Original Research Article

Journal of Cross-Cultural Psychology
1–28

DOI: 10.1177/00220221251355914
journals.sagepub.com/home/jcc
Sage

The Bittersweetness of Self-Representations: Cross-Cultural Insights That the Self Triggers Both Positive and Negative Implicit Evaluations

Minghui Ni[1], Randy T. Lee[1], and Vivian Zayas[1]

Abstract

Positive implicit self-esteem has been argued to be a universal characteristic of the human mind. Yet, recent work using U.S. samples finds evidence suggesting that the self is implicitly bivalent. The bivalent-priming elicited by the self was markedly different from the univalent-priming elicited by personally significant, positive objects. The findings of implicit self-bivalence align with theoretical perspectives acknowledging the self's affective complexity. The present preregistered study examines whether implicit self-bivalence characterizes only Western, individualistic cultures, or if it reflects a more general feature of the human mind that also characterizes Eastern, collectivistic cultures (e.g., China). Replicating past work, in a U.S. sample ($N = 125$), priming the self in an evaluative priming task (EPT) facilitated *both* positive *and* negative target categorization (bivalent-priming), whereas priming a personally significant, liked object *only* facilitated positive target categorization (univalent-priming). Critically, in a Chinese sample ($N = 104$), we translated the EPT into Chinese (and validated it through back-translation) and observed the same result pattern. We also explored potential cultural differences in the magnitude of implicit self-bivalence (i.e., implicit self-positivity and self-negativity, respectively) but did not find statistically significant cross-cultural differences. The current work provides the first evidence that implicit self-bivalence is not culturally specific to Western samples. The clear demonstration of implicit self-bivalence in an Eastern sample provides initial evidence of its cross-cultural generality. We discuss the implications for understanding the nature of self-representations, the functions of implicit self-bivalence, and directions for future work.

Keywords

implicit social cognition, implicit attitude, ambivalence, indirect measures, cross-cultural research

The self has long been of interest across diverse psychological disciplines (James, 1890). Work in social cognition has been particularly interested in evaluations of the self that are triggered spontaneously, automatically, and often outside of conscious awareness (Farnham et al., 1999; Greenwald & Banaji, 1995). A growing body of research finds that implicit self-evaluations are

[1]Cornell University, Ithaca, NY, USA

Corresponding Author:
Vivian Zayas, Department of Psychology, Cornell University, Ithaca, NY 14853, USA.
Email: vz29@cornell.edu

overwhelmingly positive, and that such implicit self-positivity is observed across cultures (Buhrmester et al., 2011; Cai et al., 2007; Falk & Heine, 2015; Farnham et al., 1999; Hetts et al., 1999; Koole et al., 2001; Sedikides et al., 2015; Tafarodi et al., 2011; Yamaguchi et al., 2007). The robustness of such findings has led to claims that implicit self-positivity is a universal feature of the mind (Sedikides et al., 2015; Yamaguchi et al., 2007).

But is the implicit self *entirely* positive? Recent work finds that activating the mental representation of the self elicits implicit positivity, but also elicits *some* implicit negativity (Zayas et al., 2022). That is, instead of being *entirely* positive, new findings raise the possibility that the self is characterized by bivalence, encompassing *both* positive *and* negative implicit evaluations. To date, however, evidence for this phenomenon, which we refer to as *implicit self-bivalence*, has only been demonstrated in samples from the United States. This raises the question: Is implicit self-bivalence *only* observed in Western, individualistic cultures, or is it *also* evident in Eastern, collectivistic cultures, which would suggest that implicit self-bivalence is a more general feature of the human mind? And if it is evident in an Eastern culture, is the magnitude of implicit positivity, implicit negativity, or both, the same?

In the current preregistered research, we aim to replicate implicit self-bivalence in a U.S. sample and extend previous findings by examining its presence in a Chinese sample. Our hypotheses are informed by attachment theory (Bowlby, 1969). Just as the attachment behavioral system is considered universal (Bowlby, 1988; Hazan & Shaver, 1994) but its behavioral expression varies across cultures (e.g., Keller, 2013; Miyake et al., 1985; Polek, 2008; van IJzendoorn & Kroonenberg, 1988; van Ijzendoorn & Sagi-Schwartz, 2008), we reason that implicit self-bivalence may be a fundamental characteristic of the human mind. Our primary preregistered prediction is that implicit self-bivalence will be observed in both U.S. and Chinese samples, consistent with the idea that the self activates both positive and negative implicit evaluations across cultures. Importantly, the presence of implicit self-bivalence does not preclude the possibility of cultural variation in its magnitude or in its individual evaluative components. Given prior research on cultural differences in the self (e.g., self-construal, self-enhancement, dialectical thinking), implicit self-evaluations may vary across cultures. At the time of preregistration, we were agnostic as to whether such differences would emerge, including their direction. Accordingly, we conducted exploratory analyses examining potential cultural differences in the strength of implicit self-positivity, implicit self-negativity, or both.

Implicit Self-Positivity

Implicit self-evaluations are typically assessed using indirect measures that infer participants' self-evaluations through assessing their evaluations of self-relevant stimuli (e.g., Implicit Association Test; Farnham et al., 1999; Greenwald & Farnham, 2000; Name Letter Paradigm; Nuttin, 1985, 1987). For example, the Name Letter Paradigm was originally designed to assess implicit self-evaluations by presenting participants with letter pairs that consist of one letter in their name and one letter not in their name, and asking them to select the more attractive one. Participants' preference for name-letters, as compared with non-name letters, is used to index implicit self-evaluations. Similarly, follow-up research adapted the original paradigm and asked participants to rate the attractiveness of letters from the alphabet. The relative liking of name-letters compared with non-name letters, while taking into account the baseline attractiveness of the letters, is used to index implicit self-evaluations. As another example, the Evaluative Priming Task (EPT; Fazio et al., 1986) can also be used to assess implicit self-evaluations by assessing the extent to which self-referring primes (e.g., "ME") facilitate the categorization of subsequently presented positive versus negative target words (Bar-Anan & Nosek, 2014; Hetts et al., 1999). Participants' response time to positive relative to negative target words following self-referring primes, compared with control primes, is used to index implicit self-evaluations. In short, with

indirect measurements, participants are never directly asked to report their self-evaluations. Instead, such evaluations are inferred indirectly.

Across diverse indirect measures, researchers consistently and robustly show that the self-representation is implicitly positive (see Greenwald & Banaji, 1995, for a review). For example, in the Name Letter Paradigm described above, participants reliably favor name-letters over non-name letters, with the effect particularly pronounced for their initials (Johnson, 1986; Kitayama & Karasawa, 1997; Koole et al., 2001). Similarly, participants also show more favorable evaluations toward their birth date numbers compared with other numbers (Kitayama & Karasawa, 1997; Koole et al., 2001). In the EPT, participants respond faster to positive target words relative to negative target words, following self-referring primes as compared with control primes (Hetts et al., 1999).

Importantly, evidence of implicit self-positivity has been observed across diverse cultures (e.g., the United States, Canada, China, Japan; Kitayama & Karasawa, 1997; Szeto et al., 2009). Despite some exceptions, overall, cross-cultural research suggests comparable levels of implicit self-positivity across Eastern and Western cultures (Cai et al., 2007; Falk et al., 2009; Kitayama & Karasawa, 1997; Kitayama & Uchida, 2003; Kobayashi & Greenwald, 2003; Yamaguchi et al., 2007; see Falk & Heine, 2015, for a review). Such findings have led to the conclusion that implicit self-positivity is a universal characteristic of the human mind (e.g., Yamaguchi et al., 2007; c.f., Boucher, 2010), reflecting human beings' fundamental self-enhancement motivation—the need for high self-esteem and to maintain a positive self-image (Alicke & Sedikides, 2009; Falk & Heine, 2015; Greenwald & Banaji, 1995; Heine & Hamamura, 2007; Koole et al., 2001).

But Is the Self Entirely Positive?

In contrast to perspectives that emphasize the fundamental motivation for self-enhancement, other theoretical perspectives, research, and basic intuition all suggest that self-representations are affectively complex and that the self may not be entirely positive (Andersen & Chen, 2002; Andersen & Glassman, 1996; Baumgardner, 1990; Higgins, 1987; Klein, 1955; Kohut, 1966; Linehan, 1993). Various perspectives underscore that self-representations develop from and are maintained by repeated interactions with caregivers (Baldwin, 1994; Baldwin et al., 1990; Bowlby, 1969, 1973; Pietromonaco & Barrett, 2000). As in all interpersonal relationships, interactions with caregivers are inherently affectively complex (Murray et al., 2015; Murray, Holmes, et al., 2006). Even the most responsive and supportive caregivers can be unavailable, inattentive, or rejecting at times (Zayas et al., 2017; Zayas & Shoda, 2015). Thus, instead of simply being characterized by unconditional love and praise, instances of caregiver's social approval (vs. disapproval), as well as their responsiveness (vs. unresponsiveness) are the basis by which the self develops the idea of both the "good me" and the "bad me" (Sullivan, 1953).

Beyond early childhood, theories also highlight the role of self-evaluations in navigating social interactions. Social interactions are affectively complex, filled with opportunities for encountering social rewards (e.g., acceptance, affiliation, and belonging) and social threats (e.g., disapproval, rejection, and exclusion). Responses from the social world influence the accessibility and valence of self-evaluations (Baldwin, 1994; Baldwin et al., 1990; Leary, 2005; Leary & Baumeister, 2000; Murray, Griffin, et al., 2006; Williams, 2009). Most relevant, the sociometer theory (Leary, 2005; Leary & Baumeister, 2000) posits that self-evaluations serve as a gauge of one's relational worth. When individuals' encounter a social threat, their self-evaluations become more negative, signaling a loss of relational value, which may prompt responses to reintegrate oneself into the group (Leary & Baumeister, 2000; Williams, 2009).

To the extent that the affective complexity of interpersonal relationships give rise to the construction, development, and maintenance of the self-representation (Zayas et al., 2022), the self

should be characterized by implicit bivalence, being associated with both positive *and* negative implicit evaluations. Surprisingly, this possibility has received scant empirical attention.

Assessing Implicit Evaluations With Relative Versus Absolute Measures

One reason for the lack of attention to this question is that commonly used measures of implicit self-esteem have largely focused on the self's *relative* implicit positivity. Such relative measures assess evaluations on a single continuum, ranging from negative to positive (Bosson et al., 2000; Dehart et al., 2011; Dewitte et al., 2008; Greenwald & Farnham, 2000; McNulty et al., 2014; Zayas & Shoda, 2005). Without a doubt, measures of relative implicit positivity are highly valuable and informative in answering many research questions (e.g., Izuma et al., 2018; Wegener et al., 2015; Winter et al., 2018). For example, in work examining individual differences, researchers are often interested in how higher (vs. lower) net implicit self-positivity affects downstream outcomes, such as self-confident behaviors and responses to failure (e.g., Greenwald & Farnham, 2000; Krause et al., 2016).

