**Perceived Authenticity as a Justification for Prejudice**

**A Brief History of the Psychological Study of Authenticity**

Philosophers have been interested in the concept of being authentic or true to oneself for thousands of years, but the psychological study of authenticity was first embraced by humanistic psychologists (Kernis & Goldman, 2004). In this tradition, authenticity was vaguely defined as living in accordance with one’s true, core, inner self. These psychologists saw the inner core of people as being fundamentally good and living in accordance to this core to be instinctual, so they saw believed the job of the therapist was to facilitate clients living in accordance with their inner selves. Many ideas in the humanist tradition were treated synonymously with “authenticity,” even if they were not always labeled as such.

*Congruence* was a fundamental concept to the way Carl Rogers saw psychotherapy. The idea was to align how a client thinks about themselves with their ideal version of themselves. Rogers (1961) quoted Kierkegaard to exemplify the goal he set for his clients: “to be that self which one really is.” Although he did not explicitly call this authenticity, it matches with contemporary definitions of the term, and he places centrality on closely-related concepts like being genuine, true, real, and not “putting up facades” to others.

One of the central concepts in Abraham Maslow’s work was *self-actualizing*, which he defined as an “acceptance and expression of the inner core or self” (Maslow, 1968, p. 197). Again, this is similar to modern notions of authenticity; he did define an authentic person as one who resists influences to deviate from their inner core.

Sidney Jourard defined authenticity as something fundamentally social: He argued that we learn at a young age, through punishments and rules, to suppress our true feelings. If done chronically and inappropriately, this leads to negative psychological functioning. He saw *self-disclosure* of the real self to others—being *transparent*—as vital for well-being. He defines the authentic person as one who is “being oneself, honestly, in one’s relations with his [sic] fellows” (Jourard, 1964; p. 153). Much of his work reiterates that one of the main goals of a therapist is to help the client live an authentic life, for the client to let people see them how they see themselves, to not allow one’s true self to be suppressed by external pressures.

These humanistic ideas about authenticity were not subjected to many empirical tests, however, partly due to the elusiveness of determining what the “true self” that one was meant to be authentic to. Indeed, “it is unclear what the self is being true to, when it is being true to itself” (Sheldon, 2009, p. 75). Positive psychologists (Sheldon & King, 2001) study authenticity by doing away with the idea that there is an *objective* inner self, given that the self is socially constructed, changing with one’s goals, contexts, and social roles. Sheldon, Ryan, Rawsthorne, & Ilardi (1997) defined authentic behavior as that which is “*phenomenally* experienced as being authentic by the self… or internally caused,” and that to act authentically is to act “with a full sense of choice and expression” (p. 1381, emphasis mine). Authenticity is thus a *subjective* experience. This definition is remarkably similar to the psychological need for autonomyin the context of self-determination theory, where autonomy “connotes an inner endorsement of one’s actions” and “people experience themselves as initiators of their own behavior” (Deci & Ryan, 1987, p. 1025). Research conducted under the umbrella of self-determination theory has found consistent positive consequences for *feeling* authentic, such as increased happiness and well-being (e.g., Kernis & Goldman, 2004; Ryan & Deci, 2004; Sheldon & Elliot, 1999).

**Perceived Authenticity**

There are many academic definitions for authenticity (Sheldon, 2009), and many researchers note that researchers often define it vaguely or not at all (e.g., Beverland, 2005; Kadirov, 2015; Molleda & Jain, 2013). But the present research question concerns how the lay person perceives authenticity. How do people define authenticity and judge authenticity? How do perceptions of authenticity affect attitudes and behavior? Evidence answering each of these questions is demonstrated across various fields of study, such as psychology, sociology, marketing, and communications.

There is no one definition for authenticity. What is perceived as authentic depends on the person and the context; perceptions of authenticity are a “socially constructed interpretation… of what is observed rather than properties inherent in the object” (Beverland & Farrelly, 2010, p. 839; see also Grayson & Martinec, 2004). Qualitative research best shows the breadth of ways people understand authenticity.

Beverland and Farrelly (2010) asked participants how they defined authenticity and presented them with about 100 images depicting various brands, cultural icons, tourist sites, etc., to facilitate discussion about what they perceived to be authentic. Among many other reasons, interviewees attributed authenticity to products because the product did what it was advertised to do (shampoo) or gave people greater ability to develop knowledge (Apple computers), while they ascribed inauthenticity to products for being expensive and exploitatively produced (Nike) or being unhealthy (cigarettes).

The perceived authenticity of items in museums can literally be authentic in that they are *actual* historical objects from an important time or relating to an important figure (Grayson & Martinec, 2004; Hede, Garma, Josiassen, & Thyne, 2014). Interestingly, however, people are very willing to assign authenticity to objects that they *know* are contrived attempts to represent a fiction. Grayson and Martinec (2004) interviewed people visiting Shakespeare’s birthplace and the Sherlock Holmes museum, finding that people used different definitions of authenticity in their judgments in labeling each of these attractions as authentic. The authors define *indexical authenticity* to refer to an object that is the original—not a copy or replica—while *iconic authenticity* refers to an object that accurately represents the original, without actually being the original. For example, an indexically authentic basketball jersey is one a player wore while playing in a game, while an iconically authentic jersey is one made *exactly* like the player’s, but was never actually game worn.

