative in valence; on the propositional view, however, the drug should become positive because something that prevents negative events is usually inferred to be positive.

On balance, meta-analytic evidence favors the propositional view [44], with implicit attitudes often showing flexible updating, including in response to how stimuli are related to each other (e.g., a drug causing vs. preventing an illness). However, important questions remain underexplored, including the following. Can temporary malleability in implicit attitudes induced by propositional influences translate into enduring change? Do propositional influences generalize to real group targets (e.g., race and gender), or are they confined to novel targets? And, finally, do propositional influences emerge under the relatively noisy conditions of the real world, or only in carefully controlled experimental studies where participants are motivated and able to attend to all relevant information?

learning algorithms [47,48], which can mathematically formalize the cognitive mechanisms and processes underlying implicit attitude change. Computational approaches are set to further increase in importance given that future research will require precise, computationally tractable hypotheses about how individual-level learning processes may ultimately give rise to emergent change at the collective level of analysis (see below).

Long-term experimental change

Compared with the wealth of evidence on short-term, experimentally induced change within individuals, there is a surprising dearth of empirical work on whether such changes persist over longer periods of time. Long-term change is of interest for theoretical reasons. It can reveal whether change represents working memory-dependent processes of transient malleability or, instead, the formation of new evaluative representations in long-term memory. Additionally, whether experimentally induced change is durable has obvious relevance to translationally oriented intervention work aimed at alleviating pernicious social group-based biases [49].

A crucial distinction in this cluster can be drawn between studies using novel targets (fictitious groups, such as Laapians and Niffians [50]) versus real group targets (differing along race, gender, and other preexisting social categories). Studies using novel targets have produced solid evidence for the persistence of experimentally induced changes in implicit attitudes over hours [51], days [52,53], and even weeks [54]. However, no convincing parallel evidence exists for real group targets [55,56]. Given this empirical dissociation, a major challenge going forward will be to bridge the gap between studies of novel and real group targets (for more details, see [Box 2](#)).

Short-term ontogenetic change

In line with the experimental work reviewed above, studies of short-term ontogenetic individual change (i.e., naturally occurring, transient changes within individuals measured over delays ranging

Box 2. Implicit attitude change in studies of fictitious and real targets

Given an interest in internally valid inferences about mechanisms of learning and memory, theoretically oriented social cognition work often uses novel targets, including groups such as Laapians and Niffians [50], individuals such as Bob or Francis West [36,42], and even nonsocial stimuli such as nonwords [110], to study implicit attitude change. By contrast, translationally oriented intervention work tends to rely on real group targets, differing along dimensions such as race or gender [34,56], with a focus on effectiveness rather than cognitive mechanisms.

This disconnect is visible in the fact that although much has been learned about inputs to individual-level malleability in implicit attitudes [44], it is unclear whether such insights generalize to real-world targets. Conversely, although debiasing-focused work has made important inroads when it comes to establishing intervention effectiveness [34,56], it remains to be seen why certain interventions are more effective than others. Importantly, novel and real-world stimuli cannot be automatically assumed to be interchangeable with each other [44,111,112].

Perhaps the most important opportunity for synergy between theoretically oriented and debiasing-focused work exists in understanding why short-term malleability in implicit attitudes toward preexisting social targets (race, gender, and other social categories) does not usually translate into enduring, long-term change [55,56]. In line with traditional associative approaches [15–18], one possibility is that changes observed in studies of novel targets [51–54] reflect priming-like processes that persist only as long as information is maintained in working memory. Another important, and currently underexplored, possibility is that, although implicit attitudes are capable of long-term change, such change can be undermined when perceivers encounter reminders of old, negative information in their daily lives.

Preliminary evidence for the latter possibility was provided in a set of studies [102] in which experimentally created implicit negativity toward a novel target was seemingly overturned using diagnostic behavioral information. Although the effects of positive counter-attitudinal information were undiminished following a 48-h delay, initial negativity returned with full force when a mild reminder (in the form of evaluative conditioning) was administered to participants at the second timepoint. These studies suggest that although implicit attitude change can endure in an information vacuum, it can be vulnerable to reminders of the initial, negative information. Using these insights to help produce enduring change in implicit attitudes toward real social group targets remains an important priority for future work.

from a few minutes or hours to a year) usually suggest both some degree of stability and some degree of malleability in implicit attitudes. Meta-analytic estimates of **test–retest reliability**, which provide an indicator of how much a single individual naturally shifts in their implicit attitudes across close timepoints, range from $r = 0.44$ [57] to $r = 0.54$ [58]. These moderate correlations imply that the rank ordering of participants' implicit attitudes is not fully consistent but also not fully unpredictable across measurement occasions; in other words, short-term ontogenetic changes may occur but, at most, to a moderate degree. Indeed, Bayesian computational modeling can yield considerably higher estimates of within-individual temporal stability [59,60], suggesting that the meta-analytic test–retest results above may represent the lower bounds of the relevant population parameter.

Critically, it is currently unknown to what extent imperfect test–retest reliability is due to measurement error or to systematic variance due to the immediate context in which the implicit attitude measure is being administered. In line with the first possibility, implicit attitude measures tend to be more noisy than parallel explicit attitude measures even at a single time point [61]. On the other hand, given ample evidence for experimentally induced malleability in implicit attitudes (see above), the possibility of context-dependent changes also seems reasonable and should receive attention in empirical and conceptual work.