However, research using only relative measures may miss the full picture of the complex structure of self-representations. To measure the extent to which the same concept activates both positive and negative evaluations, it is essential to use measures that tap into more absolute evaluations (Olson et al., 2007; O'Shea & Wiers, 2020) that capture evaluations on unidimensional continuums: one that assesses the strength of positivity (e.g., from not at all positive to strongly positive), and a separate unidimensional continuum that assesses the strength of negativity (e.g., from not at all negative to strongly negative). As an illustration (see Sakman & Zayas, 2024, for a similar discussion), consider implicit self-evaluations being separately assessed on positivity and negativity, with 0 indicating weak positivity or negativity and 10 indicating strong positivity or negativity. A scenario where implicit self-representations are associated with extremely strong positivity (scoring 10) and moderate negativity (scoring 5) and a scenario where implicit self-representations are associated with moderate positivity (scoring 5) and almost no negativity (scoring 0) would result in the same exact net implicit self-positivity using relative measures. Yet, both intuition and theory might expect that such structural differences in self-representations would affect psychological functioning and social interactions. In short, separately assessing positivity and negativity offers additional understanding of the cognitive architecture associated with the self.

Initial Evidence of Implicit Self-Bivalence

Building on these ideas, recent research has aimed to separately assess positive and negative implicit evaluations using more absolute indirect measures. By doing so, this work has uncovered that the normative (typical) structure of self-representations is implicitly bivalent; that is, although the self elicits implicit positivity, the self also evokes *some* implicit negativity (Zayas et al., 2022).

To assess positive and negative implicit evaluations separately, the researchers (Zayas et al., 2022) used the EPT (Fazio et al., 1986), a well-established paradigm to measure implicit evaluations of attitude objects (Bar-Anan & Nosek, 2014). Although researchers typically derive a single index from performance on the EPT to assess relative implicit evaluations, the EPT allows for the separate assessment of positivity and negativity. In a standard EPT, participants are first presented with a prime word, which is the name of an attitude object, or an affectively neutral stimulus that serves as a baseline. The presentation of prime word is then quickly replaced with a target word, which is either positive (e.g., *flower*, *diamond*) or negative (e.g., *bomb*, *disease*).

Participants' task is to categorize the target word as *good* or *bad* by pressing one of the two response keys. Studies using the EPT have consistently demonstrated the standard priming effect: when the valence of the prime and target word is congruent (i.e., both positive or both negative), target word categorization is facilitated (e.g., participants are faster). However, when the valence of the prime and target word is incongruent (i.e., one is positive whereas the other one is negative), target word categorization is inhibited (e.g., participants are slower; see Fazio & Olson, 2003, for a review). For example, if coffee is a positive attitude object to the participant, priming them with the word "COFFEE" would lead to faster response latencies in categorizing a subsequent positive target word such as "flower," but slower response latencies in categorizing a subsequent negative target word, such as "bomb." In contrast, priming participants with a negative attitude object, such as "GARBAGE" would lead to faster response latencies in categorizing a subsequent negative target word, but slower response latencies in categorizing a subsequent positive target word.

What are the implicit evaluations of the self? If the self is entirely positive, then the pattern of reaction times should resemble that of a positive attitude object, like coffee. But that was not the case. In Zayas et al. (2022), participants' own names were used as prime words in the EPT to measure their implicit evaluations of the self. Results showed that the self-prime significantly facilitated positive target categorization compared with baseline, reflecting that the self is associated with implicit positivity. But interestingly and critically, the self-prime also significantly facilitated negative target categorization, reflecting that the self is simultaneously associated with some implicit negativity. This work demonstrated, for the first time, that the self is associated with bivalent (i.e., both positive and negative) implicit evaluations, consistent with the theoretical perspectives highlighting the affective complexity of the self-representation.

To be clear, if the researchers had derived an index of *relative* implicit evaluations, they would not have observed implicit self-bivalence. In other words, the researchers could have computed a relative indicator by assessing the extent to which the self facilitated the categorization of positive targets and then subtracting the extent to which the self facilitated the categorization of negative targets. This would have reflected net implicit self-positivity, and the authors would have concluded that the self was overwhelmingly positive—a similar conclusion to the vast previous work documenting the robustness of the self's positivity. But, by separately assessing positive and negative implicit evaluations, Zayas et al. (2022) uncovered the bivalent structure of implicit representations. They found that implicit bivalence was not triggered by personally significant attitude objects but was instead uniquely triggered by the self.

This past work suggests that instead of being wholly positive, the normative (typical) structure of the self is bivalent. Traditionally, past work considers the coexistence of positive and negative (often termed "ambivalence") as problematic, linked with pathological outcomes (Guidano et al., 1983). Often it is viewed as only occurring for a subset of the population (e.g., individuals who are insecurely attached or possess low self-esteem; Ainsworth, 1978; Graham & Clark, 2006; Pietromonaco & Barrett, 2000; Riketta & Ziegler, 2006). The recent findings, however, found that implicit bivalence is a more typical phenomenon, occurring for the majority of individuals in the studies. Indeed, estimating the heterogeneity of implicit self-evaluations suggested that variation across participants only accounted for 1% of variance in implicit self-positivity and less than 1% of variance in implicit self-negativity, suggesting that variability among healthy, nonclinical populations in self-evaluations may be limited (Zayas et al., 2022). Such findings could be due to methodological limitations of indirect measures, but they may also reflect that the coexistence of positive and negative implicit self-evaluations is a basic feature of the human mind, consistent with past theorizing about the self's affective complexity. In the present work, we aim to explore this possibility further.

Cultural Similarity and Differences in Self-Representations

Does such implicit self-bivalence solely characterize Western, individualistic cultures, or does it also characterize Eastern, collectivistic cultures, reflecting that it may be a more general feature of the human mind? Our primary preregistered hypothesis is that implicit self-bivalence is not a feature of the mind unique to individuals from Western cultures but is generalizable to individuals from Eastern cultures. Specifically, we predicted that the self would elicit both positive and negative implicit evaluations in both a U.S. and a Chinese sample.

At the same time, we also considered the possibility that there may be cultural differences in each of its evaluative components; that is, there may be cultural differences in the strength of implicit self-positivity and implicit self-negativity, or both. Our reasoning of the cross-cultural similarity versus differences of implicit self-bivalence is informed by attachment theory. While the attachment behavioral system is assumed to be species-universal (Bowlby, 1988; Hazan & Shaver, 1994), some aspects of its structure and function appear to be relatively invariant across cultures (e.g., basic need for connection) whereas other aspects (e.g., behavioral expressions of support seeking) can be shaped by cultural practices and norms (Keller, 2013; Miyake et al., 1985; Polek, 2008; van IJzendoorn & Kroonenberg, 1988; van Ijzendoorn & Sagi-Schwartz, 2008). By analogy, we expect implicit self-bivalence, as a feature of the architecture of self-representations, to be generalizable and observable across cultures. However, the extent to which cultural factors shape the magnitude of implicit self-positivity, implicit self-negativity, or both is an open question—one that we examine through exploratory analyses.

Implicit Self-Bivalence as a Fundamental Feature of Self-Representations

Based on the idea that the experiences that are presumed to give rise to implicit self-bivalence should be similar across cultures, we form our primary hypothesis that implicit self-bivalence is cross-culturally generalizable. Indeed, the attachment behavioral system is theorized to serve as the basis for the development of one's self-representations (Baldwin, 1994; Baldwin et al., 1990; Bowlby, 1969, 1973; Pietromonaco & Barrett, 2000). Thus, to the extent that the affectively complex nature of interpersonal relationships and the processes by which they influence the development of self-representations are universal (Andersen & Chen, 2002; Baldwin, 1994; Baldwin et al., 1993; Bowlby, 1973; Cooley, 1983; Leary, 2005; Lee et al., 2024; Mead, 1934; Sullivan, 1953; Zayas et al., 2022; Zayas & Shoda, 2015), implicit self-bivalence should also be generalizable beyond a Western culture, consistent with the idea that implicit bivalence is a basic feature of self-representations.

Does the Strength of Implicit Self Evaluations Vary Cross Culturally?

Note that our primary hypothesis—that implicit self-bivalence is a fundamental feature of self-representations—can coexist with the possibility that the magnitude of implicit self-bivalence, or the separate evaluative dimensions of implicit self-positivity and self-negativity, may differ across cultures. Drawing from the literature on cross-cultural differences in self-construal, self-enhancement, and dialectical thinking, below we consider possible ways in which cross-cultural differences may emerge.

First, cultures differ in self-construal (Markus & Kitayama, 1991). Individuals from Western cultures are characterized by an independent self-construal—the view that the self is a separate, autonomous entity of distinct attributes. In contrast, individuals from Eastern cultures, particularly East Asia, are characterized by an interdependent self-construal—the view that the self is embedded in relationships with others. These cultural differences in self-construal promote differences in self-enhancement versus self-criticism motivations and tendencies. In Western

cultures that are organized in a way that fosters and encourages independent construal of the self, people tend to have a stronger self-enhancement motivation, with general sensitivity toward positive self-relevant information (Falk & Heine, 2015). For example, research has shown that American participants tend to attribute their success to internal and stable attributes while discounting their failure to either external factors or internal but unstable factors (Miller & Ross, 1975). In contrast, East Asian individuals tend to be more self-critical—being sensitive to negative self-relevant information (Falk & Heine, 2015). Such differences in self-enhancement versus self-criticism motivations and tendencies have been typically observed in their self-reports. But this raises the possibility that these motivations and tendencies may influence people's implicit self-evaluations, leading individuals from East Asian cultures to have greater implicit self-negativity.

Relatedly, cultural differences in reasoning about contradiction and tolerance for ambiguity may lead to differences in the degree of ambivalence associated with people's self-representations (Hamamura et al., 2008; Spencer-Rodgers et al., 2004). Folk epistemologies in East Asian cultures tend to promote dialectical thinking—tolerating, rather than eschewing psychological contradiction (Peng & Nisbett, 1999). In contrast, Western folk epistemologies tend to prefer synthetic thinking—searching to resolve any incongruity (Peng & Nisbett, 1999). Such cultural differences in dialectical thinking could influence how individuals from different cultures view the self (Hamamura et al., 2008; Spencer-Rodgers et al., 2004). In line with the dialectical thinking tendency, East Asians have been found to be more likely to acknowledge and accept contradictory appraisals of the self and exhibit greater ambivalence both in explicit self-esteem measures and spontaneous self-descriptions (Choi & Choi, 2002; Spencer-Rodgers et al., 2004). However, little is known about whether and how dialectical thinking is related to implicit self-bivalence.

Despite these cultural differences in self-related processes, as stated at the outset, research has consistently found comparable levels of implicit self-positivity across Eastern and Western cultures (Cai et al., 2007; Falk et al., 2009; Kitayama & Karasawa, 1997; Kitayama & Uchida, 2003; Kobayashi & Greenwald, 2003; Sedikides et al., 2015; Yamaguchi et al., 2007; see Falk & Heine, 2015 for a review). However, an important empirical question is whether individuals from different cultures are similar to or different from each other in implicit self-*bivalence* or its individual evaluative components.

Past work has used relative measures of implicit evaluations that tap into a singular dimension of whether a person evaluates themselves more positively versus negatively. But no study has used more absolute measures that tap into two dimensions: one that taps into whether a person evaluates themselves positively versus not, and a second dimension that taps into whether a person evaluates themselves negatively versus not. Thus, cross-cultural differences in implicit self-evaluations may have been obscured in past empirical work. Consider the scenario where participants from two cultures are assessed on implicit self-positivity and implicit self-negativity separately, with 10 reflecting a strong association and 0 reflecting a weak association. Participants from one culture show both a high level of implicit self-positivity (with an average scoring of 9) and a high level of implicit self-negativity (with an average scoring of 8). Participants from a different culture shows a moderate level of implicit self-positivity (with an average scoring of 5) and a moderate level of implicit self-negativity (with an average scoring of 4). Using relative measures, researchers may simply draw the conclusion that participants from the two cultures demonstrate similar levels of overall net implicit self-positivity. Yet, with absolute measures, nuances in the cognitive architecture of self-representations between the two cultures may be revealed. By separately assessing both implicit self-positivity and self-negativity, the current research allows us to compare the magnitude of implicit self-positivity and self-negativity across the two cultures.