This concept of iconic authenticity was supported by patrons of the museum finding Sherlock Holmes’s—a fictional character—“possessions” to be authentic because: “You have a feeling as if [Holmes] really touched everything,” the museum contained “things that Sherlock Holes might have used to solve the crimes,” and that one could see “[Holmes’s] living quarters, and where he sat, did his writings and thinking” (pp. 300-301). Patrons also noted the authenticity of the Holmes museum because everything inside looked as if it belonged at that historical time; in this case, people’s perceptions of authenticity are tangled up with feelings of nostalgia for a (fictionalized) past.

Kovacs, Carroll, and Lehman (2014) asked participants to choose or nominate words that would describe a restaurant as authentic. In addition to the synonyms commonly found across authenticity research (e.g., genuine, real), participants also associated authentic restaurants with words like skilled, traditional, historical, expert, professional, and iconic. Some of these words contradict what people perceive to be authentic in political candidates, as amateurism and non-professionalism can lend an aura of authenticity to a candidate (Enli, 2017; Manning, Penfold-Mounce, Loader, Vromen, & Xenos, 2017).

Researchers across disciplines have also studied the correlates, causes, and consequences of perceived authenticity—a literature that is not tied together by an overarching theoretical framework. Psychologists have found that perceiving romantic partners as authentic relates to greater relationship quality (e.g., increased satisfaction, less avoidance; Wickham, 2013), seeing one’s coworker as authentic relates to liking and trusting that coworker (Liu & Perrew, 2006), and people high in power (i.e., a CEO) are seen as less authentic in expressing how they feel than those in less power (Kim et al., 2017).

Positive consequences have been found for perceived authenticity in marketing. The more people perceive a variety of products to be authentic, the more they are willing to pay for them (Kadirov, 2015). Kovacs, Carroll, & Lehman (2014) analyzed nearly 19,000 Yelp reviews for restaurants, finding that the more people mentioned authenticity-related words, the more positively they reviewed the establishment. Family-owned restaurants were perceived as authentic, while chain restaurants were seen as inauthentic. Restaurants could also be categorized into keywords, such as Italian, Chinese, burgers, sandwiches, pizza, etc.; the more categories restaurants belonged to, the less authenticity-related words reviewers used. Kovacs and colleagues replicated these findings in a vignette experiment, as well.

Believing that one’s Airbnb stay was authentic (i.e., representative of local way of life, conducive to interacting with the local community) was positively related to consumers agreeing they paid a reasonable price and that the stay was a good value (Lian, Choi, & Joppe, 2016). Public figures are perceived as less authentic when they hire someone to write their social media posts for them (Cohen & Tyler, 2016). In-depth interviews with business owners and political consultants show that authenticity is a concept thought to be important in brand management (Beverland, 2005; Sorazio, 2017).

Two studies on perceived authenticity are more directly related to the current research question. Hahl, Kim & Sivan (2017) examined why people could find a “lying demagogue” to be such an authentic political candidate. In a minimal-group experiment, Hahl and colleagues found that participants who read about a candidate telling obvious lies and making misogynist remarks perceived this candidate to be *more* authentic when the participant felt like they were from an aggrieved social group and that the political system was not representing their interests. The authors argued that this was because lying and misogyny flout the established norms of a political system that participants found illegitimate.

Pillow, Crabtree, Galvin, and Hale (2017) also investigated why people perceived authenticity in politicians. They argued people are motivated to see candidates they support as authentic. In a survey about five candidates from the 2016 United States presidential election, they found that the positive relationship between (a) perceiving the candidate to speak candidly and (b) perceiving the candidate to be authentic was stronger when people liked the candidate, compared to when they did not. Pillow and colleagues argued that this supports a motivated reasoning (Kunda, 1990) account; they interpreted the interaction to be due to people liking the candidate weighting perceived candor more when judging the authenticity of a candidate then those who dislike the candidate. They also demonstrated that perceiving the candidate as unfiltered and not “politically correct” predicted people thinking the candidate was rejecting external influences on their behavior, which in turn predicted perceiving the candidate as authentic. They argued that this supports a correspondent inference (Jones & Davis, 1965) account, given that people assume others are acting on their own volition when when acting against norms.

**Defining Perceived Authenticity**

I define *perceived authenticity* as an observer’s judgment of how much a target individual is acting in accordance with what the observer perceives to be the target’s actual beliefs, attitudes, goals, and desires (also see Liu & Perrew, 2006). This situates the construct wholly in the observer. For example, Daisy believes Adam supports Democrats, while John believes Adam supports Republicans; if Daisy and John see Adam at a rally for a Democratic candidate, Daisy will perceive Adam’s actions to be authentic, while John will see them as inauthentic. Whether or not Adam actually supports Republicans or Democrats is irrelevant.

**Justifying Prejudice: The Justification-Suppression Model**

Prejudice is a negative evaluation of a social group or of an individual primarily based on their presumed group membership (Crandall & Eshleman, 2003). People feel and want to express prejudice, yet it is socially unacceptable to do so—how do people manage this tension? The justification-suppression model (JSM; Crandall & Eshleman, 2003) answers this question by synthesizing the commonalities between various modern theories of prejudice into a process involving four components: genuine prejudice, suppression, justification, and expression.

*Genuine prejudice* is the pure, unalloyed underlying negative affect one feels toward a social group or member of a social group. It is a motivational state, like all affect (Brehm, 1999), that drives justification and expression. Unfortunately, it is considered to currently be unmeasurable, as any self-report will be contaminated by biases (e.g., social desirability) and one’s inability to report feelings they are not consciously aware they possess. Implicit and physiological measures are promising techniques for tapping into genuine prejudice, but there are still unresolved questions about what precisely these measures capture, given that they tend to correlate weakly with one another (Fazio & Olson, 2003) and do not demonstrate greater predictive validity than do self-reports (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013).