Long-term ontogenetic change

To date, the longest delay at which the test–retest reliability of implicit attitude measures has been investigated is one year – a far cry from humans' average life expectancy of 73 years. Whether enduring, long-term individual-level ontogenetic change in implicit attitudes is possible could be revealed using intensive, massively **longitudinal designs**. However, such studies are yet to be undertaken, in large part due to their resource requirements.

The lack of evidence from longitudinal studies has led to an overreliance on **cross-sectional designs** in the study of the development of implicit attitudes. However, such work is a better source of evidence on aggregate (collective) change than on individual-level change [62], described below. Also missing from the literature are studies that measure the same individuals across multiple occasions while also systematically manipulating the context (e.g., the information to which a person has just been exposed). This type of design could provide crucial insights into whether the imperfect test–retest reliability mentioned above reflects measurement error or systematic, situation-specific variance in implicit attitudes.

Short-term cultural change

The culture in which an individual is embedded, even temporarily, can exert powerful effects on the way in which they relate to others in their social environment. In line with this idea, a handful of studies have demonstrated the effects of language (a central cue to culture) on shifting individuals' implicit attitudes [63,64]. For instance, bilingual participants randomly assigned to complete an Implicit Association Test in Spanish showed stronger pro-Spanish implicit attitudes, while participants assigned to complete the same test in English showed stronger pro-English implicit attitudes.

Moving beyond the single feature of language, a recent study has additionally demonstrated small shifts in implicit attitudes depending on the physical location of test taking — respondents' college town of Montreal versus their respective hometowns [65]. However, the number of studies in this cluster remains small, with much work to be done on several, currently unexplored, aspects of culture that could exert effects on implicit attitudes, including norms, rituals, values, and features of the physical environment [66]. This area is ripe for future theoretical and empirical exploration.

Long-term cultural change

No studies have examined long-term individual-level implicit attitude change induced by cultural events and exposures. This is a major oversight given that the contexts in which humans are embedded are known to exert major long-term effects on virtually every area of psychological functioning [67]. When it comes to methods for studying the long-term effects of culture on individual minds, researchers could look to adjacent social sciences where strong methods already exist for identifying enduring cultural influences on the person (e.g., sociology and economics). For instance, studies might track individuals who move from one location to another to examine whether their attitudes reflect their early or more recent culture [66,68] and to study the temporal dynamics and processes of acculturation.

Collective level of analysis

Short-term experimental change

This cluster of research would include any study in which an experimental intervention (e.g., a counter-attitudinal messaging campaign) is administered to an environment (such as a company, a town, or a school), with an assessment of short-term (transient) changes in aggregate attitudes among groups of individuals. At present, no studies using this type of design exist.

One study [69] investigated how the implicit attitudes of a collective (namely, a university) changed following a short-term experimental intervention, with the data suggesting that collectives had highly stable aggregate attitudes. However, the interventions considered in this study were administered to individual participants within each university. It remains an open question whether interventions directly administered at the collective level may produce different patterns of malleability in aggregate-level implicit attitudes.

To propel future inquiries, researchers might adopt tools from **network science** that can examine how intervening on a system (e.g., by introducing a dissenting minority group) can produce short-term changes throughout the collective [70,71]. Additionally, agent-based modeling [72] would enable researchers to first simulate patterns of malleability in a collective's implicit attitudes based on formalized hypotheses about processes of change and then compare such simulations to empirical observations.

Long-term experimental change

Interventions can also have enduring effects on a collective's aggregate attitudes. However, at present, no research in this cluster exists. This lack of evidence is likely due to the methodological and logistic complexities associated with long-term tracking of cross-sections of participants in environments that received controlled interventions. Future studies could examine, for example, how an organizational policy (such as an initiative to blind resumes during the hiring process) affects employees' aggregate implicit attitudes in that organization across years. Additionally, similar to the previous cluster, network science and agent-based modeling will be helpful in formalizing hypotheses about potential ripple effects of experimental interventions across collectives over the long term.

Short-term ontogenetic change

The central question defining this cluster is whether shared stages in the lifespan (e.g., life events such as starting school or becoming a parent) can prompt temporary changes in the strength of a group's collective implicit attitudes. Such effects would be revealed by naturally occurring differences in implicit attitudes in cross-sections of successive age groups (e.g., 5- vs 6-year-olds). Indeed, at least one study suggests the possibility of such cross-sectional malleability [73]. However, most relevant work has revealed either that little collective ontogenetic change in implicit attitudes exists, with most age groups revealing similar mean levels (see below), or that ontogenetic changes are

monotonic, with enduring age-related shifts [73–75]. In short, there appears to be only rare evidence of short-term ontogenetic shifts in collective-level (cross-sectional) implicit attitudes.

Long-term ontogenetic change

In this cluster, the timescale is expanded to consider mean differences across groups with larger age differences between them. For example, relevant research may compare 6–10-year-olds and adults residing in the same community [76]. As mentioned above, most of the pertinent work has shown evidence of surprising stability, with younger children, older children, and adults all revealing similar average magnitudes in their implicit attitudes [76–80]. At the same time, such conclusions are not ubiquitous, with cross-sectional age-related differences documented in at least a handful of cases [81,82].

Given an almost total lack of evidence on long-term ontogenetic changes in implicit attitudes at the individual level (see above), cross-sectional data are often used to make arguments about developmental processes. However, even a (nonubiquitous) pattern of cross-sectional stability across age groups need not imply stability of implicit attitudes at the individual level. Rather, the similarity between children and adults may be because they are assessed in the same cultural moment and thus reflect similar average cultural experiences [62]. In addition, similar mean levels in implicit attitudes may mask differences in the processes giving rise to such attitudes, for example as domain-general abilities, such as executive function, change across the lifespan [83].