The Present Research

As a test of the of implicit self-bivalence beyond a Western sample, we collected data from two samples, one from the United States (a replication of Zayas et al., 2022) and one from China. For participants in the Chinese sample, all the instructions and materials were translated into Chinese. Author M.N. translated all the instructions and materials, which were back translated into English by two research assistants. Authors V.Z. and R.T.L. checked the back-translation to ensure that the meaning was comparable with the original English instructions and materials. See the Methods section for details.

In both samples, participants were asked to complete the EPT to assess the extent to which the self automatically elicits both positive *and* negative implicit evaluations. As a within-person comparison, we also assessed the implicit evaluations elicited by a personally significant, liked attitude object. Including the personally significant, liked object allowed us to replicate the standard priming effect with common objects (Fazio et al., 1986) and thus provide validation of our methods.

We expected self-primes to elicit bivalent implicit evaluations in both samples and expected the liked-object primes to elicit only positive implicit evaluations (Fazio et al., 1986; Zayas et al., 2022; Zayas & Shoda, 2015). We further explored whether there were differences in the two samples on implicit self-positivity, implicit self-negativity, or both. In addition, in both samples, for exploratory purposes, we collected self-report measures of explicit self-evaluations and personality traits relevant to the self-construct, including explicit self-esteem, narcissism, experienced self-ambivalence, as well as independent and interdependent self-construal. For participants in the Chinese sample, we additionally collected a measure of acculturation. Given that the focus of the present work is on examining the normative structure of implicit self-evaluations across cultures and not on exploring individual differences that predict implicit self-evaluations, which would require a larger sample size to achieve sufficient statistical power, we do not explore these measures further. Details of the self-report measures are provided in Supplemental Materials.

For data collection in both samples, we obtained approval from the Institutional Review Board for human participants (Protocol No. 0907000558). Each sample had its own preregistration. We first preregistered and collected data for the Chinese sample to extend and replicate initial evidence of implicit self-bivalence documented in U.S. samples reported by Zayas et al. (2022). Following the data collection for the Chinese sample, we preregistered and collected data for the U.S. sample reported here. We chose to collect new U.S. data, instead of relying on previously collected data from Zayas et al. (2022), to ensure methodological and temporal consistency across samples (e.g., both using online data collection procedures between 2020 and 2022), allowing for greater comparability. We presented the methods and results of the U.S. sample first as it served as a replication of prior work. The preregistrations, deidentified data, and analysis code (in R) for the combined data involving both samples are available on OSF (https://osf.io/amzq4/).

Method

Participants

The U.S. Sample. One hundred and thirty-two participants from a northeastern university in the United States signed up for the study from the university's psychology research participant pool, in exchange for extra course credit. Six participants did not complete the study and were excluded. Following the exclusion criteria specified in the preregistration, we excluded one participant who performed more than 3 SDs below the mean accuracy. Thus, the final U.S. sample consisted of 125 individuals (76.0% female; M_{age} = 20.16 years, SD_{age} = 3.63 years). Participants

self-identified their racial and ethnic identities as follows: 46.40% Asian, 40.80% White, 6.40% Black or African American, 5.60% Multi-Racial, 0.80% Other; 8.80% Hispanic/Latino.

The Chinese Sample. One hundred and ten participants from China were recruited through social media, in exchange for an opportunity to be entered in a lucky draw to win a gift card.[1] The recruitment advertisement was posted in Chinese. As mentioned before, participants also completed all study materials in Chinese. Five participants did not complete the study and were excluded. Following exclusion criteria specified in the preregistration, we excluded one participant who performed more than 3 SDs below the mean accuracy. Thus, the final Chinese sample consisted of 104 individuals (65.38% female, 30.77% male, 3.85% did not answer; M_{age} = 19.51 years, SD_{age} = 5.33 years). 97.12% of participants self-identified as Asian (2.88% participants did not answer the question about racial identity). The ethnicity question asking participants to indicate whether they were Hispanic/Latino or not was omitted in the Chinese sample as it was not applicable. The majority of participants (77.88%) reported living in China at the time of participating in the study and the rest of them reported living in the United States (12.50%), the United Kingdom (1.92%), Canada (0.96%), or South Korea (0.96%). The remaining participants (5.77%) did not answer this question. Among participants who reported living in China, they reported coming from 19 (out of 34) different provincial-level administrative divisions and 30 different cities.

Statistical Power

For each sample, the experimental design was a 2 (prime type: self vs. liked-object) × 2 (target valence: positive vs. negative) within-subjects ANOVA. Within each sample, two main statistical tests were conducted to provide evidence of the primary hypothesis of implicit self-bivalence. Following our preregistrations, first, a one-sample *t*-test against zero was used to examine the extent to which the self-prime facilitated negative target categorization compared with the neutral prime. In addition, a paired sample *t*-test was used to examine the extent to which the self-prime facilitated negative target categorization more than the liked-object prime, which was expected to elicit univalent positive implicit evaluations. Sensitivity analyses revealed that with the current sample sizes, we can detect a small to medium effect size (Cohen's d = 0.32 and Cohen's d = 0.36, respectively) with 95% statistical power for the one-sample *t*-test and the paired sample *t*-test (alpha = .05, two-tailed).[2] For exploratory analyses comparing each facilitation-inhibition score between the two cultures, sensitivity analyses indicated that we could detect a small to medium effect size of Cohen's d = 0.37 using independent samples *t*-tests with 80% statistical power (alpha = .05, two-tailed).

Measures and Procedure

Overview. The measures and procedure for the U.S. sample and the Chinese sample were the same, unless otherwise specified. The study was conducted online, programmed using Inquisit 6 software (www.millisecond.com). Participants received a link to the Inquisit study and could complete the study on their own computer device, at a time and location of their convenience. Upon granting consent, participants were given instructions about the EPT, quizzed to confirm their understanding of the task, and once they passed the quiz, completed the EPT. Following the EPT, participants completed self-report measures (e.g., explicit self-esteem, narcissism, independent and interdependent self-construal). Given that these self-report measures were not central to the aims of the work, we report the details of how they were measured in Supplemental Materials. To give readers more information about the two samples' characteristics, we report the descriptive and inferential statistics of these self-report measures in the two samples in Table 1.

Table 1. Descriptive Statistics of Self-Report Measures in the U.S. and Chinese Samples and Inferential Statistics Comparing Them.

	US sample	Chinese sample	*Contrast statistics*			
Self-report measure	*M (SD)*	*M (SD)*	*t*	*df*	*p*	*d*
Explicit self-esteem[a]	37.31 (10.73)	37.60 (10.50)	-0.20	211.13	.845	0.03
Narcissism[b*]	2.92 (0.79)	3.30 (0.84)	-3.40	203.99	< **.001**	0.46
Independent self-construal[c]	40.68 (9.43)	44.58 (7.05)	-3.51	216.33	< **.001**	0.47
Interdependent self-construal[c]	47.13 (7.88)	44.75 (8.41)	2.15	203.63	**.033**	0.29
Acculturation to Chinese culture[d]	–	6.79 (1.13)	–	–	–	–
Acculturation to Western culture[d]	–	5.74 (1.25)	–	–	–	–
Explicit feelings toward *self*						
Positive evaluation[d]	6.59 (1.45)	6.03 (1.60)	2.76	209.94	**.006**	0.37
Negative evaluation[d]	4.78 (1.88)	4.36 (1.82)	1.71	222.07	.088	0.23
Structural ambivalence	3.04 (2.49)	2.65 (2.22)	1.23	225.93	.222	0.16
Experienced ambivalence[d]	5.28 (1.69)	5.66 (1.52)	-1.78	225.63	.076	0.24
Feeling thermometer[e]	67.56 (17.99)	64.81 (20.08)	1.08	209.08	.280	0.14
Explicit feelings toward *positive object*						
Positive evaluation[d]	7.83 (0.94)	7.35 (1.43)	2.98	172.16	**.003**	0.40
Negative feeling[d]	2.15 (1.11)	2.52 (1.50)	-2.07	185.81	**.040**	0.28
Structural ambivalence	-0.69 (1.87)	0.07 (2.51)	-2.53	187.07	**.012**	0.34
Feeling thermometer[e]	86.07 (13.34)	83.77 (13.58)	1.28	215.98	.202	0.17

Note. See *Method* for details.
[a]0–60 scale; [b]0–6 scale; [c]0–72 scale; [d]1–9 scale; [e]0–99 scale.
*The scales used to measure narcissism in the U.S. sample and the Chinese sample are different.

After that, demographic information was collected. Participants in both samples were asked to report their age, gender, year in school, and race. Participants in the U.S. sample were additionally asked about their ethnicity (i.e., Hispanic/Latino or not, which does not apply for participants in China). Participants in the Chinese sample were additionally asked about if they have had any study-abroad experience, and if so, in which country. They were also asked about where they were currently living (i.e., country, state/province, and city). We did not ask the U.S. participants to report current residence of living because they were students enrolled at a northeastern university in the United States and therefore were all living in the United States. At the end of the study, participants were debriefed.

Generating Prime Stimuli for Evaluative Priming Tasks (EPT). To generate the self-prime, we followed procedures in previous work (Zayas et al., 2022; Zayas & Shoda, 2015). Specifically, participants were prompted with the instruction "please type in your first name or the name that you most associate with yourself." To generate the liked-object prime, we asked participants to "think about an object that you have very strong positive feelings and reactions towards." We further clarified that

> the "object" could refer to a personal possession that you feel fondly towards, an activity that you feel strongly and positively about, a strong food preference (e.g., you may really enjoy chocolate), or a significant place (e.g., your hometown). The only things that we would like you not to include as "objects" are actual people in your life (e.g., a romantic partner, parent).[3]

They were then asked to "type in the word that will remind you of this positive significant object."

Assessing Implicit Self and Object Evaluations: EPT. Participants then completed the EPT (Fazio et al., 1986). After presenting participants with instructions of the EPT, we asked them to answer four multiple-choice questions as a comprehension check. Participants had to answer all the four questions correctly before proceeding. Otherwise, they would be presented with the instructions again. For details, see Supplemental Materials.

After passing the comprehension check, participants proceeded to the EPT. In the EPT, each trial consisted of the following three-step sequence: (a) a prime (in uppercase letters) for 200 ms, (b) a blank screen for 100 ms, and (c) a target word (in lowercase letters). The target remained on the screen until participants categorized the target as either "pleasant" or "unpleasant" by pressing one of the two response keys. Following Fazio et al.'s (1986) original procedures, participants were instructed to hold the first word (prime) in memory and to recite it aloud after categorizing the target word as quickly and accurately as possible. They were told that audible responses would be monitored for accuracy.

Participants first completed a practice block of 16 trials in which the primes were neutral words ("GLOVE" or "SCARF" for participants in the U.S. sample; "手套" or "剪刀," which mean "GLOVE" and "SCISSORS," respectively, for participants in the Chinese sample). Participants repeated the practice block until their average response time was under 1,000 ms and their average accuracy rate was above 80%.