*Suppression* is an externally or internally motivated (Plant & Devine, 1998) attempt to inhibit the expression of one’s prejudice. The primary external motivators are social norms and institutional rules (e.g., workplace code of conduct) that proscribe the expression of prejudice. Internal motivators can include one internalizing these external motivations (Deci & Ryan, 2000) or values and beliefs that one holds dear conflicting with expressing prejudice (e.g., egalitarianism). Crucially, suppressive forces do *not* necessarily eliminate the underlying genuine prejudice—they merely keep genuine prejudice from being expressed. Crandall and Eshleman (2003) review a wide variety of research demonstrating that suppressing prejudice is taxing on an individual; suppression requires cognitive effort. Despite suppressive forces acting on an individual, people remain motivated to relax these effort and free the expression of genuine prejudice. To do so, people seek out justifications for prejudice.

*Justifications* are any social or psychological process that allows one to feel as if they can express prejudice without suffering negative external (e.g., reprimand from others; workplace punishment) or internal (e.g., guilt, shame) consequences. These justifications release prejudice *expression*. For example, when people have already demonstrated that they can be non-prejudiced, they are more likely to act in discriminatory ways (Monin & Miller, 2001); more generally, people express prejudice when they feel like they have demonstrated that they have standing to do so (Miller & Effron, 2010).

**Authenticity as Justification for Prejudice**

I propose that people who are prejudiced are more likely to perceive an expression of prejudice as “authentic.” How can we see this as a justification for prejudice? The perceived authenticity justification can be seen as examples of both *vicarious justification* and *justification by rearticulation.*

**Vicarious justification.** Anecdotal evidence abounds that people will justify another’s expression of prejudice; this is clear from studying the campaign and presidency of Donald Trump (e.g., Bump, 2018; Burton, 2018; Horsey, 2016; Scott, 2018). White and Crandall (2017) demonstrated this empirically. Across a series of experiments, participants were told that a man was fired for expressing (a) anti-Black prejudice or (b) control statements, such as negative statements about police or coworkers. Participants were then asked how much they agreed that firing the employee violated his freedom of speech (e.g., “[The employee’s] bosses disrespected his right to free speech”). Various measures of prejudice predicted relevance of freedom of speech in the anti-Black prejudice conditions (meta-analytic *r* = .43)—but not in the control conditions (meta-analytic *r* = .09). People were *not* principled in marshaling freedom of speech: People defend others’ prejudiced expressions as a function of their own. White and Crandall found that this relationship was partially due to the termination of a prejudiced employee threatening the expressive autonomy of prejudiced participants (i.e., agreement with items such as: “I feel free to express my ideas and opinions”). People engage in vicarious justification: They feel the suppression placed on similarly-prejudiced others and strategically deploy values to protect these others. One reason to believe prejudice predicts perceived authenticity of similarly-prejudiced statements is the reasoning that people can claim an expression of prejudice is authentic as a way to vicariously justify prejudice.

**Justification by rearticulation.** Omi and Winant (1994) argued that racism did not necessarily decrease in the United States after the gains of the Civil Rights Movement of the 1960s; their focus is less on if racism is “getting better” or “getting worse” and more on how racist ideology changes. They argue that racism is *rearticulated*: “The new right generally does not display *explicit* racism. It has gained political currency by rearticulating racial ideology” (p. 127). Omi and Winant discuss how conservative racial ideology in the United States rearticulated from overt racism to concepts like busing, affirmative action, and reverse-discrimination. Lee Atwater, a then aide to Republican president Ronald Reagan, explicitly described his conscious effort to rearticulate racial arguments in an interview:

You start out in 1954 by saying, “Nigger, nigger, nigger.” By 1968 you can’t say “nigger”—that hurts you, backfires. So you say stuff like, uh, forced busing, states’ rights, and all that stuff, and you’re getting so abstract. Now, you’re talking about cutting taxes, and all these things you’re talking about are totally economic things and a byproduct of them is, blacks get hurt worse than whites…. “We want to cut this,” is much more abstract that even the busing thing, uh, and a hell of a lot more abstract than “Nigger, nigger” (Perlstein, 2012).

Although Omi and Winant focus on the “new right” (focusing on Reagan-era politics) liberal politicians have also been guilty of rearticulating racism (Haney-López, 2013).

Another reason for the hypothesized relationship between prejudice and perceived authenticity is that claiming prejudice as being “authentic” is a rearticulated way of expressing prejudice. It is a justification in a sense that it allows someone to express the idea that they agree with the prejudiced statement in an ostensibly race-neutral way—that is, without ever explicitly mentioning the target group at all.

**Contributing Mechanisms**

There are a number of contributing mechanisms that could explain the process by which people come to see the expression of prejudice as authentic, beyond simple agreement. I propose four accounts for capturing these contributing mechanisms: norms, social projection, balance, and motivated reasoning.

**Norms.**

**Social projection.** People believe others are similar to them—an automatic cognitive heuristic termed *social projection* (Kreuger, 2007). The more prejudiced somebody is, the more common they think it is in society; for example, Watt and Larkin (2010) found that participants high in prejudice estimated 71% of people in their country would also be prejudiced; lowly prejudiced participants estimated this to be 51% (see also Pedersen, Griffiths, & Watt, 2008). If prejudiced people think that prejudice is common in society, then they should be more likely to think that any given member of this society is prejudiced; thus, expressing prejudice is perceived as more authentic. In contrast to the JSM, this social projection explanation for the positive relationship between prejudice and perceived authenticity is an unmotivated, cognitive one—it is a simple perceptual bias due to perceiving others are similar to oneself. The social projection account implies two hypotheses: First, participants’ own prejudice should positively predict perceived descriptive normativity in society (i.e., how many people actually feel prejudice); second, perceived descriptive normativity should increase perceived authenticity of prejudiced statements.