Future research on collective ontogenetic changes would benefit from statistical modeling approaches that seek to partial out **aging effects**, which reflect individual-level processes of change, from **cohort effects** and **period effects**, which reflect differences in cultural experiences [84]. Additionally, future work could expand the scope of inquiry from comparisons of mean-level implicit attitudes across age groups to considering other features of attitudes that may differ across ontogeny, including age-related differences in the processes of attitude formation and responsiveness to interventions (Box 3).

Box 3. Interactions across dimensions of the 3D framework

Most research on implicit attitude change uses a single method to focus on one level of analysis, one timescale, and one source of change in implicit social attitudes and thus can be neatly assigned to one of the 12 clusters of the 3D framework (see Figure 2 in main text). However, there is much promise in considering the interactions of dimensions, especially how results obtained at one level of analysis (e.g., collective) can constrain the interpretation of results at another level of analysis (e.g., individual). Although examples of this approach are rare, here we consider two types of studies that provide inroads.

First, a handful of studies have examined the intersection of collective-level ontogenetic differences with individual-level experimentally induced changes by administering the same experimental manipulation designed either to change existing attitudes [113,114] or to create new attitudes [115,116] to separate samples of children and adults. In general, such studies have provided evidence for cross-level interactions such that the same individual-level experimental intervention may be effective only at some ontogenetic stages but not at others. Specifically, children typically show more limited evidence of implicit attitude formation and change than adults do. For example, children require more verbal scaffolding for successful observational learning (e.g., via evaluative conditioning) and appear to form attitudes only with clear relational information linking social groups to evaluative attributes [116]. Future cross-level work may examine differences in individual-level intervention effectiveness across collectives going beyond age, including cultural backgrounds (such as racial diversity in one's community) and individual-level demographics (such as political ideology).

Second, future research may benefit from asking how collective-level implicit attitude change can emerge from individual-level processes in complex ways that go beyond mere aggregation or summation. Here, agent-based modeling can provide new grounds for discovery. This approach consists of two steps. First, researchers formally specify how individuals (agents) interact with each other and how such interactions give rise to attitude change. Second, they conduct simulations to observe collective-level outcomes following vast numbers of interactions [72]. Agent-based modeling research on explicit (intentional and easily observable) attitudes has started to explore how individual-level processes of social influence can create collective behaviors, such as attitude-based segregation [117], attitude tipping points [70,118], and differences across individualistic versus collectivistic cultures in the speed of change [119]. Going forward, implicit social cognition research should consider how similar processes may unfold when attitudes are specified to be less intentionally activated or less easily observable among agents.

Short-term cultural change

Research in this cluster probes how a collective's implicit attitudes may shift in the short-term following external shocks. For example, studies have shown that aggregate implicit race attitudes in the United States decreased around Black Lives Matter protests [85,86] and implicit body weight attitudes temporarily increased around fat-shaming tweets [87]. Similar short-term inflection points in regional implicit attitudes have been observed around election-related rhetoric [88], the early coronavirus disease 2019 (COVID-19) pandemic [89], and marriage equality legislation [90]. Finally, using new language-based methods to indirectly measure social group representations, research has shown how language-based biases in an organization's communications shifted following the appointment of new female leaders [91].

The unprecedented availability of big data – both from archives of aggregated attitude measures and from archives of collective natural language – has newly expanded this area of inquiry. However, collective-level causal inference about the impact of an event on an attitude trajectory remains relatively new to the cognitive sciences. As such, going forward, tools from sociology and political science will be helpful in moving from purely correlational designs toward methods that can help establish plausible causal influences (e.g., using **difference-in-difference** methods) [92].

Long-term cultural change

A collective's implicit attitudes can also change in enduring ways. For example, implicit pro-straight/anti-gay attitudes in the USA have decreased by as much as 65% over 14 years (from 2007 to 2020), with slower change observed in implicit race and skin-tone attitudes, and stability in body weight, disability, and age attitudes [30,88,93–97]. These results suggest that inputs to cultural change may affect some attitude domains more strongly than others. Additionally, long-term cultural change, when it occurs, seems widespread across demographic groups and places, implying that such change likely arises from interventions directed toward the broadest macrolevel of society, for example, changes in the media representation of lesbian, gay, and bisexual (LGB) individuals [98].

A potential methodological limitation of work in this cluster is that relevant findings have been obtained using nonrepresentative samples. However, such concerns should be allayed by data showing similar trends across these nonrepresentative data and representative surveys, the possibility to re-weight data to approximate census demographics [30], and evidence of relationships between implicit attitudes and meaningful real-world outcomes [28].

Additionally, within this cluster of research, we also place the emerging line of work using **natural language processing (NLP)** approaches, such as word embeddings (elaborated in **Box 4**) examining social representations over even longer time periods of centuries [29,99–101]. Ultimately, despite increasing evidence that long-term change in collective-level implicit attitudes is possible, questions remain regarding the processes and inputs giving rise to such enduring change. Future research would benefit from quasi-experimental methods tracking the time-locked long-term relationships between large-scale inputs (e.g., changes in demographics or media representation) and large-scale changes in attitudes.

Concluding remarks

In this review, we introduced a 3D framework with the goal to systematize empirical evidence on implicit attitude change across levels of analysis, sources of change, and timescales. In doing so, we identified both what is now well established in the literature on implicit attitude change and, equally important, highlighted major gaps of knowledge to spur new research directions.

Outstanding questions

Are there differences in processes of attitude formation versus updating across sources of change, levels of analysis, or timescales?