In the data-collection blocks, there were three types of primes: the participant's own name (i.e., self-prime), the name of the personally significant positive object (i.e., liked-object prime), and a neutral object name (i.e., neutral prime which serves as a baseline). In the U.S. sample, the neutral prime was either "TABLE" or "PENCIL," counterbalanced across participants. In the Chinese sample, the neutral prime was either "桌子" or "橱柜" (which mean "TABLE" and "CUPBOARD," respectively), counterbalanced across participants. We selected the two English words from the Affective Norms for English Words (ANEW) list (Bradley & Lang, 2010) and the

two Chinese words from a list of affective words in Chinese (Ho et al., 2015) with the goal of them being neutral in affective valence (with the mean rating close to the mid-point of 5 on a scale of 1 to 9) and having relatively high rater agreements (with the standard deviation for affective valence rating being relatively low). The two neutral primes we selected for each sample had the same number of syllables (in the U.S. sample) or characters (in the Chinese sample). As in previous work (Zayas et al., 2022; Zayas & Shoda, 2015), target words in the study were either positive (*i.e., honor, lucky, diamond, loyal, freedom, rainbow, love, honest, peace, heaven*) or negative (*i.e., evil, cancer, sickness, disaster, poverty, vomit, bomb, rotten, abuse, murder*) in valence. Therefore, the EPT had a 3 prime type (self-prime vs. liked-object prime vs. neutral prime) × 2 target valence (positive vs. negative) within-subjects design. The EPT consisted of three data-collection blocks, each with 48 trials. In each block, each unique combination of prime type and target valence was presented an equal number of times. We counterbalanced the response key assignment ("A" and "L"), such that half of the participants in each sample performed the EPT with "pleasant" assigned to the left response key and "unpleasant" to the right, and the remaining half of the sample performed the EPT with the opposite response key assignments.

Data Reduction and Analyses

Data Reduction. We followed standard data reduction procedures for reaction time (RT) in previous work (Greenwald et al., 1998; Zayas et al., 2022; Zayas & Shoda, 2015). First, we excluded trials that were incorrectly categorized and trials with RTs that were outside the normal range of time needed to categorize a single trial (<150 ms or >4,999 ms). On the remaining trials, those with RTs below 300 ms and above 3,000 ms were recoded to 300 ms and 3,000 ms, respectively. RTs were transformed into log units and all analyses were performed on log-transformed RTs (RTs transformed back to milliseconds from log-transformed units were reported for illustrative purposes). In the U.S. sample, participants had average RTs of 715 ms (range: 419–1,265 ms) and average accuracy of 96% (range: 83%–100%). In the Chinese sample, participants had average RTs of 733 ms (range: 458–1,273 ms) and average accuracy of 97% (range: 81%–100%). Table 2 presents the descriptive statistics of participants' reaction times and accuracy rates in the six types of trials (3 prime types: self-prime vs. liked-object prime vs. neutral prime × 2 target valences: positive vs. negative) in the two samples.

Following previous work using the EPT (Fazio et al., 1986), we measured the extent to which each valenced prime (self and liked-object primes) facilitated (vs. inhibited) the categorization of positive and negative targets, respectively, compared with the neutral (baseline) prime. To calculate the facilitation-inhibition scores of positive targets, we subtracted the average RT for the valenced prime (self or liked-object) → positive target trials from the average RT for the neutral prime → positive target trials. Similarly, to quantify facilitation-inhibition scores of negative targets, we subtracted the average RT for the valenced prime (self or liked-object) → negative target trials from the average RT for the neutral prime → negative target trials. Thus, each participant was characterized by four facilitation-inhibition scores, one for each prime type by target valence combination.

Data Analytic Strategy. For all analyses, we used the facilitation-inhibition scores as the dependent variable. Following our preregistrations to examine whether the self is associated with both positive and negative implicit evaluations in the U.S. sample and the Chinese sample, we conducted one-sample *t*-tests against zero for each sample to examine the extent to which the self-prime and the liked-object prime facilitated positive and negative target categorization compared with the neutral prime. Our focal analysis was a paired sample *t*-test that examined the extent to which the self-prime facilitated negative target categorization more than the liked-object prime. For completeness, we also report the analogous paired sample *t*-test on facilitation-inhibition

Table 2. Means and Standard Deviations of Reaction Times and Accuracy Rates by Trial Type in the EPT for the U.S. and Chinese Samples.

	U.S. sample				Chinese sample			
	RT		Accuracy		RT		Accuracy	
Trial type	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Self-prime positive target trials	662	143	97%	4%	690	162	97%	6%
Self-prime negative target trials	707	166	95%	6%	717	155	97%	4%
Liked-object prime positive target trials	711	175	96%	5%	729	205	98%	4%
Liked-object prime negative target trials	738	184	95%	6%	753	186	96%	6%
Neutral prime positive target trials	735	184	94%	7%	756	203	96%	5%
Neutral prime negative target trials	735	183	96%	6%	751	193	98%	4%

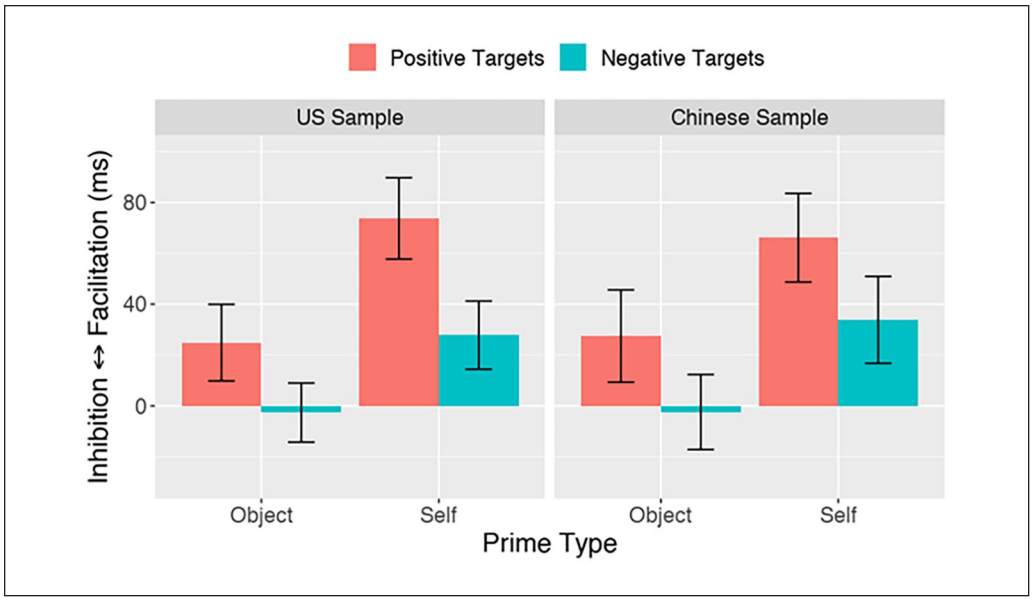

Figure 1. Bars Represent Facilitation-Inhibition Scores (Milliseconds) for Liked-Object Primes and Self-Primes as a Function of Target Valence (Positive vs. Negative) for the U.S. Sample (Left Panel) and the Chinese Sample (Right Panel).

Note. Error bars represent 95% CI. Positive numbers indicate that primes facilitated categorization of target words, while negative numbers indicate that primes inhibited categorization of target words. Means and SDs used to derive the 95% CIs were computed using log-transformed RTs and transformed back to milliseconds for illustration.

scores of positive targets, and the results of a repeated-measures ANOVA with prime type (self vs. liked-object) and target valence (positive vs. negative) as the within-subjects factors.

As mentioned earlier, although not a key focus of the current research, we explored cross-cultural differences in the *magnitude* of implicit self-positivity and implicit self-negativity. Thus, in addition to our preregistered analyses, we combined data from both samples and added culture (U.S. vs. China) as a between-subjects factor into the ANOVA model to examine whether there was a main effect or any interaction effect involving culture. We also conducted planned pairwise comparisons for each of the four facilitation-inhibition scores (2 prime types: self vs. liked-object $\times$ 2 target valences: positive vs. negative) between the two cultures.

Results

Implicit Self- and Object-Evaluations

The U.S. Sample. Results from the U.S. sample are illustrated in Figure 1 (left panel). Replicating past findings (Zayas et al., 2022), liked-object primes facilitated the categorization of positive targets, $M = 25$ ms, $SD = 85$ ms, 95% CI [10 ms, 40 ms], $t(124) = 3.69$, $p < .001$, $d = 0.33$, and did not significantly facilitate or inhibit the categorization of negative targets, $M = -3$ ms, $SD = 66$ ms, 95% CI [−14 ms, 9 ms], $t(124) = -0.50$, $p = .617$, $d = 0.04$.[4] Critically, the result pattern for self-primes was different. Self-primes significantly facilitated the categorization of positive targets, $M = 74$ ms, $SD = 90$ ms, 95% CI [58 ms, 90 ms], $t(124) = 9.89$, $p < .001$, $d = 0.88$, but, importantly, self-primes also significantly facilitated the categorization of negative targets, $M = 28$ ms, $SD = 76$ ms, 95% CI [14 ms, 41 ms], $t(124) = 4.06$, $p < .001$, $d = 0.36$.

Directly comparing the evaluations elicited by the self to those elicited by a liked object revealed that self-primes facilitated the processing of negative targets to a greater extent than liked-object primes, $M_{diff} = 30$ ms, $SD_{diff} = 82$ ms, 95% CI [16 ms, 45 ms], *paired* $t(124) = 4.12$, $p < .001$, $d = 0.37$. Self-primes also facilitated the categorization of positive targets to a greater extent than liked-object primes, $M_{diff} = 49$ ms, $SD_{diff} = 88$ ms, 95% CI [33 ms, 64 ms], *paired* $t(124) = 6.44$, $p < .001$, $d = 0.58$.

The repeated-measures ANOVA predicting the facilitation-inhibition scores showed that there was a statistically significant main effect of prime type, $F(1, 124) = 51.71$, $p < .001$, $\eta_p^2 = .29$, reflecting that participants were faster to categorize target words following self- versus liked-object primes. There was also a statistically significant main effect of target valence, $F(1, 124) = 25.00$, $p < .001$, $\eta_p^2 = .17$, reflecting that participants were faster to categorize positive (vs. negative) target words. Finally, the prime type by target valence interaction was statistically significant, $F(1, 124) = 4.12$, $p = .044$, $\eta_p^2 = .03$, suggesting that the difference in positive versus negative implicit evaluations was greater for self-primes than for liked-object primes.

Next, for descriptive purposes (Tukey, 1977), we examined the prevalence of possessing negative implicit self-evaluations within the U.S. sample. Specifically, we classified each person as possessing (or not) negative self-evaluations according to whether their facilitation-inhibition scores for negative targets were greater than zero (coded as 1, reflecting evidence of negative self-evaluations) or were equal to or less than zero (coded as 0, reflecting absence of negative self-evaluations). As a comparison, using an analogous approach, we classified each person as possessing (or not) negative evaluations for the self-nominated liked object. Self-primes facilitated the categorization of negative targets for 63.2% of the U.S. sample, which was significantly higher than 50%, $\chi^2(1) = 8.19$, $p = .004$, 95% CI [54.1%, 71.5%]. In contrast, liked-object primes facilitated the categorization of negative targets for 51.2% of the U.S. sample, which was not statistically significantly different from a 50/50 distribution, $\chi^2(1) = 0.03$, $p = .858$, 95% CI [42.1%, 60.2%]. A McNemar test showed that the percentage of possessing negative implicit evaluations was significantly different for the self and the liked object, $\chi^2(1) = 5.60$, $p = .018$.