**Balance.** Perhaps one of the simplest and elegant theories in social psychology is *balance theory* (Heider, 1958, Chapter 7). A primary contention in balance theory is that people are motivated to create a balanced state—a “harmonious state, one in which the entities comprising the situation and the feelings about them fit together without stress” (Heider, 1958, p, 180). A classic example is the “*p-o-x triad*,” in which relations between these three elements *p*, *o*, and *x* can be positive or negative. A balanced state is achieved when the signs of each of the three relations multiplies out positively. For example, Ecsch (1950) presented participants with a situation: A man named Bob (*p*) finds Jim (*o*) to be very unintelligent; however, one day Bob reads poetry (*x*) he really likes, tracks down who wrote the poems, and finds out that it was Bob who wrote them. This situation is imbalanced: Bob has a negative opinion of Jim, a positive opinion of Jim’s poems, and Jim has a positive relationship with his poems, given that he wrote them. This multiples out negatively (e.g., -++ = -), thus creating imbalance. Esch asked participants to describe what would happen. About 82% of the participants resolved the imbalance by changing the sign of one of the unit relations. For example, 46% of participants wrote some form of “Bob changes his mind about Jim. (Readers might be anecdotally familiar with the discomfort of having positive feelings toward two friends who dislike one another or only liking one member of a married couple.) People searching for affective consistency has been the impetus for a large amount of research in social psychology since and many psychological theories are consistent with balance theory’s predictions (Crandall, Silvia, N’Gabala, Tsang, & Dawson, 2007; Zajonc, 1960).

Consider the present scenario: A participant’s (*p*) relationship with a prejudiced statement (*o*) is positive if they share the same prejudice and the concept of authenticity (*x*) has a positive, moral connotation—it is virtually axiomatic across all fields that authenticity is generally a *good* thing. To achieve balance, this prejudiced person concludes the prejudiced statement was authentic, creating a positive unit relationship between the two and ensuring affective balance. Balance theory implies the following hypothesis: Presenting authenticity as a *negative* thing should *decrease* the positive relationship between self-reported prejudice and perceived authenticity.

**Motivated reasoning.** Pillow et al. (2017) found that the correlation between perceptions of a political candidate’s candor and perceived authenticity was *weaker* for those that did not like the candidate than for those who did. They argue this is evidence for motivated reasoning, because people relied on different information to make their judgments of authenticity depending on their pre-existing attitudes. I test this more directly in the present paper.

Kunda (1990) argued that “motivation may affect reasoning through reliance on a biased set of cognitive processes: strategies for accessing, constructing and evaluating beliefs” (p. 480); that is, people engaged in *motivated reasoning.* Kunda saw the motivation behind reasoning stemming from two categories of goals: accuracy goals and directional goals.

**The Present Studies**

**Study 1**

**Method**

I recruited 125 people from Amazon’s Mechanical Turk (MTurk) to participate in a “survey on perceiving others’ attitudes.” This sample size allows 80% power to detect an effect of *r* = .25 and 90% power to detect an effect of *r* = .30. These correlations were informed by being on the lower-bound of relevant past research (White & Crandall, 2017). A total of 126 people participated. Participants’ ages ranged from 19 to 69 (*M* = 34.9, *SD* = 10.95), 61% identified as male, 69% identified as White, and no participants indicated that they were Muslim.

Participants first read nine statements that were ostensibly taken from social media, comment sections, and elsewhere on the internet. Three were negative statements about Muslims (e.g., “With all that’s going on, I think it is OK for people to be suspicious of Muslims”), and three were negative statements about politicians (e.g., “All politicians really care about is themselves. They’ll do anything to get more and more power”). I included three control statements that were also negative and went against descriptive norms, but about trivial tastes and not loaded with the same moral weight as prejudiced statements. These statements were about people disliking cookies, the beach, and pizza (e.g., “What’s the big deal about pizza? I’ve never tasted a slice of pizza that tasted good”). In response to each of these statements, participants indicated on a seven-point scale (from 1, *Strongly Disagree* to 7, *Strongly Agree*) how much they thought that, in saying the statement, that person was being “true to themselves,” “honest,” “authentic,” and “genuine.” For each group of statements (Muslims, politicians, cookies, beach, and pizza), these items were averaged together to measure *perceived authenticity*.

Participants then filled out a demographic questionnaire. At the end of this page, I measured prejudice against Muslims and politicians, as well as how much participants disliked cookies, the beach, and pizza.

Anti-Muslim prejudice was measured using seven items adapted from measures of modern prejudice (Biernat & Crandall, 1999) that tapped into beliefs about Muslims (e.g., “Muslims living here should not push themselves where they are not wanted”) as well as feelings toward them (e.g., “How much do you like or dislike Muslims?”).

Anti-politician prejudice was measured using seven items, some adapted from a standard social distance questionnaire (Biernat & Crandall, 1999; “I would like a politician to be a close personal friend”) and others tapping into feelings toward politicians (e.g., “I admire politicians”).

Lastly, participants were asked how much they liked cookies, the beach, and pizza on a seven-point scale (from 1, *Not at all* to 7, *Very much so*). These items were reverse-scored to indicate control dislike statements—ones that are non-normative, but trivial tastes.

