Simultaneous ingroup and outgroup favoritism in implicit social cognition

Jordan R. Axt[a,*], Tal Moran[b], Yoav Bar-Anan[b]

[a] *University of Virginia, United States of America*
[b] *Ben-Gurion University of the Negev, Israel*

ARTICLE INFO

ABSTRACT

People like their own groups, producing ingroup favoritism, a hallmark finding of social identity theory. However, as predicted by system justification or cultural learning perspectives, outgroup favoritism among non-dominant groups is occasionally observed, particularly implicitly. The present research found that non-dominant group members displayed simultaneous ingroup *and* dominant group implicit favoritism. On indirect measures focusing on positive valence, members of non-dominant racial (Studies 1 and 4), religious (Study 2), and sexual (Study 3) groups showed ingroup favoritism. On indirect measures focusing on negative valence, members of non-dominant groups showed diminished ingroup favoritism, and sometimes favoritism towards the culturally dominant group. These results may indicate that positive self-regard forms associations between the ingroup and *positive*, whereas cultural learning and system justification form associations between non-dominant groups and *negative*. A cross-cultural design (Study 5) also found results compatible with these assumptions.

1. Introduction

Across social dimensions, people tend to have more positive attitudes towards members of their own groups (Tajfel & Turner, 1979), and towards anything associated with the self (Greenwald, 1980). Ingroup favoritism is frequently displayed by both dominant and non-dominant group members. For instance, both Latinos in America and Arabs in Israel reported greater levels of ingroup identification and equal levels of ingroup positivity compared to White and Jewish counterparts (Levin & Sidanius, 1999). Likewise, racial and religious minorities report explicit preferences for their own group at levels greater than or equal to those of Whites and Christians (Axt, Ebersole, & Nosek, 2014).

Although ingroup favoritism is pervasive, weaker ingroup favoritism or even outgroup favoritism is sometimes observed among members of socially stigmatized or non-dominant groups, particularly when using indirect measures of implicit evaluations (Jost, Banaji, & Nosek, 2004). For example, Asian-American participants exhibited weaker preferences for Asians over Whites on indirect versus direct measures of explicit evaluations (Rudman, Feinberg, & Fairchild, 2002). Similarly, a sample of Hispanic-Americans exhibited no implicit preference for Hispanics versus Whites (Uhlmann, Dasgupta, Elgueta, Greenwald, & Swanson, 2002). In other studies, African Americans showed strong ingroup preferences with direct measures, but no ingroup preference or slight outgroup preference on a Black-White

Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; Cohen's $d = 0.04$ in Jost et al., 2004; $d = 0.05$ in Nosek et al., 2007; $d = -0.16$; Nosek, Banaji, & Greenwald, 2002).

Diminished implicit ingroup favoritism is not limited to racial minorities. Gay participants revealed much weaker ingroup preference on a Gay-Straight IAT ($d = 0.11$) compared to straight participants ($d = 1.10$; Jost et al., 2004). Jewish participants also exhibited a weaker preference for Jews relative to Christians implicitly than explicitly, while overweight and low-income participants held no explicit preference for their own group but strong implicit preferences for thin and rich people, respectively (Rudman et al., 2002). Older adults showed no explicit ingroup preference between young and old people, and strongly preferred younger to older people implicitly (Gonsalkorale, Sherman, & Klauer, 2014; Nosek et al., 2002; Nosek et al., 2007). Finally, overweight and obese participants held preferences for thin over fat people explicitly ($d = 0.54$) and especially implicitly ($d = 0.91$; Schwartz, Vartanian, Nosek, & Brownell, 2006).

Such results have often been interpreted as evidence that implicit attitudes are partly shaped by culturally-based information that participants may be unaware of or explicitly disavow (e.g., Jost, Pelham, & Carvallo, 2002). From a cultural learning approach, individuals' implicit attitudes are sensitive to the stereotypes and values provided by their social context (Dasgupta, 2013). For example, women's implicit stereotypes associating men with leadership increased during time spent at a co-ed college but decreased during time spent at a women's

Fig. 1. Sample trials for an IAT (left) and a good-focal MC-IAT or BIAT (right).

college (Dasgupta & Asgari, 2004). According to system justification theory (Jost & Banaji, 1994), individuals have a need to view the world as fair and preserve the status quo, even when one's group possesses lower standing than other groups. However, due to social pressure and personal interest to express ingroup favoritism, the influence of system justification is believed to be most evident on indirect measures of implicit attitudes (Jost et al., 2004). As a result, non-dominant group members may consciously reject certain cultural values and stereotypes explicitly but retain them implicitly. In both cases, cultural learning and system justification perspectives argue that individuals' implicit attitudes should be partly shaped by the cultural values and stereotypes provided by the social context.

These two forces – group identity and cultural learning – lead to differing influences for dominant versus non-dominant group members (e.g., Blodorn, O'Brien, Cheryan, & Vick, 2016). For members of dominant groups, group identity concerns align with system justification and cultural learning influences, as both support one's ingroup. However, for non-dominant group members, group identity and system justification concerns are in opposition, with the former supporting one's own group and the latter supporting culturally dominant groups. In the present research, we report findings that non-dominant group members simultaneously show cognitions compatible with both forces in two different aspects of implicit attitudes: their implicit attitudes were compatible with the influence of group identity more than with cultural norms when measured with a focus on positive valence, but were often more compatible with cultural norms than with group identity when measured with a focus on negative valence.

1.1. The present work

The present research began with an unexpected discovery of ingroup favoritism among non-dominant group members, countering prior evidence. In Axt et al. (2014), racial (Asian, Black, Hispanic; $N = 9668$) and religious (Jewish, Buddhist, Hindu, Muslim; $N = 11,994$) non-dominant participants showed robust implicit ingroup favoritism towards their own group relative to the dominant group (Whites and Christians). The magnitude of the implicit ingroup favoritism was slightly weaker than the dominant group's for race ($d = 0.25$ for non-dominant group members; $d = 0.30$ for Whites) and moderately weaker for religion ($d = 0.62$ for non-dominant group members; $d = 0.95$ for Christians).

The difference between these results and prior research (Jost et al., 2004; Nosek et al., 2002, 2007) is surprising because similar populations were examined. The main difference, however, was the indirect measure – the IAT in previous research and the Multi-Category Implicit Association Test (MC-IAT, Sriram & Greenwald, 2009) in Axt et al.

(2014). Both measures compare two conditions in which participants categorize stimuli into groups as fast as possible with two keys. In one condition, participants categorize stimuli representing two categories (e.g., Black faces and Bad words) with one key and stimuli representing complementary categories (i.e., White faces and Good words) with the other key. In the other condition, the key assignments change for two categories such that Black faces and Good words are categorized with one key and White faces and Bad words with the other key. For both measures, the effect is assessed as the relative difficulty in categorizing items in one condition compared to the other.

The IAT versus the MC-IAT differ in one key respect. In the IAT, all four categories (e.g., Black, White, Good, Bad) are identified explicitly with category labels during all blocks to facilitate categorization. In contrast, the MC-IAT uses an innovation first presented in the Brief Implicit Association (BIAT; Sriram & Greenwald, 2009): only two categories are named explicitly (e.g., Black and Good) for categorizing with one key, and the other key is labeled "all else". In each block, the MC-IAT/BIAT encourages participants to focus on two "focal" categories rather than all categories simultaneously. This structural change affects participants' focus; Sriram and Greenwald (2009) observed faster responses to stimuli assigned to the focal category, indicating that participants selectively attended to the focal categories in each block. See Fig. 1 for sample trials of an IAT and MC-IAT.