The Chinese Sample. Results from the Chinese sample are illustrated in Figure 1 (right panel). Similar to results in the U.S. sample, liked-object primes facilitated the categorization of positive targets, $M = 28$ ms, $SD = 93$ ms, 95% CI [9 ms, 46 ms], $t(103) = 3.68$, $p < .001$, $d = 0.36$, and did not significantly facilitate or inhibit the categorization of negative targets, $M = -3$ ms, $SD = 76$ ms, 95% CI [−17 ms, 12 ms], $t(103) = -0.60$, $p = .549$, $d = 0.06$. Critically, the result pattern for self-primes was different from that of liked-object primes, as in the U.S. sample. Self-primes significantly facilitated the categorization of positive targets, $M = 66$ ms, $SD = 90$ ms, 95% CI [49 ms, 84 ms], $t(103) = 7.78$, $p < .001$, $d = 0.76$, but, importantly, self-primes also significantly facilitated the categorization of negative targets, $M = 34$ ms, $SD = 88$ ms, 95% CI [17 ms, 51 ms], $t(103) = 3.54$, $p < .001$, $d = 0.35$.

Similarly, we then directly compared the evaluations elicited by the self to those elicited by a liked object. Self-primes facilitated the processing of negative targets to a greater extent than liked-object primes, *paired* $t(103) = 4.83$, $p < .001$, $M_{diff} = 36$ ms, $SD_{diff} = 74$ ms, 95% CI [22 ms, 51 ms], $d = 0.47$. Self-primes also facilitated the categorization of positive targets to a greater extent than liked-object primes, *paired* $t(103) = 3.64$, $p < .001$, $M_{diff} = 39$ ms, $SD_{diff} = 106$ ms, 95% CI [18 ms, 59 ms], $d = 0.36$.

Repeated-measures ANOVA showed that there was a statistically significant main effect of prime type, $F(1, 103) = 26.42$, $p < .001$, $\eta_p^2 = .20$, reflecting that participants were faster to categorize target words following self- versus liked-object primes. There was also a statistically significant main effect of target valence, $F(1, 103) = 15.59$, $p < .001$, $\eta_p^2 = .13$, reflecting that

participants were faster to categorize positive (vs. negative) target words. The prime type by target valence interaction was not statistically significant, $F(1, 113) = 0.02$, $p = .879$, $\eta_p^2 < .001$.

As in the analyses for the U.S. sample, we examined the prevalence of possessing negative implicit self-evaluations within the Chinese sample. Self-primes facilitated the categorization of negative targets for the majority of participants (60.6%), which was significantly different from a 50/50 distribution, $\chi^2(1) = 4.24$, $p = .039$, 95% CI [50.5%, 69.9%]. In contrast, liked-object primes facilitated the categorization of negative targets for 47.1% of the sample, which was not statistically significantly different from a 50/50 distribution, $\chi^2(1) = 0.24$, $p = .624$, 95% CI [37.3%, 57.1%]. A McNemar test showed that the percentage of possessing negative implicit evaluations was significantly different for the self and the liked object, $\chi^2(1) = 6.04$, $p = .014$.

Cross-Cultural Comparisons

As shown above, results from both the U.S. sample and the Chinese sample showed that the self spontaneously activates both positive and negative implicit evaluations. In contrast, a personally significant, liked object only activates positive evaluations, but not negative evaluations. We then statistically tested whether implicit self- and implicit liked-object evaluations were significantly different between the two cultures. To do so, we combined the data from both samples and conducted a mixed-measures ANOVA with culture (US vs. China) as a between-subjects factor, prime type (self vs. liked-object) and target valence (positive vs. negative) as within-subjects factors, and the facilitation-inhibition scores as the dependent variable. Results showed that there was no statistically significant main effect ($p = .817$, $\eta_p^2 = .0002$) or higher-order interactions involving culture (ps $\geq .189$, $\eta_p^2 \leq .008$). Planned pairwise comparisons showed no statistically significant differences between the two cultures in any of the four facilitation-inhibition scores (ps $\geq .316$, ds $\leq .13$).

Robustness Check

Could our findings—showing that the self triggers implicit bivalence in the United States as well as the Chinese sample—partly reflect similarities in participants' characteristics? In particular, some participants in the Chinese sample reported studying abroad. If so, one could argue that they may be more similar to those in the U.S. sample than to individuals born and raised in China without such experiences.

Given that participants in the Chinese sample completed all study materials in Chinese—including self-report measures and the EPT, which requires fast and accurate responding—this provides strong evidence that they were at least fluent, if not native, in Chinese. Given that language is closely tied to cultural beliefs, their proficiency implies a greater likelihood of possessing beliefs associated with Chinese culture. Nonetheless, to conduct a more conservative test of whether implicit self-bivalence emerges in a non-Western sample, we created two subgroups: Chinese participants with study-abroad experience ($n = 53$) and those without ($n = 45$). For completeness, we also examined the potential influence of Eastern cultural background on implicit self-bivalence in the U.S. sample by creating two additional subgroups: U.S. participants who self-identified as Asian ($n = 58$) and those who did not ($n = 67$). We reran our key analyses to test whether implicit self-bivalence emerged within each of these four subgroups.

As shown in Figure 2, across all four subgroups, we find evidence of implicit self-bivalence. Self-primes significantly facilitated the categorization of both positive and negative targets compared with neutral primes (ps $\leq .032$, ds ≥ 0.28, one-tailed tests). Critically, self-primes significantly facilitated the categorization of negative targets to a greater extent than the liked-object primes in all the four subgroups (ps $\leq .046$, ds ≥ 0.30, two-tailed tests). These results suggest that implicit self-bivalence was demonstrated in both subgroups of the Chinese sample—regardless of

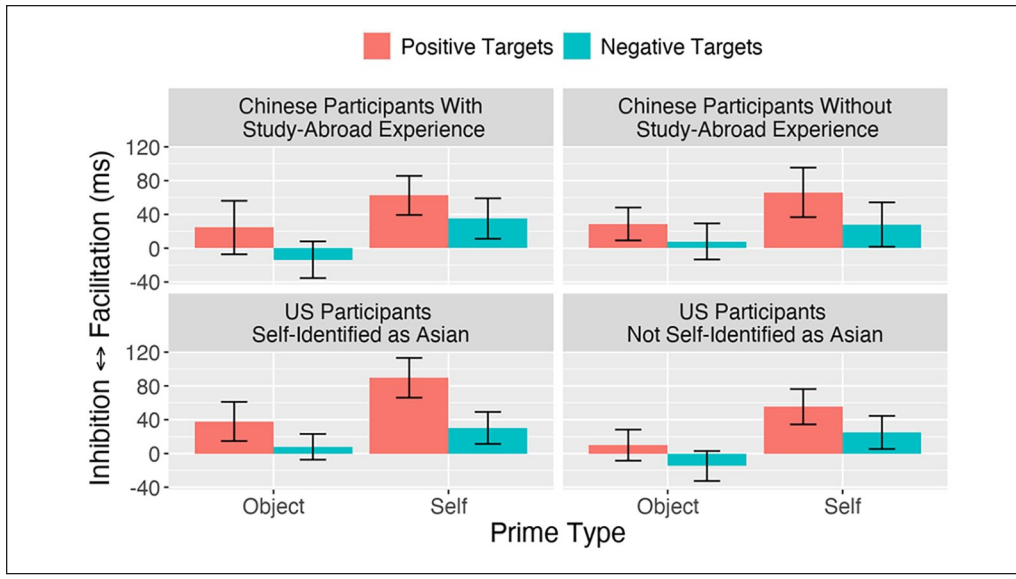

Figure 2. Bars Represent Facilitation-Inhibition Scores (Milliseconds) for Liked-Object Primes and Self-Primes as a Function of Target Valence (Positive vs. Negative) for the Four Subgroups: Chinese Participants Who Had Study-Abroad Experience (Upper Left Panel), Chinese Participants Who Did Not Have Study-Abroad Experience (Upper Right Panel), U.S. Participants Self-Identified as Asian (Lower Left Panel), and U.S. Participants Who Did Not Self-Identify as Asian (Lower Right Panel).
Note. Error bars represent 95% CI. Positive numbers indicate that primes facilitated categorization of target words, while negative numbers indicate that primes inhibited categorization of target words. Means and SDs used to derive the 95% CIs were computed using log-transformed RTs and transformed back to milliseconds for illustration.

study-abroad experience—and in both subgroups of the U.S. sample—regardless of whether participants self-identified as Asian or not Asian.

In addition, we revisited our exploration of cross-cultural differences in implicit self-evaluations by examining whether the subgroups differed significantly in their facilitation-inhibition scores. One-way ANOVAs showed that the four subgroups did not differ significantly in any facilitation-inhibition scores elicited by self-primes or liked-object primes ($ps \geq .065$, $\eta_p^2 \leq .032$). Pairwise comparisons between any of the two groups also showed no statistically significant differences in any facilitation-inhibition scores ($ps \geq .123$, $ds \leq .419$, Bonferroni corrected).

Discussion

The self is one of the most important constructs for understanding human behavior and experience. Despite the enormous amount of research across psychological disciplines and beyond, questions about the basic structure of the self-representation remain. Namely, although research in social cognition finds that implicit self-evaluations are overwhelmingly positive, recent work provides new evidence that the self is associated with both positive and negative implicit evaluations (Zayas et al., 2022). Such findings raise a host of new unanswered questions, one of which is the extent to which implicit self-bivalence is generalizable beyond a U.S. sample.

The current research is the first to demonstrate that implicit self-bivalence—the phenomenon that the self is associated with both positive and negative implicit evaluations—is not specific to Western samples. Previous work (Zayas et al., 2022) only examined implicit self-bivalence with

U.S. participants. The current work not only replicates the findings with a U.S. sample, but importantly also extends it to a Chinese sample. Specifically, we translated the instructions and materials of the EPT in Chinese (and validated the translation through back-translation), and used this Chinese version of the EPT to demonstrate implicit self-bivalence in a Chinese sample.

In both samples, we find that for the majority of the participants, self-primes facilitate the categorization of negative targets (63.2% in the U.S. sample and 60.6% in the Chinese sample, respectively). The prevalence of possessing negative implicit self-evaluations is comparable to the prevalence found in previous studies (Study 1: 63.4%; Study 2: 60.7%; Study 3: 65%; Zayas et al., 2022). In addition, we find that the extent to which the self is associated with implicit positivity and implicit negativity was not statistically significant across participants from the two cultures.

The fact that the current work finds empirical evidence of implicit self-bivalence in a Chinese sample, and that the magnitudes of implicit positive and negative self-evaluations were statistically indistinguishable provides the first empirical support for the hypothesis that the coexistence of positive and negative implicit self-evaluations reflects a normative, basic feature of the personality system.

Initial Evidence of Generality

The present work is motivated by testing the proposition that the structure of the self is linked with positive and negative evaluations, and that this is a basic universal feature of the mind. Finding evidence of implicit self-bivalence in a Chinese sample provides initial evidence that implicit self-bivalence is not a phenomenon that is unique to Western cultures.

But one could argue that perhaps the reason that we observe implicit self-bivalence in the Chinese sample is because the sample characteristics may have been similar to those in the U.S. sample. If the samples are similar, then it is not surprising that we find evidence of implicit self-bivalence in the Chinese sample. However, we do not believe this is the case for several reasons.