How can short-term malleability in implicit attitudes at the individual level translate into long-term change?

Are instances of contextual modulation and genuine change in implicit attitudes distinct? Or can repeated and sufficiently generalizable contextual modulation eventually create genuine implicit attitude change?

What are the contributions of measurement error and systematic contextual influences to imperfect test-retest reliability in implicit attitude measures?

How much do implicit attitudes durably change within individuals over ontogenesis?

Can culture induce enduring changes in individuals' implicit attitudes? If so, how quickly and via what processes?

Do multiple short-term cultural events (such as elections) accumulate to create enduring changes in a collective's implicit attitudes? Or, do enduring changes require a fundamentally different type of input (such as changes in media representation)?

Do short-term inputs lead to short-term changes in implicit attitudes and long-term inputs to long-term changes? Or may long-term changes be achievable even via short-term inputs?

Do interventions administered at the individual level (e.g., bias education) aggregate up to create emergent patterns of collective change? Do interventions administered at the collective level (e.g., hiring policies) trickle down to create changes in individuals' implicit attitudes? If so, how?

How much of the apparent cross-sectional stability in implicit attitudes is due to genuine ontogenetic stability vs. similarity in cultural inputs?

What are the most effective experimental interventions to induce short- and long-term shifts in aggregate implicit attitudes?

Box 4. Natural language processing in the study of implicit attitude change

Until recently, implicit attitude research relied exclusively on data collected from individual participants in the lab or online. However, the possibilities of relevant research expanded massively when NLP tools were used for the first time to extract relatively hidden social biases from patterns of language use across more than 800 billion words from the Internet [120]. Since this initial study, the adoption of such NLP tools, especially word embeddings (compressed quantitative indicators of word co-occurrence information), has seen rapid progress throughout psychology and especially in the study of social group representations [121,122].

Given our present focus, two questions arise regarding the ability of NLP tools to advance understanding of implicit attitude change. First, to what extent do such methods capture implicit processes? Language possesses features of both relatively explicit representations (to a large extent, speakers communicate what they communicate intentionally) and relatively implicit representations (word embeddings reflect conceptual associations indirectly computed from word co-occurrences that may well reflect unintentional processes) [29,121]. Thus, although the implicit versus explicit nature of word embeddings is being actively discussed, NLP arguably provides a window into unintentionally expressed (implicit) social group representations.

Second, how can such tools then be used to reveal implicit attitude change and along what dimensions of the 3D framework? To date, NLP has been used to study long-term collective cultural change [29,100] and long-term collective ontogenetic change [123]. Both lines of work have provided evidence for the possibility of implicit attitude change, along with meaningful variability in the extent of such change. For example, representations of racial and ethnic groups seem to be relatively malleable, whereas some gender stereotypes appear relatively recalcitrant.

Going forward, we are excited to see new applications of NLP tools across all clusters emerging from the 3D framework. For instance, NLP models trained on text data across shorter successive timescales could be used to reveal collective short-term change in implicit attitudes [124]. At the individual level, NLP tools could also be used to probe how individuals shift in their language use (for instance, the valence or sentiment of their language when communicating about a social group) following an experimental intervention [125]. In summary, while the use of NLP tools to study implicit attitude change is in its early stages, such tools have enormous potential to produce novel insights going forward.

Contrary to early theoretical ideas about the recalcitrance of implicit attitudes in the face of new information [15–18], implicit attitudes seem amenable to change both in the short term and in the long term, as one would expect from a cognitive system evolved to drive adaptive, automatic responding to social targets [22]. The two clusters of studies that seem to deviate from this overall pattern of change are cross-sectional studies of long-term developmental change (in which children's mean-level attitudes are generally found to mirror those of adults) and experimental studies of individual-level long-term change involving real group targets (in which no compelling evidence of enduring change exists, unlike for novel targets).

We believe that, in both cases, the evidence for stability likely reflects features of the information available to participants rather than cognitive constraints inherent to social information processing. In the former case, apparent stability may be due to shared cultural influences across age groups [62] and in the latter, due to persistent reminders of initial negativity encountered by participants following counter-attitudinal interventions [102]. Indeed, given evidence that cultural-level shifts in implicit attitudes are not exclusively due to **cohort replacement**, but at least partly attributable to changes likely occurring within individuals [30,88], it seems probable that, given appropriate longitudinal and repeated-intervention designs, research will produce persuasive evidence for long-term experimental and developmental change even within individuals.

Beyond these open questions, much else remains unknown about when, how, and why implicit attitudes do or do not reveal change (for a list of future research directions, see [Outstanding questions](#)). In [Figure 2](#) we use yellow backgrounds to mark those clusters in which little to no empirical evidence is available. Understandably given methodological constraints and existing research traditions, the bulk of evidence is concentrated in two clusters where the level of analysis, source of change, and timescale are aligned with each other: individual-level studies focusing on short-term experimental change and collective-level studies focusing on long-term

What are the similarities and differences between processes of change affecting implicit and explicit attitudes, especially at the collective level and in response to ontogenetic and cultural inputs?

cultural change. However, a wealth of methods is now available to conduct work in the remaining clusters that exist at the off-diagonals of research traditions; we hope that the field will take up this challenge.

Even in the clusters where compelling evidence for implicit attitude change exists, the inputs to, and especially processes of, change continue to be relatively poorly understood. This is true even for experimental studies of individual-level change – the area where theories are most well developed – (Boxes 1 and 2), but even more so for studies of long-term cultural change at the collective level where mechanistic understanding of why change occurs in some cases and not in others is almost completely absent. In this context, it will be important to probe how inputs to change might be similar or different between implicit (unintentional) and explicit (intentional) attitudes, especially beyond the relatively limited domain of experimental studies of short-term change within individuals, which has been the sole focus of relevant accounts [43,103,104].