Axt et al. (2014) used a "Good-focal" MC-IAT: for instance, the focal category labels were "Black and Good" in one block and "White and Good" in another block. The "Bad" stimuli were always in the "Everything else" category. We surmise that this procedural difference may have activated associations showing ingroup favoritism, unlike the typical IAT results among minorities. Perhaps a focus on good valence elicits stronger ingroup preferences and a focus on bad valence elicits stronger dominant group preferences. These distinct effects may be masked on the IAT, which explicitly refers to the concepts *good* and *bad* simultaneously. Indeed, an initial study reported fully in the online supplement (Study S1; $N = 200$) found that an IAT measuring implicit evaluations of White versus Black people was related with parallel good-focal and bad-focal BIATs, sharing unique variance with each. These results are compatible with the hypothesis that the IAT reflects a combination of the associations measured with the bad-focal and good-focal BIATs.

If the good-focal and bad-focal BIATs measure different constructs, confounded in the IAT, what are these constructs? In a second supplemental study (Study S2; see online supplement), we tested one hypothesis – that positive information is more impactful on good-focal BIATs and negative information more impactful on bad-focal BIATs. Participants ($N = 195$) read either 10 positive or 10 negative pieces of information about a target person, then completed good-focal and bad-

focal BIATs measuring implicit preferences between the target and a novel person. When participants read negative information about the target, preference for the novel person was stronger on the bad-focal BIAT than on the good-focal BIAT ($p = .039$, $d = 0.21$). When participants read positive information about the target, the preference for the target was stronger on the good-focal BIAT than the bad-focal BIAT ($p = .002$, $d = 0.32$). In other words, there was an interaction ($p < .001$, $\eta_p^2 = 0.07$) between type of information presented (positive versus negative) and the type of indirect measure used (good-focal versus bad-focal), reflecting more sensitivity of good-focal BIATs to positive than negative information, and more sensitivity of bad-focal BIATs to negative than positive information.

Perhaps Axt et al.'s (2014) finding that good-focal BIATs show ingroup favoritism among members of non-dominant groups uncovers an implicit cognition that existing measures—without the ability to assess good and bad associations separately—missed. A focus on good valence may produce more ingroup favoritism because people associate positive attributes with their ingroup more than with outgroups. Individuals maintain a strong need for positive self-regard (Sedikides, Gaertner, & Toguchi, 2003), and use their group membership as one source of this positivity (e.g., Brewer, 1999; Greenwald & Pettigrew, 2014). Members of both dominant and non-dominant groups may then find their group membership as primarily a source for positive information, and this information is then more influential in shaping performance on indirect measures focused on positive valence.

Conversely, a focus on bad valence may produce more dominant group favoritism because cultural values accentuate negative information about non-dominant groups. For instance, non-dominant groups are overrepresented as threats in the media (e.g., Tukachinsky, Mastro, & Yarchi, 2015) and non-dominant group members frequently receive negative labels from dominant group members (e.g., Verkuyten & Thijs, 2002). At the same time, dominant group identities are often "unmarked categories" that serve as the silent default, and are rarely identified explicitly (Hartmann, Gerteis, & Croll, 2009; Pratto & Stewart, 2012). That is, cultural norms do not explicitly present the members of the dominant groups positively so much as they present members of non-dominant groups negatively, and these cultural norms may then be more influential in shaping performance on indirect measures focused on negative valence.

As a result, dominant and non-dominant group members may derive mostly positive associations with their ingroup from their group identity, and derive mostly negative associations with non-dominant groups from cultural messages, even if they consciously disavow such messages. Moreover, this framework anticipates that positive associations will be more related to explicit, conscious group evaluations than negative associations because both share the same source – positive self-regard. Because members of both non-dominant and dominant groups may not endorse the negative depiction of the non-dominant groups presented in one's culture (e.g., through media depictions), self-reported ingroup evaluation would reflect people's positive self-identity rather than the negative information the culture provides about non-dominant groups (Brewer & Gardner, 1996). The associations formed by negative information would then be detected more by measures focused on negative valence and would have less influence on individuals' explicit ingroup and outgroup evaluations.

If positive information about the ingroup induces ingroup favoritism on measures of positive valence, both members of dominant and non-dominant social groups should show ingroup favoritism on measures that are more sensitive to positive information. If negative information about non-dominant group affects performance mostly on measures of negative valence, then members of non-dominant groups should show weaker ingroup favoritism on measures that are more sensitive to negative information than on measures more sensitive to positive information. For dominant group members, both measures should reveal ingroup favoritism, though one measure may do so more strongly than the other.

The present research investigated whether good-focal and bad-focal association measures reflect these two distinct associations. Dominant[1] and non-dominant group members completed indirect measures of evaluations concerning race (Studies 1 and 4), religion (Studies 2 and 5) and sexual orientation (Study 3). In each study, non-dominant participants displayed ingroup favoritism on measures focusing on positive valence and diminished ingroup favoritism or even outgroup favoritism towards the culturally dominant group on measures focusing on negative valence, whereas dominant group members displayed ingroup favoritism on both forms of measurement. These results further highlight how positive and negative associations are distinct components of implicit social cognition, and clarify how implicit attitudes may be dually shaped by social identity and cultural learning perspectives.

2. Study 1

2.1. Method

2.1.1. Participants

In Studies 1–4, due to the possible cultural specificity of effects, we included only American citizens or residents in analysis. See https://osf.io/u7tkp for all data, materials and the online supplement. In all studies, we report all measures, manipulations and exclusions. In Study 1, a specific sample size was not determined beforehand, and data collection ended once we believed the sample size was sufficiently large. For all other studies, sample size was determined before data analysis.

Of the 38,316 participants who volunteered, consented, and provided data as part of a featured task at Project Implicit (implicit.harvard.edu; Nosek, 2005), we included 23,764 Americans ($M_{age} = 31.9$, $SD_{age} = 14.50$; 73.4% female).[2] We classified participants as White ($n = 15,673$), Black ($n = 2064$), or East Asian ($n = 890$) if they reported that race, and reported their ethnicity as not Hispanic, and classified as Hispanic ($n = 2320$) if they reported their ethnicity as Hispanic. This sample size provided at least 80% power at detecting a small between-subjects effect size of Cohen's $d = 0.20$ for all comparisons.

2.1.2. Procedure

The study consisted of four components completed in randomized order: two surveys assessing racial and political attitudes not analyzed for this manuscript, a demographics questionnaire with explicit racial preference measures, and a race MC-IAT, which was either good-focal or bad-focal.

2.1.2.1. Demographics and survey items.
Participants completed a 15-item demographics questionnaire. We only analyzed the items relating to race, ethnicity, gender, and age. After demographics, participants completed six items concerning preferences for Black, White, Asian, and Hispanic people responding on a 7-point scale ranging from "I strongly prefer X people to Y people" (-3) to "I strongly prefer Y people to X people" ($+3$) for all six possible pairings.

2.1.2.2. Indirect measure.
The MC-IAT measured evaluation strengths between four racial groups: Black, White, Asian, and Hispanic. Participants were randomly assigned to complete a good-focal or bad-focal MC-IAT.

---

[1] In the present work, dominant groups (i.e., those having the most power or status) were identical to groups holding a numerical majority. As a result, we cannot disentangle whether observed effects are due to a group's status or size. Since most theories in social psychology differentiate between the social groups used here by referring to their social status, we refer to such groups as either "dominant" or "non-dominant".

[2] The study had 44,956 started sessions, with 38,316 providing data, and 24,400 completing the study (54.3% completion rate).

In the first block of a good-focal MC-IAT (16 trials), participants pressed the "I" key for all Good words (Love, Pleasant, Great, Wonderful) and the "E" key for other words (Hate, Unpleasant, Awful, Terrible). In the second block (16 trials), participants pressed the "I" key for all Good words and for faces (two male, two female with pro-totypical surnames, e.g., "N. Chang" below the face) belonging to the focal category (Asian, Black, Hispanic or White), and the "E" key for "any other images and words." The other items were the same negative words and faces from one of the other three racial groups. These first two blocks were practice and not analyzed. For the remaining 12 blocks (16 trials each), the structure was the same as the second block with the target and other racial group rotating between all 12 possible combinations. For example, there were three blocks for which participants hit the "I" key for Asian faces, and the other faces were Black, Hispanic, or White faces each for one of those blocks. Randomization was constrained so that each racial group appeared as a target once every four blocks. Participants were randomly assigned to one out of 24 possible block orders. Bad-focal MC-IATs had the same design, with the only change being that the "I" key was used to classify all Bad words and the "E" key was used for other words (the good words).