First, our Chinese participants were recruited in China with advertisements in Chinese and completed all study materials in Chinese. Completing the EPT in Chinese is not a trivial task; it requires fast and accurate responding, while holding the prime word in mind. This provides strong evidence that they were at least fluent, if not native, in Chinese. Language fluency is closely linked with cultural practices and norms. In addition, given that language affects the self-representation that is most accessible in a given moment, and that using the Chinese language among bilingual speakers activates more of a collectivistic mindset (see Ji et al., 2004; Lee et al., 2010; Wang et al., 2010), observing implicit self-bivalence in our Chinese sample is a particularly strong demonstration.

Second, we conducted what we believe is an overly conservative test of our primary question. We examined evidence of implicit self-bivalence in four subgroups: Chinese participants with or without study-abroad experience, and for completeness, U.S. participants self-identified as Asian or not Asian. All four subgroups showed evidence of implicit self-bivalence. Such a test is overly conservative because many of the Chinese participants with study-abroad experience likely identified with the Chinese culture, and many of our U.S. Asian participants likely did not. Even with this more conservative test, we found evidence of implicit self-bivalence in the four subgroups and did not find statistically significant difference among these groups in the magnitude of implicit self-positivity or implicit self-negativity.

Why Does Implicit Self-Bivalence Emerge in Both the U.S. and Chinese Samples?

The present work documents implicit self-bivalence in a Chinese sample, providing initial evidence that the phenomenon is not limited to a Western culture. Such a proposition is in line with

the social cognitive literature delineating pathways in which self-representations may acquire positive and negative implicit evaluations. Implicit evaluations may develop as a result of repeated experiences encountered over time (Wilson et al., 2000). Evaluations that are automatically triggered when one interacts with the stimulus gradually become etched in the memory structure associated with it. Implicit evaluations may also arise from propositional information, including those acquired through inferential reasoning (Cone & Ferguson, 2015; Gawronski & Bodenhausen, 2006; Kurdi & Dunham, 2020). Given that the construction, development, and maintenance of the self-representation occurs within dyadic contexts (Baldwin et al., 1990; Bowlby, 1969, 1973) and that relationships, especially significant ones, are characterized by their affective complexity (Murray, Griffin, et al., 2006; Murray & Holmes, 2009), the self should be characterized by implicit bivalence, being associated with both positive *and* negative implicit evaluations (Zayas et al., 2022).

We reason that implicit self-bivalence not only *reflects* the complexity of relationships, but like other memory structures, it affects how people perceive, interpret, and react to future situations (Sakman & Zayas, 2024). In this way, implicit self-bivalence may play a role in helping people navigate the ongoing and dynamic affective complexities that they routinely encounter in daily life (Zayas et al., 2022). Possessing both positive and negative implicit self-evaluations might enable individuals to detect, monitor, process, and respond to momentary instances of disconnection and misunderstanding that are common in social environments (Murray & Holmes, 2009). Having some negative implicit self-evaluations might alert individuals to instances (actual or imagined) of social disapproval or rejection. Although aversive, such attunement to social threats is a prerequisite to adjust one's behavior accordingly, and thus resume social belonging (Leary, 2005; Williams, 2009). Consistent with these ideas, relational and developmental perspectives highlight the interconnections between self-evaluations and feelings of belonging and how the ubiquity of social threats necessitates the need to manage instances of encountering actual or imagined negative self information (Baldwin, 1994; Baldwin et al., 1990; Leary, 2005; Leary & Baumeister, 2000; Murray, Griffin, et al., 2006; Williams, 2009).

Thus, to the extent that people inevitably and routinely encounter and think about both rewarding and threatening social interactions, which can trigger self-evaluations and encourage them to draw either positive or negative inferences about the self, the self-representation is likely to develop and maintain a bivalent structure. Findings from the current work showing that this implicit self-bivalence in both a U.S. and a Chinese sample is consistent with the view that some of the affective complexity experienced and learned about the self may be etched in the self-representation.

Are Implicit Positivity and Implicit Negativity Comparable Across Cultures?

Although we expected that individuals from China would show evidence of implicit self-bivalence, we were agnostic as to whether there would be cross-cultural differences in the magnitude of implicit self-positivity and implicit self-negativity. Some work using relative measures of implicit self-evaluations has not found appreciable differences across cultures (e.g., Cai et al., 2007; Falk et al., 2009; Yamaguchi et al., 2007). Yet, given work on cross-cultural differences in self-construal, self-enhancement, and dialectical thinking, we might expect cross-cultural differences in the magnitude of positive and negative implicit evaluations. Our results showed no evidence of statistically significant differences in the strength of either implicit positive or negative self-evaluations across cultures. There are several possibilities for such findings.

First, null findings are difficult to interpret because they may reflect insufficient statistical power to detect an effect. Although sensitivity analysis revealed that the current sample sizes allowed us to detect a small to medium effect of Cohen's $d = 0.37$ between the two cultures with 80% statistical power (assuming alpha = .05, two-tailed test), it is possible that any existing

cross-cultural differences are smaller than this threshold. Consequently, cross-cultural differences in implicit self-positivity, implicit self-negativity, or both may exist, but our sample sizes may have been too small to detect these relativity smaller differences.

Another possible reason for the absence of statistically significant differences between the two cultures is that the U.S. and Chinese sample may have been more similar than expected. However, when we created four subgroups (Chinese participants with or without study-abroad experience, and U.S. participants self-identified as Asian or not Asian), we found no statistically significant difference among the four subgroups in implicit positivity or implicit negativity. Sensitivity analysis showed that we had 80% power to detect a small to medium effect size of $f = 0.22$ (Cohen's $d = 0.44$), with the sample size $N = 223$ for the one-way, four group ANOVA. Nonetheless, we cannot rule out other ways that our U.S. and Chinese participants may have had overlapping experiences. Globalization and the rising use of social media have increased people's exposure to different cultures and expanded shared experiences, which means that individuals from different cultures will now have more common experiences (Arnett, 2002; Jenkins & Deuze, 2008).

Even if we assume that the magnitude of implicit self-positivity and implicit self-negativity is comparable across cultures, the current work still raises important questions about how implicit self-bivalence may be expressed differently across cultural contexts. That is, just like the attachment behavioral system is universal, but its expression is open to cultural influences, we reason that implicit bivalence is a universal feature of the self-representation, but its expression may vary depending on cultural influences.

From this perspective, implicit self-bivalence may generally help individuals navigate complex social dynamics, but different cultures may emphasize different aspects of these interactions. In individualistic cultures, which are characterized by an independent self-construal and emphasize separateness and autonomy (Markus & Kitayama, 1991; Triandis, 2001), individuals are motivated to accurately gauge their worthiness and competence compared with others. This requires the ability to make and detect positive and negative evaluations of the self, allowing individuals to monitor, maintain, and adapt cognitions and behavior, in the service of maintaining social standing. Individualistic cultures, marked by weaker normative constraints and greater expressive freedom, may give rise to self-bivalence through internal role conflict or cultural tolerance for psychological inconsistency (Gelfand et al., 2006; Triandis, 1995). In collectivistic cultures, which are characterized by an interdependent self-construal and emphasize connectedness and harmony (Markus & Kitayama, 1991; Triandis, 1995), individuals are motivated to be sensitive to the needs of others. Both positive and negative self-evaluations can independently serve as cues for individuals to adjust their behavior in ways that promote harmony and cohesion within groups. Similarly, Triandis (1995) proposed that collectivistic cultures foster context-dependent and multifaceted self-representations, which may naturally accommodate both positive and negative self-views. Despite cross-cultural differences in the priority of goals in social interactions, maintaining both positive and negative implicit self-evaluations may enable individuals to adaptively navigate and react to the dynamic, affectively complex, and ever-changing social environment.

Implications for Measuring Implicit Evaluations

The current work also provides further implications for the measurement of implicit evaluations. Various theoretical perspectives have acknowledged the affectively complex nature of our evaluations. For example, the Evaluative Space Model (Cacioppo & Berntson, 1994; Cacioppo et al., 1997) posits that evaluations of positivity and negativity reflect two distinct and separate neural systems. In early stages of perception, individuals engage in a series of rapid evaluations in

which positivity *and* negativity are assessed in parallel, before integrating them into behavior of either approach *or* avoidance.

Empirical research, however, does not fully take this theoretical backdrop into account. Although researchers have studied ambivalent attitudes toward a variety of targets (e.g., Alvarez & Brehm, 1995; Gallucci et al., 2020; Lipkus et al., 2001; Maio et al., 2000), ambivalence is typically conceptualized as an *explicit* phenomenon and operationalized through *explicit* measures (Newby-Clark et al., 2002). Despite this, widely adopted indirect measures of attitudes, such as the Implicit Association Test (Greenwald et al., 1998), assess *implicit* evaluations on a single evaluative dimension, ranging from negative on one end to positive on the other end. As a result, much work has assessed *relative* evaluations (positive vs. negative) and has not been well positioned to assess the extent to which the same object elicits *both* positive *and* negative evaluations.

Even with the EPT used in the current study, traditionally, researchers would derive a single index that reflects the extent to which the attitude object facilitates the categorization of positive targets relative to negative targets (Fazio et al., 1986; Wentura & Degner, 2010). If we were to apply the typical measurement in the current research, we would draw the conclusion in the present research that both the self and a personally significant positive object are associated with implicit positivity (i.e., they facilitated the categorization of positive targets more than negative objects). The nuance of the bivalent versus univalent structure of representations associated with the self versus liked objects would have gone unnoticed. The current work therefore suggests the importance of examining *both* positive *and* negative implicit associations triggered by the attitude object of interest to enrich our understanding of the underlying cognitive architecture.

Constraints on Generality and Future Research

The current research extends the phenomenon of implicit self-bivalence beyond a Western sample. However, there are important constraints on generality that should be considered (Simons et al., 2017). First, the current study involves only two samples, both composed of young adults. Therefore, while our findings suggest that implicit self-bivalence is not a feature of the mind unique to individuals from Western cultures and can be observed in an Eastern culture as well, we do not claim cultural universality. Future research should include participants from other geographic regions or cultural backgrounds, as well as a wider age range, to more fully assess the generalizability of implicit self-bivalence.

Relatedly, we did not find any statistically significant difference between our Chinese and U.S. participants in implicit self-evaluations. Although such findings are consistent with other work (e.g., Cai et al., 2007; Falk et al., 2009; Yamaguchi et al., 2007), future research should examine cross-cultural differences using larger samples of participants, as well as recruiting samples that are more culturally different from one another.

In addition, a promising route for future research is to examine how situational factors (e.g., failure, rejections, social comparisons) may influence state-level implicit evaluations triggered by the self, and how such effects may differ across cultures. It is possible, for example, that the experience of social exclusion might particularly influence the implicit self-negativity among participants from collectivistic cultures characterized by an interdependent self-construal (Uskul & Over, 2017).

Finally, future research should empirically test the functional significance of implicit self-bivalence. Given our reasoning above, interesting questions for future research include, for example, how does implicit self-positivity, implicit self-negativity, or both change as a function of the social interaction dynamics experienced at-the-moment? How do individual differences in implicit self-evaluations predict sensitivity to social rejections?

Concluding Remarks

Previous research has claimed that implicit self-positivity is a universal feature (see Yamaguchi et al., 2007). The current work replicates recent findings (Zayas et al., 2022) suggesting that the self is implicitly bivalent: the self elicits positive implicit evaluations, but it also simultaneously elicits some negative implicit evaluations. More importantly, it demonstrates for the first time that such implicit self-bivalence is not specific to Western samples but can be exhibited in an Eastern sample. The affectively complex, bivalent nature of self-representations may be a universal and robust characteristic of the human mind.