**Results**

Correlations that test the primary hypothesis are reported in the diagonal of Table 1; this table presents the correlations between each combination of dislike and perceived authenticity measures. The more prejudice people reported toward Muslims, the more they perceived similarly-prejudiced statements about Muslims to be authentic, *r* = .38, *p* < .001, and the more prejudice people reported towards politicians, the more they thought prejudiced statements about politicians were authentic, *r* = .18, *p* = .048. However, control dislikes did not correlate with thinking the control dislike statements were authentic.

Was this due to a *general* tendency for people who report prejudice to also think negative statements represent people’s authentic selves? EXPLAIN SUCCINCTLY HERE. Confidence intervals for these differences in correlations were calculated using Zou’s (2007) method (Diedenhofen & Musch, 2015). The correlation between prejudice against Muslims and perceived authenticity of anti-Muslim statements was .38, while the correlation between prejudice against Muslims and perceived authenticity of anti-politician statements was .05. The difference between these two was .33 [.16, .50]; the relationship between Muslim prejudice and thinking anti-Muslim statements are authentic were stronger than the relationship between prejudice against Muslims and perceived authenticity of anti-politician statements. This was also the case with anti-politician prejudice and perceived authenticity of anti-Muslim statements, .24 [.06, .41].

Table 1

*Correlations Between Dislike and Perceived Authenticity*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Perceived Authenticity | | | | |
| Dislike | Muslims | Politicians | Pizza | Beach | Cookies |
| Muslims | **.38** | .05 | -.16 | -.05 | -.16 |
| Politicians | -.06 | **.18** | -.02 | .02 | -.05 |
| Pizza | -.13 | -.15 | -.03 | .01 | -.02 |
| Beach | **-.24** | **-.28** | -.10 | -.17 | -.06 |
| Cookies | -.05 | -.12 | .05 | -.09 | .05 |

Bolded *r*s, *p* < .05.

**Study 2**

**Method**

Participants were recruited as they were walking around the University of Kansas campus. Research assistants approached passers by and asked if they would like to fill out a short, one-page survey in exchange for a piece of candy. Sample size was determined by how many people I could recruit by the end of the semester. A total of 221 people participated, but 7 participants were excluded for partial responding.

Participants were randomly assigned to one of two conditions: An *illegal immigrant condition* or a *Kansas State condition*. In the former, participants read two negative statements about illegal immigrants (e.g., “With all that’s going on, I think it is OK for people to be suspicious of illegal immigrants”); in the latter, participants read two negative statements about Kansas State students—a rival school of the University of Kansas (e.g., “Students that go to Kansas State smell weird”). The same measure of perceived authenticity was used in this study as in Study 1, and participants were again told that these statements came from social media websites and comment sections on the internet. Participants then reported how they felt about illegal immigrants and Kansas State students.

Five of the items used in Study 1 for Muslims were employed here, adapted to illegal immigrants (e.g., “Illegal immigrants here teach their children values and skills different from those required to be successful in the United States”); five of the items used in Study 1 for politicians were employed here, adapted to Kansas State students (e.g., “Kansas state students or graduates are similar to me”). In order to keep the survey at a one-page limit, demographic questions were not asked.

**Results**

I predicted that prejudice against illegal immigrants would predict perceived authenticity, but only in the illegal immigrant condition, while anti-Kansas State prejudice would only predict perceived authenticity in the Kansas State condition. These two prejudices were positively correlated, *r* = .28, *p* < .001, so I use the irrelevant prejudice as a control in all analyses.

First, I regressed perceived authenticity on anti-Kansas State prejudice, anti-illegal immigrant prejudice, condition, and an interaction between the latter two predictors. The condition by anti-illegal immigrant prejudice was significant, *b* = -.44, *SE* = 0.18, *t*(209) = -2.46, *p* = .015. Prejudice against illegal immigrants was positively related to authenticity in the illegal immigrant condition, *b* = .36. *SE* = .12, *t*(209) = 2.98, *p* = .003; however, it was not related to authenticity in the Kansas State condition, *b* = -.08, *SE* = .13, *t*(209) = -0.56, *p* = .575.

Next, I regressed authenticity on anti-illegal immigrant prejudice, anti-Kansas State prejudice, condition, and the interaction between the latter two predictors. The condition by anti-Kansas State prejudice interaction was significant, *b* = .45, *SE* = .16, *t*(209) = 2.75, *p* = .007. Prejudice against Kansas State students was positively related to authenticity in the Kansas State condition, *b* = .53, *SE* = .12, *t*(209) = 4.44, *p* < .001; however, it was not related to authenticity in the illegal immigrant condition, *b* = .08, *SE* = .12, *t*(209) = 0.66, *p* = .513.

**Study 3**

**Method**

I recruited 200 people from MTurk to participate in a “survey on perceiving other peoples’ attitudes.” Throughout this paper, participants who completed one of these studies were barred from participating in a subsequent one (Litman, Robinson, & Atterbock, 2017). The analyses for this study involves multilevel modeling, and I was unsure of proper a priori expected population parameters to choose for a power analysis, so sample size was determined from a subjective decision of what seemed appropriate. I recruited 200 participants, and each participant contributed 10 data points; I found a level one *n* = 2000 and a level two *n* = 200 to be reasonable. Participants’ ages ranged from 19 to 70 (*M* = 34.15, *SD* = 11.56), 54% identified as male, and 76% identified as White. Participants answered a questionnaire in the following order.

**Target groups.** Most of the following questions were asked once for each target group. These target groups were: Black people, transgender people, fat people, police officers, lawyers, business people, prostitutes, drug dealers, blind people, and deaf people.