In all studies, good-focal and bad-focal MC-IATs were scored using the *D* algorithm, recommended by Nosek, Bar-Anan, Sriram, Axt, and Greenwald (2014).[3] The MC-IAT allowed computing six *D* scores representing each paired comparison of racial groups (White vs. Black, White vs. Asian, White vs. Hispanic, Asian vs. Black, Asian vs. Hispanic, Black vs. Hispanic). For good-focal BIATs, higher values denoted stronger associations of positive valence with the dominant group (Whites in Studies 1 and 4, Christianity in Study 2, straight people in Study 3, Judaism in Study 5) than with the non-dominant group. For example, higher values for a White versus Black *D* score in a good-focal BIAT indicates participants were faster when the focal categories were White people and Good words than when the focal categories were Black people and Good words. For bad-focal BIATs, higher values denoted stronger associations of negative valence with the non-dominant than the dominant group. For example, higher values for a White versus Black *D* score in a bad-focal BIAT indicated participants were faster when the two focal categories were Black people and Bad words than White people and Bad words.

In Study 1, we excluded the MC-IAT data of participants (3.2%) who had > 10% of their responses faster than 300 ms (Nosek et al., 2014).

2.2. Results

Explicit attitudes were scored such that more positive scores indicate greater preferences for the dominant group. We focused on the three scores within the MC-IAT measuring associations towards White people (White vs. Asian, White vs. Black, White vs. Hispanic) but all scores are available in the online dataset.

2.2.1. Reliabilities

We computed Cronbach's alpha (Cronbach, 1951) from three data parcels for each measure. The first parcel included the first trial of each triplet of consecutive trials, the second parcel included the second trial, and the third parcel included the third trial. Good-focal and Bad-focal MC-IATs were comparably reliable for White-Asian (Good: α = 0.65, Bad: α = 0.65), White-Black (Good: α = 0.70, Bad: α = 0.69), and White-Hispanic (Good: α = 0.65, Bad: α = 0.66) associations.

2.2.2. Implicit evaluations

For Studies 1–2, we tested for differences between good-focal and bad-focal MC-IATs using a series of independent samples *t*-tests. Among participants from non-dominant racial groups, good-focal and bad-focal

MC-IATs reliably differed from each other (see Table 1 for descriptive and test statistics).

Across studies, we also tested for the presence of ingroup or outgroup favoritism in implicit evaluations through a series of one-sample *t*-tests against a neutral value of zero. In Study 1, participants from non-dominant racial groups showed reliable implicit ingroup favoritism on good-focal MC-IATs, and reliable implicit outgroup favoritism on bad-focal MC-IATs (see Table 2 for test statistics), with the one exception being Hispanic participants showing only moderate evidence of ingroup favoritism on good-focal MC-IATs.

White participants showed more implicit ingroup favoritism on bad-focal than good-focal MC-IATs (see Table 1), but showed reliable implicit ingroup favoritism with both good-focal and bad-focal measures across all comparison races (see Table 2).

2.2.3. Explicit evaluations

See Table 3 for descriptive statistics of explicit evaluations. Across studies, we tested for the presence of explicit ingroup favoritism through a series of one-sample *t*-tests against a neutral value of zero. Participants from non-dominant racial groups preferred their own group relative to White people (all *t*'s > 12.92, all *p*'s < .001, all *d*'s > 0.32; see online supplement for individual tests), and White participants preferred White people to each other racial group (all *t*'s > 59.55, all *p*'s < .001, all *d*'s > 0.48; see online supplement).

2.2.4. Relationship between explicit and implicit evaluations

For Studies 1–2, we used Fisher's *r* to *Z* transformation (Fisher, 1921) to compare independent correlations. Using the entire sample, for White vs. Black associations, good-focal MC-IATs ($r(9,730) = 0.20$, $p < .001$) were more related to explicit preferences than bad-focal MC-IATs ($r(8,740) = 0.14$, $p < .001$), Fisher's $z = 4.05$, $p < .001$. For White vs. Hispanic associations, good-focal MC-IATs ($r(9739) = 0.15$, $p < .001$) were more related to explicit preferences than bad-focal MC-IATs ($r(8736) = 0.10$, $p < .001$), Fisher's $z = 3.17$, $p = .002$. For White vs. Asian associations, there were no reliable differences for good-focal MC-IATs ($r(9735) = 0.12$, $p < .001$) vs. bad-focal MC-IATs ($r(8750) = 0.13$, $p < .001$) in correlations with explicit preferences, Fisher's $z = -0.48$, $p = .629$.

2.3. Discussion

Participants from non-dominant racial groups showed implicit ingroup favoritism with measures focusing on positive valence (though very weakly for Hispanic participants) and implicit outgroup favoritism with measures focusing on negative valence. White participants showed implicit ingroup favoritism in both types of measures, though the effect was stronger with measures focusing on negative valence than with measures focusing on positive valence.

Though participants categorized the same stimuli in both tasks, focusing on positive valence revealed ingroup favoritism, whereas focusing on negative valence revealed dominant group favoritism among participants from non-dominant racial groups. Further, good-focal measures were generally more strongly correlated with explicit preferences than were bad-focal measures, despite similar internal consistencies. These results are consistent with an interpretation that bad-focal measures are more sensitive to cultural evaluations favoring the dominant group (Whites), and good-focal measures are more sensitive to evaluations experienced as one's own. Study 2 repeated the same test in another domain, religion.

3. Study 2

3.1. Method

3.1.1. Participants

We collected data until there were at least 40 participants per

---

[3] We use *D* to refer to outcome scores for IATs, MC-IATs or BIATs, and *d* to refer to the Cohen's *d* effect size (Cohen, 1988).

Table 1
Means, standard deviations, and test statistics for comparing good-focal and bad-focal measures in Studies 1, 2 and 4.

Study 1

	Good N	Bad N	Good-focal	Bad-focal	t	p	d [95% CI]
White participants							
White-Asian associations	6573	5948	0.11 (0.59)	0.33 (0.58)	21.10	< .001	0.38 [0.34, 0.41]
White-Black associations	6567	5943	0.16 (0.60)	0.39 (0.60)	21.85	< .001	0.39 [0.36, 0.43]
White-Hispanic associations	6562	5940	0.17 (0.58)	0.43 (0.57)	25.19	< .001	0.45 [0.42, 0.93]
Asian participants							
White-Asian associations	359	361	−0.08 (0.60)	0.15 (0.60)	5.26	< .001	0.39 [0.24, 0.54]
Black participants							
White-Black associations	856	751	−0.12 (0.60)	0.16 (0.64)	9.01	< .001	0.45 [0.35, 0.55]
Hispanic participants							
White-Hispanic associations	1042	881	−0.03 (0.60)	0.26 (0.59)	10.62	< .001	0.49 [0.40, 0.58]

Study 2

	Good N	Bad N	Good-focal	Bad-focal	t	p	d [95% CI]
Christian participants							
Christianity-Judaism associations	493	450	0.45 (0.43)	0.28 (0.48)	5.71	< .001	0.37 [0.24, 0.50]
Christianity-Hinduism associations	498	455	0.56 (0.47)	0.32 (0.53)	7.65	< .001	0.50 [0.37, 0.63]
Christianity-Islam associations	489	426	0.56 (0.46)	0.37 (0.50)	5.83	< .001	0.39 [0.26, 0.52]
Jewish participants							
Christianity-Judaism associations	110	91	−0.28 (0.50)	0.05 (0.52)	3.18	.002	0.45 [0.17, 0.73]
Hindu participants							
Christianity-Hinduism associations	50	41	−0.27 (0.44)	0.02 (0.61)	2.27	.025	0.47 [0.06, 0.90]
Muslim participants							
Christianity-Islam associations	77	48	−0.30 (0.48)	0.07 (0.53)	3.97	< .001	0.72 [0.36, 1.10]