Data Availability Statement

Preregistrations, deidentified data, and analysis syntax (in R) are available at OSF (https://osf.io/amzq4/).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by U.S. National Science Foundation award 2234932 awarded to Vivian Zayas. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. National Science Foundation.

ORCID iDs

Minghui Ni https://orcid.org/0000-0002-8383-9412
Randy T. Lee https://orcid.org/0000-0003-3396-858X
Vivian Zayas https://orcid.org/0000-0002-9534-3721

Supplemental Material

Supplemental material for this article is available online.

Notes

1. For the study with Chinese participants, we expected to be limited in our data collection capacities. Thus, we preregistered that "we aim to have approximately 60 participants in the study, or as many participants as we can possibly run during a one-month period." However, with three waves of data-collection efforts, we were able to recruit a greater number of participants than originally expected and increased the sample size so that it was comparable to the sample size of the U.S. sample. Results were not modified by the wave of data collection. See *Supplemental Materials* for details.
2. Given our directional hypotheses that self-primes would *facilitate* negative target categorization and do so *more than* the liked-object primes, we could have conducted one-tailed tests (Lakens, 2016). But we decided to be more conservative and conducted two-tailed tests.
3. In both the U.S. and the Chinese sample, some participants used long phrases or animals as liked-object primes. In the Chinese sample, some participants typed in English as self-generated primes, even though the entire experimental instructions and materials were in Chinese. We report results of analyses including these participants per our preregistration plan. But considering the potential influence of such primes on participants' responses, we further excluded those participants ($n = 14$ from the U.S. sample, $n = 12$ from the Chinese sample). Focusing on data of the remaining participants, our results still held. See *Supplemental Materials* for details.

4. RTs and associated confidence intervals are displayed in milliseconds for illustrative purposes. However, all analyses were performed on log-transformed RTs following standard procedures (Bar-Anan & Nosek, 2014; Greenwald et al., 1998; Zayas & Shoda, 2015).

References

Ainsworth, M. D. S. (1978). The Bowlby-Ainsworth attachment theory. *Behavioral and Brain Sciences*, *1*(3), 436–438. https://doi.org/10.1017/S0140525X00075828

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*(1), 1–48. https://doi.org/10.1080/10463280802613866

Alvarez, R. M., & Brehm, J. (1995). American ambivalence towards abortion policy: Development of a heteroskedastic probit model of competing values. *American Journal of Political Science*, 1055–1082. https://doi.org/10.2307/2111669

Andersen, S. M., & Chen, S. (2002). The relational self: An interpersonal social-cognitive theory. *Psychological Review*, *109*(4), 619–645. https://doi.org/10.1037/0033-295X.109.4.619

Andersen, S. M., & Glassman, N. S. (1996). Responding to significant others when they are not there: Effects on interpersonal inference, motivation, and affect. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition, Vol. 3. The interpersonal context* (pp. 262–321). The Guilford Press.

Arnett, J. J. (2002). The psychology of globalization. *American Psychologist*, *57*(10), 774–783. https://doi.org/10.1037/0003-066X.57.10.774

Baldwin, M. W. (1994). Primed relational schemas as a source of self-evaluative reactions. *Journal of Social and Clinical Psychology*, *13*(4), 380–403. https://doi.org/10.1521/jscp.1994.13.4.380

Baldwin, M. W., Carrell, S. E., & Lopez, D. F. (1990). Priming relationship schemas: My advisor and the Pope are watching me from the back of my mind. *Journal of Experimental Social Psychology*, *26*(5), 435–454. https://doi.org/10.1016/0022-1031(90)90068-W

Baldwin, M. W., Fehr, B., Keedian, E., Seidel, M., & Thomson, D. W. (1993). An exploration of the relational schemata underlying attachment styles: Self-report and lexical decision approaches. *Personality and Social Psychology Bulletin*, *19*(6), 746–754. https://doi.org/10.1177/0146167293196010

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*, 668–688. https://doi.org/10.3758/s13428-013-0410-6

Baumgardner, A. H. (1990). To know oneself is to like oneself: Self-certainty and self-affect. *Journal of Personality and Social Psychology*, *58*(6), 1062–1072. https://doi.org/10.1037/0022-3514.58.6.1062

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*(4), 631–643. https://doi.org/10.1037/0022-3514.79.4.631

Boucher, H. C. (2010). Understanding Western-East Asian differences and similarities in self-enhancement. *Social and Personality Psychology Compass*, *4*(5), 304–317. https://doi.org/10.1111/j.1751-9004.2010.00266.x

Bowlby, J. (1969). *Attachment and loss* (Issue 79). Random House.

Bowlby, J. (1973). Attachment and loss: Volume II: Separation, anxiety and anger. In *Attachment and loss: Volume II: Separation, anxiety and anger* (pp. 1–429). The Hogarth press and the Institute of Psycho-Analysis.

Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development*. Basic Books.

Bradley, M. M., & Lang, P. J. (2010). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings* (Technical Report C-2). University of Florida.

Buhrmester, M. D., Blanton, H., & Swann, W. B., Jr. (2011). Implicit self-esteem: Nature, measurement, and a new way forward. *Journal of Personality and Social Psychology*, *100*(2), 365–385. https://doi.org/10.1037/a0021341

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*(3), 401–423. https://doi.org/10.1037/0033-2909.115.3.401

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*(1), 3–25. https://doi.org/10.1207/s15327957pspr0101_2

Cai, H., Brown, J. D., Deng, C., & Oakes, M. A. (2007). Self-esteem and culture: Differences in cognitive self-evaluations or affective self-regard?*Asian Journal of Social Psychology*, *10*(3), 162–170. https://doi.org/10.1111/j.1467-839X.2007.00222.x

Choi, I., & Choi, Y. (2002). Culture and self-concept flexibility. *Personality and Social Psychology Bulletin*, *28*(11), 1508–1517. https://doi.org/10.1177/014616702237578

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37–57. https://doi.org/10.1037/pspa0000014

Cooley, C. H. (1983). *Human Nature and the Social Order* (1st ed.). Routledge. https://doi.org/10.4323/9780203789513

Dehart, T., Pelham, B., Fiedorowicz, L., Carvallo, M., & Gabriel, S. (2011). Including others in the implicit self: Implicit evaluation of significant others. *Self and Identity*, *10*(1), 127–135. https://doi.org/10.1080/15298861003687880

Dewitte, M., De Houwer, J., Buysse, A., & Koster, E. H. W. (2008). Proximity seeking in adult attachment: Examining the role of automatic approach–avoidance tendencies. *British Journal of Social Psychology*, *47*(4), 557–573. https://doi.org/10.1348/014466607X265148

Falk, C. F., & Heine, S. J. (2015). What is implicit self-esteem, and does it vary across cultures?*Personality and Social Psychology Review*, *19*(2), 177–198. https://doi.org/10.1177/1088868314544693

Falk, C. F., Heine, S. J., Yuki, M., & Takemurá, K. (2009). Why do westerners self-enhance more than East Asians?*European Journal of Personality*, *23*(3), 183–203. https://doi.org/10.1002/per.715

Farnham, S. D., Greenwald, A. G., & Banaji, M. R. (1999). Implicit self-esteem. In D. Abrams & M. A. Hogg (Eds.), *Social identity and social cognition* (pp. 230–248). Blackwell Publishing.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297–327. https://doi.org/10.1146/annurev.psych.54.101601.145225

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238. https://doi.org/10.1037/0022-3514.50.2.229

Gallucci, A., Del Mauro, L., Pisoni, A., Lauro, L. J. R., & Mattavelli, G. (2020). A systematic review of implicit attitudes and their neural correlates in eating behaviour. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsaa156

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gelfand, M. J., Nishii, L. H., & Raver, J. L. (2006). On the nature and importance of cultural tightness-looseness. *Journal of Applied Psychology*, *91*(6), 1225–1244. https://doi.org/10.1037/0021-9010.91.6.1225.

Graham, S. M., & Clark, M. S. (2006). Self-esteem and organization of valenced information about others: The" Jekyll and Hyde"-ing of relationship partners. *Journal of Personality and Social Psychology*, *90*(4), 652–665. https://doi.org/10.1037/0022-3514.90.4.652.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022–1038. https://doi.org/10.1037/0022-3514.79.6.1022

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Guidano, V. F., Liotti, G., & Mahoney, M. J. (1983). *Cognitive processes and emotional disorders: A structural approach to psychotherapy*. The Guilford Press.

Hamamura, T., Heine, S. J., & Paulhus, D. L. (2008). Cultural differences in response styles: The role of dialectical thinking. *Personality and Individual Differences*, *44*(4), 932–942. https://doi.org/10.1016/j.paid.2007.10.034

Hazan, C., & Shaver, P. R. (1994). Attachment as an organizational framework for research on close relationships. *Psychological Inquiry*, *5*(1), 1–22. https://doi.org/10.1207/s15327965pli0501_1

Heine, S. J., & Hamamura, T. (2007). In search of east Asian self-enhancement. *Personality and Social Psychology Review*, *11*(1), 4–27. https://doi.org/10.1177/1088868306294587

Hetts, J. J., Sakuma, M., & Pelham, B. W. (1999). Two roads to positive regard: Implicit and explicit self-evaluation and culture. *Journal of Experimental Social Psychology*, *35*(6), 512–559. https://doi.org/10.1006/jesp.1999.1391

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*(3), 319–340. https://doi.org/10.1037/0033-295X.94.3.319

Ho, S. M., Mak, C. W., Yeung, D., Duan, W., Tang, S., Yeung, J. C., & Ching, R. (2015). Emotional valence, arousal, and threat ratings of 160 Chinese words among adolescents. *PLOS ONE*, *10*(7), e0132294. https://doi.org/10.1371/journal.pone.0132294

Izuma, K., Kennedy, K., Fitzjohn, A., Sedikides, C., & Shibata, K. (2018). Neural activity in the reward-related brain regions predicts implicit self-esteem: A novel validity test of psychological measures using neuroimaging. *Journal of Personality and Social Psychology*, *114*(3), 343–357. https://doi.org/10.1037/pspa0000114

James, W. (1890). *The principles of psychology* (Vol. 1). Henry Holt and Co. https://doi.org/10.1037/10538-000

Jenkins, H., & Deuze, M. (2008). Convergence culture. *Convergence*, *14*(1), 5–12.

Ji, L. J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology*, *87*(1), 57–65. https://doi.org/10.1037/0022-3514.87.1.57

Johnson, M. M. S. (1986). *The initial letter effect: Ego-attachment or mere exposure?* [Unpublished doctoral dissertation], Ohio State University.