Notably, when it comes to such theory development, it should not be assumed that hypotheses formulated at one level of analysis will necessarily generalize to another [105]. The few available theories that have attempted to jointly consider both individual and collective levels of change have treated the two as antagonistic to each other, asking whether change in implicit attitudes is best explained at the level of individuals or collectives [27,69]. We look forward to a new generation of accounts that expressly consider interactions across levels of analysis (e.g., person-by-situation interactions), and how evidence obtained at one level of analysis can help constrain and inform the interpretation of evidence at another level [106] (Box 3).

Declaration of interest

Benedek Kurdi is a member of the Scientific Advisory Board of Project Implicit, a 501(c)(3) nonprofit organization and international collaborative of researchers who are interested in implicit social cognition.

References

- Devine, P.G. (1989) Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18
- Fazio, R.H. *et al.* (1986) On the automatic activation of attitudes. *J. Pers. Soc. Psychol.* 50, 229–238
- Bargh, J.A. (1989) Conditional automaticity: varieties of automatic influence in social perception and cognition. In *Unintended Thought* (Uleman, J.S. and Bargh, J.A., eds), pp. 3–51, Guilford Press
- Greenwald, A.G. and Banaji, M.R. (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102, 4–27
- Melnikoff, D.E. and Kurdi, B. (2022) What implicit measures of bias can do. *Psychol. Inq.* 33, 185–192
- Fazio, R.H. and Olson, M.A. (2003) Implicit measures in social cognition research: their meaning and use. *Annu. Rev. Psychol.* 54, 297–327
- Gawronski, B. *et al.* (2022) Implicit bias ≠ bias on implicit measures. *Psychol. Inq.* 33, 139–155
- Banaji, M.R. and Hardin, C.D. (1996) Automatic stereotyping. *Psychol. Sci.* 7, 136–141
- Greenwald, A.G. *et al.* (1998) Measuring individual differences in implicit cognition: the Implicit Association Test. *J. Pers. Soc. Psychol.* 74, 1464–1480
- Payne, B.K. *et al.* (2005) An inkblot for attitudes: affect misattribution as implicit measurement. *J. Pers. Soc. Psychol.* 89, 277–293
- Kurdi, B. *et al.* (2019) Relationship between the Implicit Association Test and intergroup behavior: a meta-analysis. *Am. Psychol.* 74, 569–586
- Meissner, F. *et al.* (2019) Predicting behavior with implicit measures: disillusioning findings, reasonable explanations, and sophisticated solutions. *Front. Psychol.* 10, 2483
- Greenwald, A.G. *et al.* (2022) Implicit-bias remedies: treating discriminatory bias as a public-health problem. *Psychol. Sci. Public Interest* 23, 7–40
- Collins, A.M. and Loftus, E.F. (1975) A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428
- Bargh, J.A. (1999) The cognitive monster: the case against the controllability of automatic stereotype effects. In *Dual-Process Theories in Social Psychology* (Chaiken, S. and Trope, Y., eds), pp. 361–382, Guilford Press
- McConnell, A.R. and Rydell, R.J. (2014) The systems of evaluation model: a dual-systems approach to attitudes. In *Dual-Process Theories of the Social Mind* (Sherman, J.W. *et al.*, eds), pp. 204–217, Guilford Press
- Strack, F. and Deutsch, R. (2004) Reflective and impulsive determinants of social behavior. *Personal. Soc. Psychol. Rev.* 8, 220–247
- Smith, E.R. and DeCoster, J. (2000) Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Personal. Soc. Psychol. Rev.* 4, 108–131
- Blair, I.V. (2002) The malleability of automatic stereotypes and prejudice. *Personal. Soc. Psychol. Rev.* 6, 242–261
- De Houwer, J. *et al.* (2020) Attitudes beyond associations: on the role of propositional representations in stimulus evaluation. *Adv. Exp. Soc. Psychol.* 61, 127–183
- De Houwer, J. *et al.* (2021) Attitudes as propositional representations. *Trends Cogn. Sci.* 25, 870–882
- Kurdi, B. and Dunham, Y. (2020) Propositional accounts of implicit evaluation: taking stock and looking ahead. *Soc. Cogn.* 38, s42–s67
- Cone, J. *et al.* (2017) Changing our implicit minds: how, when, and why implicit evaluations can be rapidly revised. *Adv. Exp. Soc. Psychol.* 56, 131–199