Study 4

	N	Good-focal	Bad-focal	t	p	d [95% CI]
White participants						
White-Asian associations	303	0.17 (0.48)	0.37 (0.41)	5.78	< .001	0.33 [0.22, 0.45]
White-Black associations	278	0.25 (0.45)	0.36 (0.44)	3.72	< .001	0.22 [0.10, 0.34]
White-Hispanic associations	271	0.17 (0.50)	0.41 (0.45)	7.18	< .001	0.44 [0.31, 0.56]
Asian participants						
White-Asian associations	113	−0.13 (0.41)	0.20 (0.44)	6.45	< .001	0.63 [0.42, 0.83]
Black participants						
White-Black associations	202	−0.11 (0.46)	0.004 (0.48)	2.62	.010	0.18 [0.04, 0.32]
Hispanic participants						
White-Hispanic associations	180	−0.07 (0.49)	0.15 (0.48)	4.90	< .001	0.36 [0.21, 0.51]

Note. Means and standard deviations (in parentheses). All BIAT's scored such that higher values mean more positive associations with group listed first. *Good N* = sample size in good-focal condition. *Bad N* = sample size in bad-focal condition. *N* = sample size for study using within-subjects design. *Good-Focal* = good-focal BIAT *D* score. *Bad-Focal* = bad-focal BIAT *D* score.

condition from each religion. Of the 2740 participants who volunteered, consented and provided data for the study at the Project Implicit research pool, we included 1855 Americans (M_{age} = 31.10, SD_{age} = 14.1, 56.8% female).[4]

Participants were only assigned the study if they reported being Christian (n = 1285), Jewish (n = 272), Hindu (n = 120), or Muslim (n = 165) when registering. Given the difficulty in recruiting minority populations, power was only moderate for comparisons among participants from non-dominant groups. The final sample size provided 80% power at detecting a between-subjects effect size of d = 0.39 for Jewish participants, d = 0.52 for Muslim participants, d = 0.60 for Hindu participants. The sample provided 80% power at detecting an effect of d = 0.20 for Christian participants.

3.1.2. Procedure

The study consisted of two components completed in a randomized order: an attitudes survey and either a good-focal or bad-focal religions MC-IAT.

3.1.2.1. Attitudes survey. Participants completed the same explicit

---

[4] The study had 3443 started sessions, with 2740 providing data, and 2254 completing the study (64.4% completion rate).

preference items used in Study 1, now concerning Christian, Jewish, Islamic, and Hindu people.

3.1.2.2. Indirect measure. The MC-IAT was the same as in Study 1, but with different categories: Christianity (Jesus, Christian, Gospel, Church), Hinduism (Krishna, Hindu, Dharma, Mantra), Islam (Muhammad, Muslim, Koran, Allah), and Judaism (Abraham, Jew, Torah, Yahweh). We excluded 8.4% of the participants for the same exclusion rule as in Study 1.

3.2. Results

3.2.1. Reliabilities

Good-focal and bad-focal MC-IATs were comparably reliable, for Christianity-Judaism (Good: α = 0.67, Bad: α = 0.63), Christian-Muslim (Good: α = 0.73, Bad: α = 0.64), and Christianity-Hinduism (Good: α = 0.71, Bad: α = 0.68) associations.

3.2.2. Implicit evaluations

Using the same analysis strategy as Study 1, participants from non-dominant religious groups showed reliable differences on good-focal versus bad-focal MC-IATs (see Table 1 for descriptive and test statistics). Participants from non-dominant religious groups showed implicit ingroup favoritism with good-focal MC-IATs but showed a *lack* of

Table 2

Test statistics for ingroup and outgroup favoritism in good-focal and bad-focal measures for Studies 1, 2 and 4.

Study 1

	Good-focal			Bad-focal		
	t	p	d [95% CI]	t	p	d [95% CI]
White participants						
White-Asian associations	15.49	< .001	0.19 [0.17, 0.22]	44.45	< .001	0.58 [0.55, 0.60]
White-Black associations	21.10	< .001	0.26 [0.24, 0.29]	50.25	< .001	0.65 [0.62, 0.68]
White-Hispanic associations	23.12	< .001	0.29 [0.26, 0.31]	57.17	< .001	0.74 [0.71, 0.77]
Asian participants						
White-Asian associations	−2.61	.010	−0.14 [−0.24, −0.04]	4.82	< .001	0.25 [0.15, 0.36]
Black participants						
White-Black associations	−5.79	< .001	−0.20 [−0.27, −0.13]	6.86	< .001	0.25 [0.18, 0.33]
Hispanic participants						
White-Hispanic associations	−1.83	.067	−0.06 [−0.12, 0.003]	12.81	< .001	0.43 [0.36, 0.50]

Study 2

	Good-focal			Bad-focal		
	t	p	d [95% CI]	t	p	d [95% CI]
Christian participants						
Christianity-Judaism associations	23.20	< .001	1.05 [0.94, 1.15]	12.34	< .001	0.58 [0.48, 0.68]
Christianity-Hinduism associations	26.98	< .001	1.21 [1.09, 1.32]	12.70	< .001	0.60 [0.50, 0.69]
Christianity-Islam associations	26.81	< .001	1.21 [1.10, 1.33]	15.30	< .001	0.74 [0.63, 0.85]
Jewish participants						
Christianity-Judaism associations	−5.84	< .001	−0.57 [−0.76, −0.36]	−0.90	.376	−0.09 [−0.29, 0.11]
Hindu participants						
Christianity-Hinduism associations	−4.26	< .001	−0.60 [−0.90, −0.31]	−0.16	.876	−0.02 [−0.16, 0.11]
Muslim participants						
Christianity-Islam associations	−5.45	< .001	−0.62 [−0.86, −0.38]	0.87	.386	0.13 [−0.16, 0.41]

Study 4

	Good-focal			Bad-focal		
	t	p	d [95% CI]	t	p	d [95% CI]
White participants						
White-Asian associations	6.47	< .001	0.37 [0.25, 0.48]	15.80	< .001	0.90 [0.77, 1.03]
White-Black associations	9.17	< .001	0.55 [0.42, 0.67]	13.37	< .001	0.80 [0.67, 0.93]
White-Hispanic associations	5.68	< .001	0.34 [0.22, 0.47]	15.25	< .001	0.92 [0.78, 1.06]
Asian participants						
White-Asian associations	−3.31	.001	−0.31 [−0.50, −0.12]	4.45	< .001	0.41 [0.22, 0.60]
Black participants						
White-Black associations	−3.51	.001	−0.25 [−0.38, −0.11]	0.22	.828	0.02 [−0.12, 0.15]
Hispanic participants						
White-Hispanic associations	−1.75	.083	−0.13 [−0.28, 0.02]	3.94	< .001	0.29 [0.14, 0.44]

Note. All BIAT's scored such that higher values mean more positive associations with group listed first. Test statistics report one-sample *t*-tests against zero. See Table 1 for sample size information.

implicit ingroup or outgroup favoritism on bad-focal MC-IATs (see Table 2 for test statistics).

Unlike Whites with racial attitudes, Christian participants showed more ingroup favoritism on good-focal than bad-focal MC-IATs (see Table 1), though both good-focal and bad-focal measures showed reliable implicit ingroup favoritism (see Table 2).

3.2.3. Explicit evaluations

See Table 3 for descriptive statistics of explicit evaluations. Participants from each non-dominant religious group reported preferring people from their own religion relative to Christian people (all *t*'s > 5.71, all *p*'s < .001, all *d*'s > 0.53; see online supplement for individual tests), and Christian participants reported preferring Christian people to people from each other religion (all *t*'s > 19.36, all *p*'s < .001, all *d*'s > 0.57; see online supplement).