**Perceived descriptive normativity.** Participants were asked to think about Americans in general and indicate “what percentage of Americans, if they were being truly and totally honest with themselves, would admit they feel negatively” toward each of the target groups; participants indicated their responses on a 0 to 100 scale.

**Perceived prescriptive normativity.** Participants were then asked to pivot from thinking about what “Americans *actually* feel” to think about “how Americans think people *should* feel.” They were asked to indicate on the same 0 to 100 scale what percentage of Americans “think it is OK to feel negatively toward these groups?”

**Perceived authenticity.** Ten quotations, ostensibly “taken from social media posts or comments on the internet,” were presented to participants. Each target group had one corresponding negative statement about them (e.g., “Blacks are the people causing the racial tension in America today,” “Business people don’t care about anyone but themselves and making lots of money”); the same four items employed in Studies 1 and 2 were asked for each statement.

**Prejudice.** Participants were asked how they feel about each of the target groups on a scale from 0 (*Very negatively*) to 100 (*Very positively*). These items were reverse-scored such that higher scores indicated more prejudice toward the group.

**Concern about political correctness.** The belief that there are “PC crusaders” in society censoring speech, changing history, and dividing the country with identity politics were measured using six items (Lalonde, Doan, & Patterson, 2000); participants were asked how much they agree with these statements (e.g., “I believe that there are ‘PC’ crusaders who want to censor people’s speech”) on a seven-point scale.

**Demographics.** Lastly, participants indicated their age, gender identity, race, political outlook (1, *Very liberal* to 7, *Very conservative*), and political affiliation (1, *Strongly Republican* to 7, *Strongly Democrat*). Political outlook and reverse-scored political affiliation were averaged together to measure right-wing political identification (*r* = .85).

**Analysis details.** As constructs were measured on very different scales (i.e., seven- versus 101-point), all measures were standardized (across, not within, individuals) before analyses. Measurements were considered at the first level, while individuals were at the second. Thus, perceived descriptive and prescriptive normativity, perceived authenticity, and prejudice were at the measurement level, while political correctness concern and right-wing political identification were at the person level. All regression coefficients, when measured at the first level, were allowed to differ by individual; that is, random slopes were defined for all level-one predictors. Lastly, I used Satterthwaite’s approximation for degrees of freedom for all *t*-tests of regression coefficients (Kuznetsova, Brockhoff, & Christensen, 2017).

**Results**

Prejudice again correlated positively with perceived authenticity of negative statements, *b* = .24, *SE* = .03, *t*(181.53) = 8.31, *p* < .001. Additionally, the *more* that participants thought others had the prejudice (i.e., descriptive normativity), the *more* authentic they perceived the speaker to be, *b* = .20, *SE* = .02, *t*(186.24) = 8.60, *p* < .001.

**Moderator analyses.** I tested if the relationship between prejudice and perceived authenticity was moderated by (a) perceived prescriptive normativity, (b) concern about political correctness, or (c) right-wing political identification. For each of these moderators, I regressed perceived authenticity on prejudice, the moderator, and the interaction between the two. Each of the three models included prejudice as a random slope; when testing the interaction by prescriptive normativity, coefficients for both prescriptive normativity and the interaction were included as random slopes, as well.

The prejudice by prescriptive normativity interaction on authenticity was significant, *b* = -.06, *SE* = .02, *t*(156.39) = -2.82, *p =* .005. Probing this interaction (Preacher, Curran, & Bauer, 2006) showed that, when participants reported *low* perceived prescriptively normativity (i.e., a standard deviation below the mean), the relationship between prejudice and authenticity was *b* = 28, *SE* = .05, *z* = 6.13, *p* < .001. When the group was high perceived prescriptive normativity (i.e., a standard deviation above the mean), the relationship was about half as strong, *b* = .15, *SE* = .03, *z* = 4.96, *p* < .001.

The prejudice by political correctness concern on authenticity was *not* significant, *b* = .005, *SE* = .03, *t*(172.5) = 0.17, *p* = 0.864.

**Study 4**

TOUCH ON SPECIFICS OF WHY I CHOSE THIS MANIPULATION INSTEAD OF OTHER, STANDARD ONES.

**Method**

**Results**

**Study 5**

**Method**

I recruited 200 participants from MTurk to participate in a “study on person perception.” As this study aimed to eliminate the correlation between prejudice and perceived authenticity with an experimental manipulation, sample size was determined by simulating data where the correlation between two variables was *r* = .40 for half of the participants and *r* = .00 for the other half, then choosing the sample size that led to 80% power. This power analysis was used to inform sample sizes in all subsequent studies. Two participants failed to complete the writing task (described below); they were excluded from all analyses, leaving a final sample size of 198. Participants’ ages ranged from 19 to 77 (*M* = 36.39, *SD* = 11.80), 52% identified as male, and 76% identified as White.

Participants were told that the study was aimed at the question, “Why do we think that others are the way they are?” They were told that they would answer a few questions and then comment on previous participants’ responses to those same questions. Participants were then randomly assigned to an *expression condition* or a *suppression condition*.

In both conditions, participants were told that they were in the version of the survey about “fat people,” and they were asked to give a number of reasons why people might be fat. They were given ten blank lines below the instructions to do so. In the *expression condition*, they were told that it was “important that you feel free to write whatever reasons” that they think of, whether they agree with them, whether they think the reasons are nice or mean. In the *suppression condition*, they were told that it was important that their reasons “are not focused on blaming fat people for their bodies,” because “quite a lot of research shows that blaming people for their weight is a sign of prejudice.” In this latter condition, an additional question asked them to look over their answers again, making sure that nothing they said blamed fat people for their weight. After double-checking, they were instructed to select a button that read, “Yes, I followed the directions.”