Keller, H. (2013). Attachment and culture. *Journal of Cross-Cultural Psychology*, *44*(2), 175–194. https://doi.org/10.1177/0022022112472253

Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin*, *23*(7), 736–742. https://doi.org/10.1177/0146167297237006

Kitayama, S., & Uchida, Y. (2003). Explicit self-criticism and implicit self-regard: Evaluating self and friend in two cultures. *Journal of Experimental Social Psychology*, *39*(5), 476–482. https://doi.org/10.1016/S0022-1031(03)00026-X

Klein, M. (1955). The psychoanalytic play technique. *American Journal of Orthopsychiatry*, *25*(2), 223–237. https://doi.org/10.1111/j.1939-0025.1955.tb00131.x

Kobayashi, C., & Greenwald, A. G. (2003). Implicit-explicit differences in self-enhancement for Americans and Japanese. *Journal of Cross-Cultural Psychology*, *34*(5), 522–541. https://doi.org/10.1177/0022022103257855

Kohut, H. (1966). Forms and transformations of narcissism. *Journal of the American Psychoanalytic Association*, *14*(2), 243–272. https://doi.org/10.1177/000306516601400201

Koole, S. L., Dijksterhuis, A., & Van Knippenberg, A. D. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*(4), 669–685. https://doi.org/10.1037//0022-3514.80.4.669

Krause, S., Back, M. D., Egloff, B., & Schmukle, S. C. (2016). Predicting self–confident behaviour with implicit and explicit self–esteem measures. *European Journal of Personality*, *30*(6), 648–662. https://doi.org/10.1002/per.2076

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition*, *38*(Suppl.), s42–s67. https://doi.org/10.1521/soco.2020.38.supp.s42

Lakens, D. (2016, March 19). *One-sided tests: Efficient and underused*. The 20% Statistician. http://daniellakens.blogspot.com/2016/03/one-sided-tests-efficient-and-underused.html

Leary, M. R. (2005). Sociometer theory and the pursuit of relational value: Getting to the root of self-esteem. *European Review of Social Psychology*, *16*(1), 75–111. https://doi.org/10.1080/10463280540000007

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. *Advances in Experimental Social Psychology 32*, 1–62. https://doi.org/10.1016/S0065-2601(00)80003-9

Lee, R. T., Ni, M., Fang, W. M., Ravreby, I., Shoda, Y., & Zayas, V. (2024). An integrative framework for capturing emotion and emotion regulation in daily life. *Affective Science*. Advance online publication. https://doi.org/10.1007/s42761-024-00262-0

Lee, S. W. S., Oyserman, D., & Bond, M. H. (2010). Am I doing better than you? That depends on whether you ask me in English or Chinese: Self-enhancement effects of language as a cultural mindset prime. *Journal of Experimental Social Psychology*, *46*(5), 785–791. https://doi.org/10.1016/j.jesp.2010.04.005

Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. Guilford Press.

Lipkus, I. M., Green, J. D., Feaganes, J. R., & Sedikides, C. (2001). The relationship between attitudinal ambivalence and desire to quit smoking among college smokers 1. *Journal of Applied Social Psychology*, *31*(1), 113–133. https://doi.org/10.1111/j.1559-1816.2001.tb02485.x

Maio, G. R., Fincham, F. D., & Lycett, E. J. (2000). Attitudinal ambivalence toward parents and attachment style. *Personality and Social Psychology Bulletin*, *26*(12), 1451–1464. https://doi.org/10.1177/01461672002612001

Markus, H. R., & Kitayama, S. (1991). Cultural variation in the self-concept. In *The self: Interdisciplinary approaches* (pp. 18–48). Springer. https://doi.org/10.1007/978-1-4684-8264-5_2

McNulty, J. K., Baker, L. R., & Olson, M. A. (2014). Implicit self-evaluations predict changes in implicit partner evaluations. *Psychological Science*, *25*(8), 1649–1657. https://doi.org/10.1177/0956797614537833

Mead, G. H. (1934). *Mind, self, and society: From the standpoint of a social behaviorist* (C. W. Morris, Ed.). University of Chicago Press.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, *82*(2), 213–225. https://doi.org/10.1037/h0076486

Miyake, K., Chen, S-j., & Campos, J. J. (1985). Infant temperament, mother's mode of interaction, and attachment in Japan: An interim report. *Monographs of the Society for Research in Child Development*, *50*(1–2), 276–297. https://doi.org/10.2307/3333838

Murray, S. L., Griffin, D. W., Rose, P., & Bellavia, G. (2006). For better or worse? Self-esteem and the contingencies of acceptance in marriage. *Personality and Social Psychology Bulletin*, *32*(7), 866–880. https://doi.org/10.1177/0146167206286756

Murray, S. L., & Holmes, J. G. (2009). The architecture of interdependent minds: A Motivation-management theory of mutual responsiveness. *Psychological Review*, *116*(4), 908–928. https://doi.org/10.1037/a0017015

Murray, S. L., Holmes, J. G., & Collins, N. L. (2006). Optimizing assurance: The risk regulation system in relationships. *Psychological Bulletin*, *132*(5), 641–666. https://doi.org/10.1037/0033-2909.132.5.641

Murray, S. L., Holmes, J. G., Griffin, D. W., & Derrick, J. L. (2015). The equilibrium model of relationship maintenance. *Journal of Personality and Social Psychology*, *108*(1), 93–113. https://doi.org/10.1037/pspi0000004

Newby-Clark, I. R., McGregor, I., & Zanna, M. P. (2002). Thinking and caring about cognitive inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable?*Journal of Personality and Social Psychology*, *82*(2), 157–166. https://doi.org/10.1037/0022-3514.82.2.157

Nuttin, J. M., Jr. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology*, *15*(3), 353–361. https://doi.org/10.1002/ejsp.2420150309

Nuttin, J. M., Jr. (1987). Affective consequences of mere ownership: The name letter effect in twelve European languages. *European Journal of Social Psychology*, *17*(4), 381–402. https://doi.org/10.1002/ejsp.2420170402

Olson, J. M., Goffin, R. D., & Haynes, G. A. (2007). Relative versus absolute measures of explicit attitudes: Implications for predicting diverse attitude-relevant criteria. *Journal of Personality and Social Psychology*, *93*(6), 907–926. https://doi.org/10.1037/0022-3514.93.6.907

O'Shea, B. A., & Wiers, R. W. (2020). Moving beyond the relative assessment of implicit biases: Navigating the complexities of absolute measurement. *Social Cognition*, *38*(Suppl.), s187–s207. https://doi.org/10.1521/soco.2020.38.supp.s187

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, *54*(9), 741–754. https://doi.org/10.1037/0003-066X.54.9.741

Pietromonaco, P. R., & Barrett, L. F. (2000). The internal working models concept: What do we really know about the self in relation to others?*Review of General Psychology*, *4*(2), 155–175. https://doi.org/10.1037/1089-2680.4.2.155

Polek, E. (2008). *Attachment in cultural context: Differences in attachment between Eastern and Western Europeans* [Thesis fully internal (DIV)]. University of Groningen, Faculty of Behavioural and Social Sciences.

Riketta, M., & Ziegler, R. (2006). Self-ambivalence and self-esteem. *Current Psychology*, *25*, 192–211. https://doi.org/10.1007/s12144-006-1003-7

Sakman, E., & Zayas, V. (2024). Negative speaks louder than positive: Negative implicit partner evaluations forecast destructive daily interactions and relationship decline. *Journal of Social and Personal Relationships*, *41*(6), 1415–1438. https://doi.org/10.1177/02654075231224803

Sedikides, C., Gaertner, L., & Cai, H. (2015). On the panculturality of self-enhancement and self-protection motivation: The case for the universality of self-esteem. *Advances in Motivation Science*, *2*, 185–241. https://doi.org/10.1016/bs.adms.2015.04.002

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Spencer-Rodgers, J., Peng, K., Wang, L., & Hou, Y. (2004). Dialectical self-esteem and East-West differences in psychological well-being. *Personality and Social Psychology Bulletin*, *30*(11), 1416–1432. https://doi.org/10.1177/0146167204264243

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. Routledge. https://doi.org/10.4324/9781315014029

Szeto, A. C. H., Sorrentino, R. M., Yasunaga, S., Otsubo, Y., Kouhara, S., & Sasayama, I. (2009). Using the implicit association test across cultures: A case of implicit self-esteem in Japan and Canada. *Asian Journal of Social Psychology*, *12*(3), 211–220. https://doi.org/10.1111/j.1467-839X.2009.01286.x

Tafarodi, R. W., Shaughnessy, S. C., Yamaguchi, S., & Murakoshi, A. (2011). The reporting of self-esteem in Japan and Canada. *Journal of Cross-Cultural Psychology*, *42*(1), 155–164. https://doi.org/10.1177/0022022110386373

Triandis, H. C. (1995). *Individualism and collectivism* (1st ed.). Routledge. https://doi.org/10.4324/9780429499845

Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of Personality*, *69*(6), 907–924. https://doi.org/10.1111/1467-6494.696169

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131–160). Addison-wesley.

Uskul, A. K., & Over, H. (2017). Culture, social interdependence, and ostracism. *Current Directions in Psychological Science*, *26*(4), 371–376. https://doi.org/10.1177/0963721417699300

Van IJzendoorn, M. H., & Kroonenberg, P. M. (1988). Cross-cultural patterns of attachment: A meta-analysis of the strange situation. *Child Development*, *59*(1), 147–156. https://doi.org/10.2307/1130396

Van Ijzendoorn, M. H., & Sagi-Schwartz, A. (2008). Cross-cultural patterns of attachment: Universal and contextual dimensions. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 880–905). The Guilford Press.

Wang, Q., Shao, Y., & Li, Y. J. (2010). "My way or mom's way?" The bilingual and bicultural self in Hong Kong Chinese children and adolescents. *Child Development*, *81*(2), 555–567. https://doi.org/10.1111/j.1467-8624.2009.01415.x

Wegener, I., Geiser, F., Alfter, S., Mierke, J., Imbierowicz, K., Kleiman, A., Koch, A. S., & Conrad, R. (2015). Changes of explicitly and implicitly measured self-esteem in the treatment of major depression: Evidence for implicit self-esteem compensation. *Comprehensive Psychiatry*, *58*, 57–67. https://doi.org/10.1016/j.comppsych.2014.12.001

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). The Guilford Press.

Williams, K. D. (2009). Ostracism: A temporal need-threat model. *Advances in Experimental Social Psychology*, *41*, 275–314. https://doi.org/10.1016/S0065-2601(08)00406-1

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101.

Winter, D., Steeb, L., Herbert, C., Sedikides, C., Schmahl, C., Bohus, M., & Lis, S. (2018). Lower self-positivity and its association with self-esteem in women with borderline personality disorder. *Behaviour Research and Therapy*, *109*, 84–93. https://doi.org/10.1016/j.brat.2018.07.008

Yamaguchi, S., Greenwald, A. G., Banaji, M. R., Murakami, F., Chen, D., Shiomura, K., . . .Krendl, A. (2007). Apparent universality of positive implicit self-esteem. *Psychological Science*, *18*(6), 498–500. https://doi.org/10.1111/j.1467-9280.2007.01928.x

Zayas, V., & Shoda, Y. (2005). Do automatic reactions elicited by thoughts of romantic partner, mother, and self relate to adult romantic attachment?*Personality and Social Psychology Bulletin*, *31*(8), 1011–1025. https://doi.org/10.1177/0146167204274100

Zayas, V., & Shoda, Y. (2015). Love you? Hate you? Maybe it's both: Evidence that significant others trigger bivalent-priming. *Social Psychological and Personality Science*, *6*(1), 56–64. https://doi.org/10.1177/1948550614541297

Zayas, V., Surenkok, G., & Pandey, G. (2017). Implicit ambivalence of significant others: Significant others trigger positive and negative evaluations. *Social and Personality Psychology Compass*, *11*(11), e12360. https://doi.org/10.1111/spc3.12360

Zayas, V., Wang, A. M., & McCalla, J. D. (2022). Me as good and me as bad: Priming the self triggers positive and negative implicit evaluations. *Journal of Personality and Social Psychology*, *122*(1), 106–134. https://doi.org/10.1037/pspp0000332