3.2.4. Relationship between explicit and implicit evaluations

Good-focal MC-IATs were more related to explicit preferences than bad-focal MC-IATs (Christianity vs. Judaism: Good: $r(674) = 0.32$, $p < .001$, Bad: $r(596) = 0.13$, $p = .001$, Fisher's $z = 3.49$, $p < .001$, Christianity vs. Hindu: Good: $r(674) = 0.36$, $p < .001$, Bad: $r(601) = 0.12$, $p = .002$, Fisher's $z = 4.44$, $p < .001$, Christianity vs. Islam: Good: $r(669) = 0.37$, $p < .001$, Bad: $r(571) = 0.15$, $p < .001$,

Fisher's $z = 4.18$, $p < .001$).

3.3. Discussion

Participants from non-dominant religious groups showed implicit ingroup favoritism with measures focusing on good valence, and no evidence of ingroup or outgroup favoritism with measures focusing on bad valence. Unlike White participants, Christian participants showed stronger implicit ingroup favoritism with good-focal measures than with bad-focal measures, though both measures revealed ingroup favoritism.

Across all studies, religion is the only case where the dominant group showed stronger implicit ingroup favoritism with good-focal than bad-focal measures. If cultural negativity towards non-Christian religions is weaker than Christian's individual positive attitudes towards Christianity, then these data are consistent with a cultural learning or a system justification hypothesis. Supporting this possibility, explicit attitudes were more strongly related with good-focal than bad-focal measures. We return to this issue and highlight possible explanations for this discrepancy in the General Discussion.

Study 3 tested whether good-focal and bad-focal measures differed within both non-dominant and dominant members in another domain, sexual orientation.

Table 3
Means and standard deviations for explicit evaluations in Studies 1 and 2.

Study 1	Explicit M (SD)
White participants	
White-Asian preferences	0.43 (0.88)
White-Black preferences	0.52 (0.92)
White-Hispanic preferences	0.48 (0.93)
Asian participants	
White-Asian preferences	−0.56 (1.30)
Black participants	
White-Black preferences	−0.76 (1.28)
Hispanic participants	
White-Hispanic preferences	−0.40 (1.20)

Study 2	Explicit M (SD)
Christian participants	
Christian-Jewish preferences	0.65 (1.11)
Christian-Hindu preferences	0.89 (1.21)
Christian-Muslim preferences	1.05 (1.28)
Jewish participants	
Christian-Jewish preferences	−1.06 (1.21)
Hindu participants	
Christian-Hindu preferences	−0.65 (1.20)
Muslim participants	
Christian-Muslim preferences	−0.82 (1.26)

Note. Means and standard deviations (in parentheses). All explicit evaluations scored such that higher values mean more positive associations with group listed first.

4. Study 3

4.1. Method

4.1.1. Participants

We collected data until there were 40 gay participants with eligible BIAT data. Of the 1621 participants who volunteered, consented and provided data for the study at the Project Implicit research pool, we included 1093 Americans (M_{age} = 30.3, SD_{age} = 13.5, 67.0% female).[5] The final sample size provided 80% power at detecting a within-subjects effect size of d = 0.10 for straight participants and d = 0.45 for gay participants.

4.1.2. Procedure

The study consisted of three components: an explicit attitudes questionnaire followed by a sexual orientation demographic item, and both good-focal and bad-focal sexual orientation BIATs. The two BIATs were completed in sequence, but all other components were randomized.

4.1.2.1. Demographics and attitudes survey. Participants first completed three items concerning attitudes towards gay and straight people. One item assessed relative preferences for gay versus straight people similar to those used in previous studies, and the other two items separately gauged feelings of warmth towards gay and straight people (an 11-point scale ranging from "Extremely cold" to "Extremely warm"). Participants reported their sexual orientation using five response options ("Heterosexual, straight", "Gay or Lesbian", "Bisexual", "Asexual", "Questioning").

4.1.2.2. Indirect measure. Participants completed both a good-focal and bad-focal BIATs measuring implicit evaluation of gay and straight

people. Stimuli for each sexuality category consisted of two words (Gay, Homosexual, Straight, Heterosexual) and two drawings of hand-holding (same-sex or opposite-sex) couples.

Each BIAT contained five blocks. In the first (practice) block of a good-focal BIAT (16 trials), participants pressed the "I" key for all Good words (Love, Pleasant, Great, Wonderful) and the "E" key for "other words", which were Bad words (Hate, Unpleasant, Awful, Terrible). In the second block (20 trials), participants pressed the "I" key for all Good words and words or images related to Gay people, and the "E" key for "any other images and words." These other stimuli were Bad words and words or images related to Straight people. The third block (20 trials) had the same design, but the "I" key was for Good words and words or images related to Straight people and the "E" key for "any other images or words". The fourth and fifth blocks repeated the second and third blocks respectively. The key assignments described for the second/fourth and third/fifth blocks were randomized between participants. Bad-focal BIATs had the same design, the only change being the "I" key was used to categorize all Bad words and members of the focal group, and the "E" key was used "for other words", which were always Good words and the other group.

Participants were assigned to complete the two BIATs in one of four orders, randomizing whether good-focal or bad-focal BIATs were completed first and whether blocks pairing Gay or Straight people with the focal category occurred first. We excluded 3.9% of the participants for the same exclusion rule as in Study 1.

4.2. Results

4.2.1. Reliabilities

Good-focal and bad-focal BIATs were comparably reliable (Good: α = 0.82, Bad: α = 0.81).[6]

4.2.2. Implicit evaluations

For Studies 3–5, we tested for differences between good-focal and bad-focal BIATs using a series of dependent samples t-tests. Among gay participants, good-focal and bad-focal BIATs reliably differed from each other, $t(39)$ = 4.28, p < .001, d = 0.68. Gay participants showed ingroup favoritism on the good-focal BIAT (M = −0.24, SD = 0.47, $t(39)$ = −3.80, p = .001, d = −0.51), and no reliable preference on the bad-focal BIAT (M = 0.07, SD = 0.45, $t(39)$ = 1.03, p = .310, d = 0.16).

Straight participants also showed reliable differences in good-focal and bad-focal BIATs, $t(775)$ = 8.38, p < .001, d = 0.30. Straight participants showed implicit ingroup favoritism in both measures, though more strongly on bad-focal (M = 0.47, SD = 0.47, $t(775)$ = 27.96, p < .001, d = 1.00) than on good-focal (M = 0.31, SD = 0.47, $t(775)$ = 18.58, p < .001, d = 0.66) BIATs.

Participants grouped from all other sexual orientation categories (bisexual, asexual, questioning; n = 71) showed reliable differences between good-focal and bad-focal BIATs, $t(70)$ = 3.52, p = .001, d = 0.42. Participants showed a lack of implicit favoritism on the good-focal BIAT (M = −0.02, SD = 0.49, $t(70)$ = −0.41, p = .681, d = −0.04) but implicit straight favoritism on the bad-focal BIAT (M = 0.22, SD = 0.47, $t(70)$ = 3.86, p < .001, d = 0.47).

4.2.3. Explicit evaluations

Gay (M = −0.98, SD = 1.27 for the preference item) and straight participants (M = 0.66, SD = 1.13) both reported explicit ingroup preference (Gay participants: $t(39)$ = −4.85, p < .001, d = −0.76; Straight participants: $t(781)$ = 16.38, p < .001, d = 0.58). Participants grouped from all other sexual orientations had a slight but not reliable

---

[5] The study had 2037 started sessions, with 1621 providing data, and 1320 completing the study (64.8% completion rate).

[6] Reliabilities for the BIATs in Studies 3–5 were higher than reliabilities of the MC-IATs used in Studies 1–2 as BIATs had twice as many trials for each contrast *D* score (e.g., evaluations of White vs. Black people) as MC-IATs.

preference for gay versus straight people ($M = -0.13$, $SD = 0.63$, $d = -0.21$; $t(70) = -1.69$, $p = .095$).