24. Ferguson, M.J. *et al.* (2019) When and how implicit first impressions can be updated. *Curr. Dir. Psychol. Sci.* 28, 331–336
25. Dunham, Y. and Banaji, M.R. (2010) Platonic blindness and the challenge of understanding context. In *The Mind in Context* (Mesquita, B. *et al.*, eds), pp. 201–213
26. Payne, B.K. and Hannay, J.W. (2021) Implicit bias reflects systemic racism. *Trends Cogn. Sci.* 25, 927–936
27. Payne, B.K. *et al.* (2017) The bias of crowds: how implicit bias bridges personal and systemic prejudice. *Psychol. Inq.* 28, 233–248
28. Calanchini, J. *et al.* (2022) Regional intergroup bias. *Adv. Exp. Soc. Psychol.* 66, 281–337
29. Charlesworth, T.E.S. *et al.* (2022) Historical representations of social groups across 200 years of word embeddings from Google Books. *Proc. Natl. Acad. Sci.* 119, e2121798119
30. Charlesworth, T.E.S. and Banaji, M.R. (2019) Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychol. Sci.* 30, 174–192
31. Charlesworth, T.E.S. *et al.* (2022) The Project Implicit International Dataset: measuring implicit and explicit social group attitudes and stereotypes across 34 countries (2009–2019). *Behav. Res. Methods* 55, 1413–1440
32. Nosek, B.A. *et al.* (2007) Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18, 36–88
33. Ratliff, K.A. *et al.* (2020) Documenting bias from 2007–2015: pervasiveness and correlates of implicit attitudes and stereotypes II. *PsyArXiv* Published online February 28, 2020. <https://osf.io/jeyc7>
34. Lai, C.K. *et al.* (2014) Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *J. Exp. Psychol. Gen.* 143, 1765–1785
35. Van Dessel, P. *et al.* (2020) On the effectiveness of approach-avoidance instructions and training for changing evaluations of social groups. *J. Pers. Soc. Psychol.* 119, e1–e14
36. Cone, J. and Ferguson, M.J. (2015) He did what? The role of diagnosticity in revising implicit evaluations. *J. Pers. Soc. Psychol.* 108, 37–57
37. Van Dessel, P. *et al.* (2019) Changing deep-rooted implicit evaluation in the blink of an eye: Negative verbal information shifts automatic liking of Gandhi. *Soc. Psychol. Personal. Sci.* 10, 266–273
38. Shen, X. *et al.* (2020) Beware a dishonest face? Updating face-based implicit impressions using diagnostic behavioral information. *J. Exp. Soc. Psychol.* 86, 103888
39. Moran, T. *et al.* (2020) Can (instructions about) stimulus pairings influence automatic and self-reported evaluations in the presence of more diagnostic evaluative information? *Pers. Soc. Psychol. B* 47, 1249–1263
40. Melnikoff, D.E. and Bailey, A.H. (2018) Preferences for moral vs. immoral traits in others are conditional. *Proc. Natl. Acad. Sci.* 115, E592–E600
41. Kurdi, B. *et al.* (2021) Persuading the implicit mind: changing negative implicit evaluations with an 8-minute podcast. *Soc. Psychol. Pers. Sci.* 13, 688–697
42. Mann, T.C. and Ferguson, M.J. (2015) Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *J. Pers. Soc. Psychol.* 108, 823–849
43. Gawronski, B. and Bodenhausen, G.V. (2006) Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychol. Bull.* 132, 692–731
44. Kurdi, B. *et al.* (2022) How do explicit and implicit evaluations shift? A preregistered meta-analysis of the effects of co-occurrence and relational information. *J. Pers. Soc. Psychol.* 124, 1174–1202
45. Calanchini, J. *et al.* (2020) Reducing implicit racial preferences: III. A process-level examination of changes in implicit preferences. *J. Pers. Soc. Psychol.* 121, 796–818
46. Cone, J. and Calanchini, J. (2020) A process dissociation model of implicit rapid revision in response to diagnostic revelations. *Personal. Soc. Psychol. Bull.* 47, 201–215
47. Kurdi, B. *et al.* (2019) Model-free and model-based learning processes in the updating of explicit and implicit evaluations. *Proc. Natl. Acad. Sci.* 116, 6035–6044
48. Stillerman, B. *et al.* (2022) Societal stereotypes shape learning to produce group-based preferences. *PsyArXiv* Published online May 20, 2022. <https://doi.org/10.31234/osf.io/mwztc>