On the next page, participants were told: “Some people already answered the *exact same* question that you just answered. Here is one of the reasons that they gave for people having obesity...” This was repeated four times, each with a new statement. Two blamed fat people for their weight (i.e., “they have no willpower,” and “they’re too lazy to exercise”)—I refer to these as the *negative statements*. The other two—the *neural statements*—did not (i.e., “their genes make them overweight,” and “environmental things like poverty or bad parenting”). Participants were asked five questions about each statement.

**Perceived authenticity.** The same four questions used in Studies 1 – 4 were used to measure the perceived authenticity of the speaker. The eight items for the negative statements were averaged together, while the eight items for the neutral statements were averaged together to measure perceived authenticity.

**Manipulation check.** Participants were asked how much they agreed with the statement, “This answer follows the rules of the task,” on a seven-point scale (1, *Strongly disagree* to 7, *Strongly agree*). The items for the positive and neutral statements were again averaged together separately.

Participants then answered a demographics questionnaire, followed by Crandall’s (1994) anti-fat attitudes questionnaire. In line with the theory behind the scale, I used the “Dislike” subscale as my measure of anti-fat prejudice. An example item reads: “I really don’t like fat people much.” Participants indicated how much they agreed with these statements on a seven-point scale (1, *Strongly disagree* to 7, *Strongly agree*).

**Results**

The two negative statements were seen as following the rules less in the suppression condition (*M* = 1.89, *SD* = 1.27) than in the expression condition (*M* = 6.1, *SD* = 0.85), *t*(158.31) = 27.12, *p* < .001, *d* = 3.86 [3.39, 4.43]. There was no difference between the two conditions for the neutral statements, *t*(170.52) = 1.08, *p* = .281, *d* = 0.15 [-0.13, 0.43]. Note that Welch’s *t*-test was employed due to unequal variances across conditions (Delacre, Lakens, & Leys, 2017).

I tested my hypothesis by regressing perceived authenticity of the negative statements on anti-fat prejudice, condition, and the interaction between the two. The interaction was significant, *b* = .36, *SE* = .12, *t*(194) = 3.16, *p* = .002. Probing the interaction with simple slopes analyses (Preacher, Curran, & Bauer, 2006) showed that prejudice was positively correlated with perceived authenticity in the suppression condition, *b* = .34, *SE* = .08, *t*(194) = 4.41, *p* < .001. There was no relationship between the two in the expression condition, *b* = -.02, *SE* = .08, *t*(194) = -0.23, *p* = .815.

This interaction held even after including the other two anti-fat attitudes subscales—fear and willpower—as additional predictors, *b* = .21, *SE* = .11, *t*(192) = 1.99, *p* = .048.Regressing the perceived authenticity of the neutral statements on anti-fat prejudice, condition, and the interaction between the two yielded a nonsignificant interaction, *b* = .10, *SE* = .10, *t*(194) = 1.05, *p* = .293.

**Study 6**

**Method**

I recruited 200 people from MTurk to participate in a “study on person perception.” A total of 202 people participated, but one was dropped from analyses for partial responding. Participants’ ages ranged from 18 to 70 (*M* = 36.71, *SD* = 11.70), 42% identified as male, and 77% identified as White.

The procedure was identical to Study 5, and demographics and prejudice were measured the same as in Study 5. Only the measures after reading both negative and both neutral statements changed. After reading each statement, participants indicated on a seven-point scale how “authentic and genuine,” “true and honest to themselves,” “politically correct,” and “overly careful and too polite” they believed the person was being. The former two items were designed to measure perceived authenticity, with the latter two measuring perceived political correctness.

Considering items in response to negative and neutral statements separately, an exploratory principal axis factor analysis with an oblimin rotation supported this two-factor solution and with simple structure (using Kaiser-Guttman criteria, a parallel analysis, and examining scree plots). The four authenticity items for the negative statements were averaged together, while the four items for the neutral statements were averaged together to measure perceived authenticity. The same was done for perceived political correctness.

**Results**

The primary hypothesis was again tested by regressing perceived authenticity of the negative statements on anti-fat prejudice, condition, and the interaction between the two. The interaction was not significant in this study, *b* = .14, *SE* = .12, *t*(197) = 1.14, *p* = .255. However, the simple slopes followed the same pattern as in Study 4: Prejudice predicted perceived authenticity in the suppression condition, *b* = .27, *SE* = .09, *t*(197) = 3.11, *p* = .002, but not in the expression condition, *b* = .14, *SE* = .08, *t*(197) = 1.73, *p* = .085. Regressing perceived authenticity of the neutral statements on prejudice, condition, and their interaction yielded a nonsignificant interaction as in Study 4, *b* = .03, *SE* = .12, *t*(197) = 0.24, *p* = .808.

I also predicted that perceived political correctness of negative statements would be negatively related to perceptions of their authenticity; this was not the case, *r* = -.02, *p* = .818. Regressing perceived political correctness of the two negative statements on anti-fat prejudice, condition, and their interaction did not yield a significant interaction, *b* = -.07, *SE* = .13, *t*(197), *p* = .609. There was a significant zero-order correlation between prejudice and perceived political correctness, however, *r* = .30, *p* < .001.