4.2.4. Relationship between explicit and implicit evaluations

In Studies 3–5, we used a Williams' t-test (Steiger, 1980) to compare dependent correlations. Good-focal BIATs were more correlated with explicit evaluations than bad-focal BIATs (Explicit preference: Good r (888) = 0.41, $p < .001$, Bad $r(888) = 0.25$, $p < .001$, Williams' t (885) = 4.56, $p < .001$, Warmth towards gay people: Good r (887) = -0.25, $p < .001$, Bad $r(887) = -0.16$, $p < .001$, Williams' t (884) = 2.46, $p = .014$, Warmth towards straight people: Good r (886) = 0.22, $p < .001$, Bad $r(886) = 0.15$, $p < .001$, Williams' t (883) = 2.03, $p = .043$).

Since each participant completed both good- and bad-focal BIATs, we also tested whether each measure accounted for some variation in explicit evaluations using simultaneous linear regression. Both good-focal and bad-focal BIAT scores uniquely predicted explicit preferences (Good $\beta = 0.36$, $t = 11.02$, $p < .001$; Bad $\beta = 0.11$, $t = 3.38$, $p = .001$), warmth towards gay people (Good $\beta = -0.22$, $t = -6.29$, $p < .001$; Bad $\beta = -0.08$, $t = -2.16$, $p = .031$), and warmth towards straight people (Good $\beta = 0.20$, $t = 5.50$, $p < .001$; Bad $\beta = 0.08$, $t = 2.11$, $p = .035$).

4.3. Discussion

Gay participants showed implicit ingroup favoritism with measures focusing on good valence but no evidence of ingroup or outgroup favoritism with measures focusing on bad valence. Straight participants showed stronger implicit ingroup favoritism with good-focal measures than with bad-focal measures, though both revealed ingroup favoritism. Participants from other sexual orientations showed a lack of ingroup or outgroup favoritism on good-focal measures but pro-straight preferences on bad-focal measures. Explicit evaluations were more related to good-focal than bad-focal indirect measures, though both measures uniquely predicted explicit evaluations.

5. Study 4

In Studies 1–3, racial, religious, and sexual minorities showed implicit ingroup favoritism with measures of positive valence, and no preference or implicit outgroup favoritism with measures of negative valence. Moreover, good-focal measures were a stronger predictor of explicit evaluations across dominant and non-dominant group members. These results conflict with those of dominant group members, who showed implicit ingroup favoritism on good-focal and bad-focal measures.

Having established that the effect occurs across three social domains, we next turned to testing the hypothesis that bad associations show diminished ingroup (or even outgroup) favoritism among members of non-dominant groups because they are more sensitive to cultural context. In Study 4, we tested that hypothesis using self-reported cultural knowledge. We investigated whether self-reports of perceived racial preferences in American culture would be more strongly related to bad-focal than good-focal measures. Though previous investigations have not found reliable associations between self-reports of perceived cultural preferences and the IAT (Nosek & Hansen, 2008a), it is possible that measures of good and bad implicit associations may reveal different outcomes, much as they did when finding differing levels of ingroup and outgroup favoritism among participants from non-dominant groups in Studies 1–3.

5.1. Method

5.1.1. Participants

The study was restricted to American citizens or residents who were either White, Asian, Black or Hispanic. 1563 participants consented and provided data for the study at the Project Implicit research pool (Mage = 35.1, SDage = 14.5, 64.6% female).[7] For analyses, participants were classified as White ($n = 981$), Black ($n = 242$), or East Asian ($n = 134$) if reporting that race and ethnicity as not Hispanic, and as Hispanic ($n = 206$) if reporting ethnicity as Hispanic. This sample size provided at least 80% power at detecting a within-subjects effect size of $d = 0.30$.

5.1.2. Procedure

Participants completed different BIATs depending on their race or ethnicity. Asian, Black and Hispanic participants completed both good-focal and bad-focal BIATs using their own group and White people as categories. White participants completed both good and bad-focal BIATs using one randomly chosen pair of categories out of White and Hispanic people, White and Black people, or White and Asian people.

The study consisted of three components: an attitudes survey dealing with personal and cultural preferences, and both good-focal and bad-focal BIATs. The BIATs were completed in sequence with order randomized between subjects. All other study components were presented in a randomized order.

5.1.2.1. Attitudes survey. Participants completed three components in a random order. To measure race preferences, participants completed the 7-point single-item measure from Study 1 for the two racial groups in that condition. To measure cultural preferences, participants completed three items adapted from Nosek and Hansen (2008a). In the first item, participants reported which statement best described American culture (-3 = "The American culture strongly prefers X to White people", $+3$ = "The American culture strongly prefers White to X people"). In the two additional items, participants reported how much the American culture liked each of the two racial groups in that condition (1 = "Strongly dislike", 7 = "Strongly like"). Participants also completed measures of self-reported positivity and negativity towards the racial groups used in each BIAT, which were not included in analyses but are available in the online dataset.

5.1.2.2. Indirect measure. The good-focal and bad-focal BIATs had the same design as in Study 3, now using the racial stimuli from Study 1. Data exclusion was the same as in Study 3 (2.4% of participants' BIAT data).

5.2. Results

5.2.1. Reliabilities

The good-focal and bad-focal BIATs were comparably reliable, for White-Asian (Good: $\alpha = 0.81$, Bad: $\alpha = 0.78$), White-Black (Good: $\alpha = 0.82$, Bad: $\alpha = 0.82$), and White-Hispanic (Good: $\alpha = 0.84$, Bad: $\alpha = 0.82$) associations.

5.2.2. Implicit evaluations

As in Study 1, participants from non-dominant racial groups showed reliable differences on good-focal and bad-focal measures (see Table 1 for descriptive and test statistics). Participants showed implicit ingroup favoritism on good-focal BIATs, (see Table 2 for test statistics), though evidence was again relatively weak among Hispanic participants. Asian and Hispanic participants showed reliable implicit outgroup favoritism on bad-focal BIATs, though this did not occur among Black participants (see Table 2).

Also replicating Study 1, White participants showed more implicit ingroup favoritism on bad-focal than good-focal BIATs (see Table 1), though both measures revealed implicit ingroup favoritism (see Table 2).

---

[7] The study had 1916 started sessions, with 1563 providing data, and 1466 completing the study (76.5% completion rate).

Table 4
Means and standard deviations for self-report items in Studies 4 and 5.

Study 4

	Exp. Pref.	Cult. Pref.	Pos. Non-Dom Group	Pos. White	Neg. Non-Dom Group	Neg. White.	Cult. Non-Dom Group	Cult. White
White participants								
White-Asian	0.43 (0.82)	1.57 (1.10)	3.41 (1.05)	3.43 (1.11)	1.63 (0.79)	1.89 (0.94)	4.18 (1.49)	5.96 (1.42)
White-Black	0.43 (0.76)	1.65 (1.09)	3.41 (1.00)	3.40 (1.07)	1.91 (0.86)	1.81 (0.84)	3.33 (1.48)	5.72 (1.39)
White-Hispanic	0.35 (0.77)	1.87 (1.01)	3.36 (1.11)	3.39 (1.10)	1.73 (0.77)	1.99 (0.92)	3.37 (1.35)	5.79 (1.47)
Minority participants								
Asian	−0.30 (1.03)	1.89 (1.02)	3.56 (1.01)	3.25 (0.97)	1.83 (0.86)	1.99 (0.96)	3.81 (1.34)	5.94 (1.44)
Black	−0.67 (1.06)	2.42 (0.91)	3.96 (1.07)	3.33 (1.07)	1.84 (0.87)	2.22 (0.99)	2.54 (1.54)	5.93(1.75)
Hispanic	−0.37 (1.07)	2.04 (1.11)	3.88 (1.04)	3.49 (1.01)	1.79 (0.90)	1.98 (1.00)	3.13 (1.44)	6.00 (1.50)
Study 5	Exp. Pref.	Cult. Pref.	Pos. Jew	Pos. Christ.	Neg. Jew	Neg. Christ.	Cult. Jew	Cult. Christ.
Sample								
American	0.89 (1.08)	−1.59 (1.25)	4.30 (0.76)	3.37 (1.00)	1.40 (0.74)	1.88 (1.02)	3.90 (1.34)	6.03 (1.10)
Israeli	1.10 (1.23)	1.95 (1.16)	4.02 (1.01)	3.15 (1.21)	1.65 (0.79)	1.67 (0.91)	6.23 (1.19)	4.32 (1.25)