49. Paluck, E.L. *et al.* (2021) Prejudice reduction: progress and challenges. *Annu. Rev. Psychol.* 72, 533–560
50. Gregg, A.P. *et al.* (2006) Easier done than undone: asymmetry in the malleability of implicit preferences. *J. Pers. Soc. Psychol.* 90, 1–20
51. Kurdi, B. and Banaji, M.R. (2019) Attitude change via repeated evaluative pairings versus evaluative statements: Shared and unique features. *J. Pers. Soc. Psychol.* 116, 681–703
52. Mann, T.C. *et al.* (2020) How effectively can implicit evaluations be updated? Using evaluative statements after aversive repeated evaluative pairings. *J. Exp. Psychol. Gen.* 149, 1169–1192
53. Mann, T.C. and Ferguson, M.J. (2017) Reversing implicit first impressions through reinterpretation after a two-day delay. *J. Exp. Soc. Psychol.* 68, 122–127
54. Cone, J. *et al.* (2021) The long-term effects of new evidence on implicit impressions of other people. *Psychol. Sci.* 32, 173–188
55. Lai, C.K. *et al.* (2016) Reducing implicit racial preferences: II. Intervention effectiveness across time. *J. Exp. Psychol. Gen.* 145, 1001–1016
56. Forscher, P.S. *et al.* (2017) Breaking the prejudice habit: mechanisms, timecourse, and longevity. *J. Exp. Soc. Psychol.* 72, 133–146
57. Greenwald, A.G. and Lai, C.K. (2020) Implicit social cognition. *Annu. Rev. Psychol.* 71, 419–445
58. Gawronski, B. *et al.* (2017) Temporal stability of implicit and explicit measures. *Personal. Soc. Psychol. Bull.* 43, 300–312
59. Haines, N. *et al.* (2022) Theoretically informed generative models can advance the psychological and brain sciences: Lessons from the reliability paradox. *PsyArXiv* Published online August 22, 2022. <https://doi.org/10.31234/osf.io/xr7y3>
60. Kvam, P.D. *et al.* (2023) Improving the reliability and validity of the IAT with a dynamic model driven by similarity. *PsyArXiv* Published online March 15, 2023. <https://psyarxiv.com/ke7cp/>
61. Bar-Anan, Y. and Nosek, B.A. (2014) A comparative investigation of seven indirect attitude measures. *Behav. Res. Methods* 46, 668–688
62. Baron, A.S. (2015) Constraints on the development of implicit intergroup attitudes. *Child Dev. Perspect.* 9, 50–54
63. Ogunnaik, O. *et al.* (2010) The language of implicit preferences. *J. Exp. Soc. Psychol.* 46, 999–1003
64. Danziger, S. and Ward, R. (2009) Language changes implicit associations between ethnic groups and evaluation in bilinguals. *Psychol. Sci.* 21, 799–800
65. Oforu, E.K. *et al.* (2022) Geographic regions exert a small effect on prejudice. In *Proceedings of the SPSP 2022 Annual Convention, San Diego, CA, USA*
66. Peterson, R.A. (1979) Revitalizing the culture concept. *Annu. Rev. Sociol.* 5, 137–166
67. Chopik, W.J. and Kitayama, S. (2018) Personality change across the life span: Insights from a cross-cultural, longitudinal study. *J. Pers.* 86, 508–521
68. Finkelstein, A. *et al.* (2016) Sources of geographic variation in health care: evidence from patient migration. *Q. J. Econ.* 131, 1681–1726
69. Vuletic, H.A. and Payne, B.K. (2019) Stability and change in implicit bias. *Psychol. Sci.* 30, 854–862
70. Centola, D. *et al.* (2018) Experimental evidence for tipping points in social convention. *Science* 360, 1116–1119
71. Centola, D. *et al.* (2021) The reduction of race and gender bias in clinical treatment recommendations using clinician peer networks in an experimental setting. *Nat. Commun.* 12, 6585
72. Jackson, J.C. *et al.* (2017) Agent-based modeling. *Soc. Psychol. Pers. Sci.* 8, 387–395
73. Qian, M.K. *et al.* (2019) Differential developmental courses of implicit and explicit biases for different other-race classes. *Dev. Psychol.* 55, 1440–1452
74. Qian, M.K. *et al.* (2015) Implicit racial biases in preschool children and adults from Asia and Africa. *Child Dev.* 87, 285–296
75. Qian, M. *et al.* (2021) Age-related differences in implicit and explicit racial biases in Cameroonians. *Dev. Psychol.* 57, 386–396
76. Baron, A.S. and Banaji, M.R. (2006) The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychol. Sci.* 17, 53–58