**Study 7**

**Method**

I recruited 230 people from MTurk to participate in a task involving “classifying statements.” A total of 229 people participated; their ages ranged from 19 to 63 (*M* = 34.13, *SD* = 9.74), 55% identified as male, and 77% identified as White.

Participants were told: “We are creating a statistical model that will be able to predict how people will perceive others. These models must first learn from humans, however; this is where we need your help! In this task today, you will be reading statements made by others and classifying them...” They were told that the specific concept of focus was “*authenticity*,” and that they would be (a) classifying statements *about* authenticity as “good” or “bad” and (b) reading statements taken from the internet and rating how authentic those people were being in expressing those statements.

After indicating that they had read the instructions carefully, participants were randomly assigned to either an *authenticity is good* or an *authenticity is bad* condition. In the former, participants were given eight positive (e.g., “Being authentic can be soothing because it reduces stress in your life. If you are acting in a natural way without trying to change your behavior that is one less thing to worry about”) and two negative (e.g., “Sometimes being authentic means ignoring the expectations and desires of people around you. If you always act the way you truly feel deep down, you can hurt people's feelings and make them uncomfortable”) statements to rate as “good,” “bad,” or “unsure.” In the latter, participants were given two positive and eight negative statements.

On the same page, participants were asked, “How much of a good thing do you think being ‘authentic’ is?” and “How good of a thing do other people think being ‘authentic’ is?” on a scale from 1 (*Extremely bad*) to 7 (*Extremely good*). These items served as a manipulation check.

Participants were given a list of ten statements and asked to rate how authentic they thought people were being, on a scale from 0 (*Not at all*) to 100 (*Very much so*). Two of the statements had to do with illegal immigrants: “With all that’s going on, I think it is OK for people to be suspicious of illegal immigrants,” and “Illegal immigrants threaten what it means to be American. It is STUPID to let them come flooding in.” These two items were averaged together to represent perceived authenticity of prejudiced statements. The other eight statements were statements one might see elsewhere on the internet (e.g., “I’m pulling for the Minnesota Twins this year, they’re a fun team”).

Participants then filled out a demographics questionnaire and then indicated on a seven-point scale how much they agreed with a number of statements that were related to what they rated as authentic. Three were items illegal immigrants used in Study 2 (e.g., “Undocumented aliens here teach their children values and skills different from those required to be successful in the United States”), which were averaged together to form a measure of prejudice against illegal immigrants.

**Results**

Participants rated the goodness of authenticity higher in the *authenticity is good* condition (*M* = 6.30, *SD* = 0.82) than in the *authenticity is bad* condition (*M* = 5.28, *SD* = 1.23), *t*(198.53) = 7.38, *p* < .001, *d* = 0.98 [0.70, 1.25]. They also indicated that they believed *others* thought authenticity was better in the *authenticity is good* (*M* = 6.14, *SD* = 0.79) condition than in the *authenticity is bad* (*M* = 3.29, *SD* = 1.46) condition, *t*(175.14) = 18.42, *p* < .001, *d* = 2.43 [2.09, 2.78]. Welch’s *t*-test was again used due to unequal variances across conditions (Delacre et al., 2017).

The primary hypothesis was tested by regressing perceived authenticity on prejudice, condition, and the interaction between the two. The interaction was not significant, *b* = 1.74, *SE* = 2.49, *t*(225) = 0.70, *p* = .486. However, I replicated the same correlation between perceived authenticity and prejudice, *r* = .21, *p* = .002.

**Study 8**

I recruited 220 people from MTurk to participate in a “survey on perceiving others.” Participants’ ages ranged from 18 to 71 (*M* = 36.24, *SD* = 11.55), 47% identified as male, and 75% identified as White.

Participants were asked to “read about someone and answer questions about something they said” and randomly assigned to either an *accuracy* motivationor *expression* motivation condition. In the accuracy condition, I asked participants to “be as accurate as possible” and to “try to guess the correct answer” when reporting their perceptions of the target person. In the expression condition, I asked participants to “answer based on expressing your own opinion” and to “respond in a way that expresses what you personally think.” To help participants adhere to these goals, participants were told that they would be getting a bonus payment upon completion and that, “We only ask in return that you [be accurate in your ratings/respond based on your own opinion].”

Participants then read about a person named Colin, as well as some information about him: where he lives, what he does for work, food he likes, and some basic personality characteristics (also see White & Crandall, 2017; White & Molina, 2016). This was meant to be general, somewhat bland information for the purpose of allowing participants to feel like they have sufficient background information to reason with and make judgments about the target. At the end of the description, participants were told that Colin recently said the following statement: “Black people are so touchy about race that it is difficult to get along with them. They can be combative and assume the worst from White people. This makes me feel uncomfortable sometimes, which is why I don’t really like to hang around them much.”

I reminded participants about their goal before measuring perceived authenticity using the same items as in Studies 1 – 5. Participants answered a demographics questionnaire. Prejudice was then measured using an eight-item, seven-point symbolic racism scale (Henry & Sears, 2002; “Irish, Italian, Jewish, and many other minorities overcame prejudice and worked their way up. African-Americans should do the same, without any special favors”).

**Results**

I tested the primary hypothesis by regressing perceived authenticity on prejudice, condition, and their interaction. The interaction term was not significant, *b* = .09, *SE* = .13, *t*(216) = 0.69, *p* = .491. However, the simple slopes showed the predicted pattern: prejudice positively predicted authenticity in the expression condition, *b* = .25, *SE* = .10, *t*(216) = 2.57, *p* = .011, but not in the accuracy condition, *b* = .15, *SE* = .09, *t*(216) = 1.65, *p* = .101.