Note. Means and standard deviations (in parentheses). *Exp. Pref* (−3 to +3) = explicit preference for own group compared to other group. *Cult. Pref* (−3 to +3) = cultural preference for White versus non-dominant racial group (preference for Jews over Christians in Study 5). *Pos. Non-Dom Group* (1 to 5) = positivity towards non-dominant group. *Pos. White* (1 to 5) = positivity towards White people. *Neg. Non-Dom Group* (1 to 5) = negativity towards non-dominant group. *Neg. White* (1 to 5) = negativity towards White people. *Cult. Non-Dom Group* (1 to 7) = cultural liking of non-dominant group. *Cult. White.* (1 to 7) = cultural liking of White people. *Pos. Jew* (1 to 5) = positivity towards Jews. *Pos. Christ.* (1 to 5) = positivity towards Christians. *Neg. Jew* (1 to 5) = negativity towards Jews. *Neg. Christ* (1 to 5) = negativity towards Christians. *Cult. Jew* (1 to 7) = cultural liking of Jews. *Cult. Christ.* (1 to 7) = cultural liking of Christians.

5.2.3. Explicit evaluations

Table 4 shows descriptive statistics for the self-report measures. As in Study 1, Asian, Black and Hispanic participants preferred people from their own group relative to White people (all *t*'s > 3.29, all *p*'s < .002, all *d*'s > 0.29; see online supplement for individual tests), and White participants preferred White people to people from each other racial group (all *t*'s > 7.97, all *p*'s < .001, all *d*'s > 0.46; see online supplement). Using another series of one-sample *t*-tests against a neutral value of 0, participants from all racial groups reported that the American culture preferred White people to the non-dominant racial group in that condition (all *t*'s > 20.93, all *p*'s < .001, all *d*'s > 1.43; see online supplement for individual tests).

5.2.4. Implicit-explicit correlations

If the distinction between implicit good- and bad-associations corresponds with the distinction between self-reported preferences and perceptions of cultural preferences, then implicit associations with good should be more related to self-reported preferences and implicit associations with bad should be more related to perceptions of cultural preferences. Unlike prior studies, for explicit preferences, there were no reliable differences in strength of correlation for good-focal versus bad-focal measures for White-Asian associations (Good *r*(402) = 0.29, *p* < .001; Bad *r*(402) = 0.20, *p* < .001; Williams' *t*(399) = 1.48, *p* = .139), White-Black associations (Good *r*(462) = 0.31, *p* < .001; Bad *r*(462) = 0.36, *p* < .001; Williams' *t*(459) = −0.99, *p* = .325) or White-Hispanic associations (Good *r*(436) = 0.33, *p* < .001; Bad *r*(436) = 0.31, *p* < .001 Williams' *t*(433) = 0.23, *p* = .816).

Moreover, greater belief that the American culture favors White people was associated with *lower* levels of pro-White implicit associations for both good-focal and bad-focal measures. There were no reliable differences in strength of correlation for good-focal versus bad-focal measures for White-Asian associations (Good *r*(404) = −0.10, *p* = .036; Bad *r*(404) = −0.10, *p* = .041; Williams' *t*(401) = −0.03, *p* = .975), White-Black associations (Good *r*(466) = −0.25, *p* < .001; Bad *r*(466) = −0.23, *p* < .001; Williams' *t*(463) = −0.38, *p* = .708) or White-Hispanic associations (Good *r*(436) = −0.16, *p* = .001; Bad *r*(436) = −0.07, *p* = .143, Williams' *t*(433) = −1.64, *p* = .102). These results suggest that although mean levels of self-reported cultural preferences showed a cultural preference for the dominant racial group, they did not capture variability in the expected direction for the dominant-group favoritism detected with the indirect measure focusing on negative valence.

5.3. Discussion

Replicating Study 1, participants from non-dominant racial groups showed implicit ingroup favoritism with measures of good associations, and implicit outgroup favoritism with measures of bad associations, though this was closer to no preference on bad associations among Black participants. White participants showed stronger implicit ingroup favoritism on indirect measures focused on negative than positive valence. Unlike prior studies, good-focal measures were not more related to explicit evaluations, though the evidence across studies still supports this claim.

We did not find evidence in support of the hypothesis that bad associations reflect self-reported perceptions of cultural values. Self-reported cultural preferences were negatively related to both bad-association and good-association measures. These negative correlations may indicate that the cultural preference measure was sensitive to factors other than simple exposure to or awareness of cultural preferences. For instance, reports of the degree to which the American culture favored Whites was positively correlated with greater liberalism, which was reported by participants when registering for the research site (White-Asian perceptions: *r*(393) = 0.26, *p* < .001; White-Black perceptions: *r*(449) = 0.39, *p* < .001; White-Hispanic perceptions: *r*(425) = 0.27, *p* < .001). These results suggest that perceptions of group-based cultural preference may capture, at least partly, individuals' ideology or motivation.

After our failure to use self-report measures to assess the factors captured by indirect measures of bad associations, we turned to comparing across cultures with differing values. Results from Study 4 suggest that asking participants to characterize their perceptions of cultural preferences may not help explain the distinction between good and bad associations. However, it is possible that cultural factors do not have substantial influence on interindividual differences within a national culture and are easier to detect when comparing across national cultures. That is, individuals may be able to broadly recognize which groups are preferred by their culture (e.g., 93% of participants from non-dominant racial groups and 87% of White participants in Study 4 reported the American culture preferred White people) but people do not have particularly accurate insight into the relative strength of that preference (i.e., whether one's culture slightly or moderately prefers White to Black people), which would make analyses conducted within only one culture uninformative. As such, comparisons across cultures with opposing social preferences may be more revealing than a

comparison within one culture.

6. Study 5

Using a cross-cultural approach, Study 5 examined Jewish participants from two cultures with divergent religious favoritism (United States vs. Israel). We examined whether—as the cultural learning and system justification account would predict—bad-focal measures would show stronger preference for the ingroup when the ingroup is the dominant versus the non-dominant group.

Study 5 provides a more direct test concerning whether the reduced ingroup favoritism displayed by members of non-dominant groups in Studies 1–3 reflected cultural values or messages about the non-dominant ingroup. In Study 5, we compared performance on good-focal and bad-focal indirect measures of implicit evaluations among two samples that shared an ingroup identity but differed in cultural context. Specifically, we compared Jewish participants living in the United States (where Jews are not the culturally dominant group) versus Israel (where Jews are the culturally dominant group). If good-focal measures are more sensitive to ingroup identity, then cultural context should not moderate the strength of ingroup favoritism shown on good-focal measures among Israeli and American Jewish participants. Conversely, if bad-focal measures are sensitive to cultural messages about group standing, then cultural context should moderate the degree of ingroup favoritism shown on bad-focal measures, with American Jews displaying weaker ingroup favoritism on bad-focal measures compared to Israeli Jews.

In this sense, Study 5 sought to complete the double dissociation that complements Studies 1–4. In Studies 1–4, participants from non-dominant and dominant groups shared a cultural context but differed in group identity. Here, participants showed greater dissimilarity on good-focal measures, where each group demonstrated ingroup favoritism, and greater similarity on bad-focal measures, where participants from non-dominant groups showed diminished ingroup favoritism or outgroup favoritism towards the culturally dominant group. In Study 5, we tested whether the opposite pattern would emerge when comparing participants with the same ingroup identity but different cultural contexts; specifically, that participants sharing a group identity but differing in cultural context would show greater dissimilarity on bad-focal measures and greater similarity on good-focal measures.