77. Charlesworth, T.E.S. and Banaji, M.R. (2022) The development of social group cognition. In *The Oxford Handbook of Social Cognition* (Carlston, D.E. et al., eds), Oxford University Press
78. Dunham, Y. et al. (2008) The development of implicit intergroup cognition. *Trends Cogn. Sci.* 12, 248–253
79. Dunham, Y. et al. (2013) Two signatures of implicit intergroup attitudes: developmental invariance and early enculturation. *Psychol. Sci.* 24, 1–27
80. Steele, J.R. et al. (2018) A cross-cultural investigation of children's implicit attitudes toward White and Black racial outgroups. *Dev. Sci.* 21, e12673
81. Williams, A. and Steele, J.R. (2017) Examining children's implicit racial attitudes using exemplar and category-based measures. *Child Dev.* 90, e322–e338
82. Perszyk, D.R. et al. (2019) Bias at the intersection of race and gender: evidence from preschool-aged children. *Dev. Sci.* 22, e12788
83. Gonsalkorale, K. et al. (2013) Measures of implicit attitudes may conceal differences in implicit associations. *Soc. Psychol. Pers. Sci.* 5, 271–278
84. Yang, Y. and Land, K.C. (2013) *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*, Taylor & Francis
85. Sawyer, J. and Gampa, A. (2018) Implicit and explicit racial attitudes changed during Black Lives Matter. *Pers. Soc. Psychol. B* 44, 1039–1059
86. Sawyer, J.E. and Gampa, A. (2023) Social movements as parsimonious explanations for implicit and explicit attitude change. *Personal. Soc. Psychol. Rev.* 27, 28–51
87. Ravary, A. et al. (2019) Shaping the body politic: mass media fat-shaming affects implicit anti-fat attitudes. *Pers. Soc. Psychol. B* 45, 1580–1589
88. Charlesworth, T.E.S. and Banaji, M.R. (2022) Patterns of implicit and explicit attitudes: IV. Change and stability from 2007 to 2020. *Psychol. Sci.* 33, 1347–1371
89. Darling-Hammond, S. et al. (2020) After 'the China virus' went viral: racially charged coronavirus coverage and trends in bias against Asian Americans. *Health Educ. Behav.* 47, 870–879
90. Ofosu, E.K. et al. (2019) Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proc. Natl. Acad. Sci.* 116, 8846–8851
91. Lawson, M.A. et al. (2022) Hiring women into senior leadership positions is associated with a reduction in gender stereotypes in organizational language. *Proc. Natl. Acad. Sci.* 119, e2026443119
92. Wing, C. et al. (2018) Designing difference in difference studies: best practices for public health policy research. *Annu. Rev. Public Health* 39, 453–469
93. Charlesworth, T.E.S. and Banaji, M.R. (2021) Patterns of implicit and explicit attitudes II. Long-term change and stability, regardless of group membership. *Am. Psychol.* 76, 851–869
94. Charlesworth, T.E.S. and Banaji, M.R. (2021) Patterns of implicit and explicit stereotypes III: Long-term change in gender stereotypes. *Soc. Psychol. Personal. Sci.* 13, 14–26
95. Schmidt, K. and Axt, J.R. (2016) Implicit and explicit attitudes toward African Americans and Barack Obama did not substantively change during Obama's presidency. *Soc. Cogn.* 34, 559–588
96. Schmidt, K. and Nosek, B.A. (2010) Implicit (and explicit) racial attitudes barely changed during Barack Obama's presidential campaign and early presidency. *J. Exp. Soc. Psychol.* 46, 308–314
97. Westgate, E.C. et al. (2015) Implicit preferences for straight people over lesbian women and gay men weakened from 2006 to 2013. *Collabra* 1, 1–10
98. Ayoub, P.M. and Garretson, J. (2017) Getting the message out: media context and global changes in attitudes toward homosexuality. *Comp. Polit. Stud.* 50, 1055–1085
99. Bhatia, N. and Bhatia, S. (2021) Changes in gender stereotypes over time: a computational analysis. *Psychol. Women Q.* 45, 106–125
100. Garg, N. et al. (2018) Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc. Natl. Acad. Sci.* 115, E3635–E3644
101. Jones, J.J. et al. (2020) Stereotypical gender associations in language have decreased over time. *Sociol. Sci.* 7, 1–35
102. Kurdi, B. et al. (2023) The prospect of long-term implicit attitude change: tests of spontaneous recovery and reinstatement. *PsyArXiv* Published online April 29, 2023. <https://doi.org/10.31234/osf.io/mwfwah>
103. Rydell, R.J. and McConnell, A.R. (2006) Understanding implicit and explicit attitude change: a systems of reasoning analysis. *J. Pers. Soc. Psychol.* 91, 995–1008
104. De Houwer, J. (2014) A propositional model of implicit evaluation. *Soc. Pers. Psychol. Compass* 8, 342–353
105. Piantadosi, S. et al. (1988) The ecological fallacy. *Am. J. Epidemiol.* 127, 893–904
106. Hatzenbuehler, M.L. et al. (2014) protective school climates and reduced risk for suicide ideation in sexual minority youths. *Am. J. Public Health* 104, 279–286
107. Meyer, D.E. and Schvaneveldt, R.W. (1971) Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *J. Exp. Psychol.* 90, 227–234
108. Neely, J.H. (1976) Semantic priming and retrieval from lexical memory: evidence for facilitatory and inhibitory processes. *Mem. Cogn.* 4, 648–654
109. Gaertner, S.L. and McLaughlin, J.P. (1983) Racial stereotypes: associations and ascriptions of positive and negative characteristics. *Soc. Psychol. Q.* 46, 23–30
110. Smith, C.T. et al. (2020) The impact of instruction- and experience-based evaluative learning on IAT performance: a quad model perspective. *Cogn. Emot.* 34, 21–41
111. Kurdi, B. and Banaji, M.R. (2022) Implicit person memory: domain-general and domain-specific processes of learning and change. In *The Handbook of Impression Formation* (Moskowitz, G.B. and Baletis, E., eds), pp. 459–488, Routledge
112. Faure, R. et al. (2020) The case for studying implicit social cognition in close relationships. *Soc. Cogn.* 38, 98–114
113. Gonzalez, A.M. et al. (2017) Reducing children's implicit racial bias through exposure to positive out-group exemplars. *Child Dev.* 88, 123–130
114. Gonzalez, A.M. et al. (2021) Developmental differences in the malleability of implicit racial bias following exposure to counterstereotypical exemplars. *Dev. Psychol.* 57, 102–113
115. Gonzalez, A.M. et al. (2017) Malleability of implicit associations across development. *Dev. Sci.* 20, e12481
116. Charlesworth, T.E.S. et al. (2020) Children's implicit attitude acquisition: evaluative statements succeed, repeated pairings fail. *Dev. Sci.* 23, e12911
117. Harton, H.C. and Bourgeois, M.J. (2003) Cultural elements emerge from dynamic social impact. In *The Psychological Foundations of Culture* (Schaller, M. and Crandall, C.S., eds), pp. 41–75, Lawrence Erlbaum Associates
118. De, S. et al. (2018) Tipping points for norm change in human cultures. *ArXiv* Published online July 2, 2018. <https://arxiv.org/abs/1804.07406>
119. Muthukrishna, M. and Schaller, M. (2020) Are collectivistic cultures more prone to rapid transformation? Computational models of cross-cultural differences, social network structure, dynamic social influence, and cultural change. *Personal. Soc. Psychol. Rev.* 24, 103–120
120. Caliskan, A. et al. (2017) Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 183–186
121. Charlesworth, T.E.S. and Banaji, M.R. (2022) Word embeddings reveal social group attitudes and stereotypes in large language corpora. In *Handbook of Language Analysis in Psychology* (Dehghani, M. and Boyd, R.L., eds), pp. 594–608, Guilford Publications
122. Jackson, J.C. et al. (2022) From text to thought: how analyzing language can advance psychological science. *Perspect. Psychol. Sci.* 17, 805–826
123. Charlesworth, T.E.S. et al. (2021) Gender stereotypes in natural language: word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychol. Sci.* 32, 218–240
124. Ashokkumar, A. and Pennebaker, J.W. (2021) Social media conversations reveal large psychological shifts caused by COVID-19's onset across U.S. cities. *Sci. Adv.* 7, eabg7843
125. Rocklage, M.D. et al. (2017) Persuasion, emotion, and language: the intent to persuade transforms language via emotionality. *Psychol. Sci.* 29, 749–760