6.1. Method

6.1.1. Participants

Jewish participants from the United States and Israel were recruited for the study. Israeli participants ($N = 94$; $M_{age} = 24.6$; 56.4% female) were undergraduates who completed the study in exchange for course credit. American participants were undergraduates ($N = 32$; $M_{age} = 20.3$; 65.6% female) recruited through a Jewish student organization who completed the study for a $5 gift card or online participants ($N = 143$; $M_{age} = 22.01$, assigned to the study only if in the age range 17–27; 61% female) at the Project Implicit research pool, preselected based on their reported religion. The final sample size allowed for 80% power to detect an interaction effect size of $d = 0.33$.

6.1.2. Procedure

The study consisted of four components: good-focal and bad-focal BIATs measuring evaluations of Christianity and Judaism, a demographics questionnaire, and an attitudes survey. The two BIATs were completed in sequence but all other study components were randomized. Americans completed the study in English and Israelis in Hebrew.

6.1.2.1. Indirect measure. The within-subjects good-focal and bad-focal BIATs had the same design as those used in Study 3, now using words related to Christianity and Judaism (from Study 2). We scored the

BIATs such that more positive scores meant more positive associations towards Judaism relative to Christianity. Data exclusion was the same as Study 3 (2.4% of participants' BIAT scores).

6.1.2.2. Demographics questionnaire and attitudes survey. Non-Project Implicit participants first completed a seven-item demographics questionnaire reporting age, gender, race, ethnicity, country of citizenship, current religious identification, and degree of religiosity. Project Implicit participants completed these items upon registering.

All participants completed the same self-report items from Study 4, updated to measure attitudes and preferences towards Christian and Jewish people. Items were scored such that more positive values indicated greater personal or cultural preference for Jewish relative to Christian people.

6.2. Results

Across the dependent measures, there were no significant differences between the online and undergraduate American participants (all p's > .10), so analyses combine the two samples.

6.2.1. Reliabilities

Good-focal and bad-focal BIATs were comparably reliable (Good: $\alpha = 0.70$; Bad: $\alpha = 0.74$).

6.2.2. Explicit evaluations

Descriptive statistics for all self-report measures are shown in Table 4. American participants believed that the American culture preferred Christians to Jews, $t(109) = -13.26$, $p < .001$, $d = 1.26$, and Israeli participants believed that the Israeli culture preferred Jews to Christians, $t(90) = 16.02$, $p < .001$, $d = 1.68$. Both American and Israeli participants explicitly preferred Jewish to Christian people (American participants: $t(109) = 8.67$, $p < .001$; $d = 0.83$; Israeli participants: $t(90) = 8.52$, $p < .001$, $d = 0.89$), and an independent samples t-test found no reliable difference between American samples in explicit religious preferences, $t(199) = 1.28$, $p = .203$, $d = 0.18$.

6.2.3. Implicit evaluations

Israeli and American Jews showed implicit ingroup favoritism on both good-focal measures (Israeli $M = 0.51$, $SD = 0.39$, $t(90) = 12.64$, $p < .001$, $d = 1.32$; American $M = 0.41$, $SD = 0.37$, $t(106) = 11.51$, $p < .001$, $d = 1.11$) and bad-focal measures (Israeli $M = 0.47$, $SD = 0.34$, $t(90) = 13.38$, $p < .001$, $d = 1.40$; American $M = 0.22$, $SD = 0.46$, $t(106) = 4.97$, $p < .001$, $d = 0.48$).

A 2 (Country: US vs. Israel) by 2 (BIAT: good vs. bad-focal) mixed-measures ANOVA revealed a significant main effect of BIAT type, $F(1,196) = 12.22$, $p = .001$, $\eta_p^2 = 0.06$, and a main effect of country, $F(1,196) = 15.57$, $p < .001$, $\eta_p^2 = 0.08$. These main effects were qualified by a significant location by BIAT interaction, $F(1,196) = 5.56$, $p = .019$, $\eta_p^2 = 0.03$ (Fig. 2). Replicating Study 2, American Jewish participants preferred Jews on the good-focal BIAT more than with the bad-focal BIAT, $t(106) = 4.11$, $p < .001$, $d = 0.40$. Conversely, Israeli participants showed no reliable differences in implicit preferences for Jews on the good-focal BIAT versus the bad-focal BIAT, $t(90) = 0.83$, $p = .410$, $d = 0.09$. Israeli Jews showed strong positivity towards Judaism with both good-focal and bad-focal BIATs while American Jews did so only on the good-focal BIAT.

6.2.4. Implicit-explicit correlations

We compared the strength of the correlation between the good-focal and bad-focal BIATs with self-report measures of explicit religious preference and perceived cultural religious preference (see online supplement for full correlation matrix). Good-focal BIATs were more strongly correlated with explicit preferences than were bad-focal BIATs (Good: $r(198) = 0.33$, $p < .001$; Bad: $r(198) = 0.17$, $p = .017$; Williams' $t(195) = 2.08$, $p = .039$). Both good-focal and bad-focal
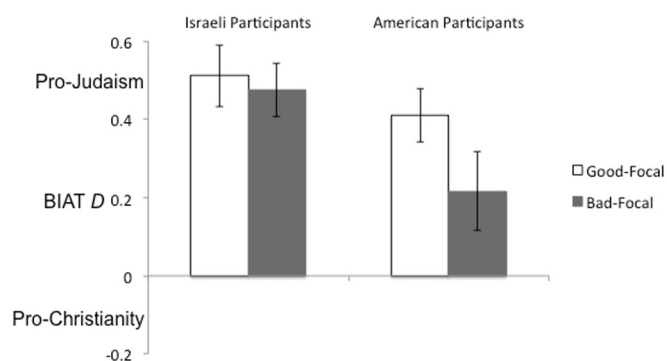
Fig. 2. Good-focal and bad-focal BIAT D scores for evaluations of Judaism vs. Christianity among Israeli and American Jewish participants in Study 5. Higher values mean more positive (less negative) evaluations of Judaism relative to Christianity.

BIATs were positively related with perceptions of cultural preferences, though with no reliable differences in the strength of these correlations (Good: $r(198) = 0.17$, $p = .014$, Bad: $r(198) = 0.25$, $p < .001$, Williams' $t(195) = 0.90$, $p = .370$). A simultaneous linear regression predicting explicit preferences found that good-focal BIATs contributed uniquely, $\beta = 0.31$, $t = 4.28$, $p < .001$, but not bad-focal BIATs, $\beta = 0.06$, $t = 0.84$, $p = .400$. Conversely, a simultaneous linear regression predicting cultural preferences found that bad-focal BIATs contributed uniquely, $\beta = 0.21$, $t = 2.84$, $p = .005$, but not good-focal BIATs, $\beta = 0.10$, $t = 1.36$, $p = .175$.

6.3. Discussion

American and Israeli Jews had opposing perceptions of their culture's preference towards Christian and Jewish people, but no reliable differences in self-reported personal preferences for Jewish versus Christian people. American Jewish participants showed stronger implicit ingroup favoritism with indirect measures focusing on positive valence than with indirect measures focusing on negative valence. However, Israeli Jews showed no reliable differences in implicit ingroup favoritism between good-focal and bad-focal measures. Good-focal measures uniquely predicted explicit preferences, whereas bad-focal measures did not; and bad-focal measures uniquely predicted perceived cultural preferences, whereas good-focal measures did not.

These results are consistent with an account that good-focal measures are more sensitive to the positive information people derive from their ingroup identity and bad-focal measures are more sensitive to the negative information about non-dominant groups that stem from cultural messages or values. However, the findings related to the self-reported cultural preferences should be interpreted with caution given the contradictory results from Study 4, which suggested that self-reported perceptions of cultural values are likely influenced by other external factors. More informative are the mean-level results illustrating that while Israeli and American Jews showed comparable levels of ingroup favoritism in explicit evaluations and good-focal measures, only those living in a culture where they were the non-dominant group demonstrated weakened ingroup favoritism on bad-focal measures.

The correlational results regarding perceived cultural preferences differ from Study 4, where self-reported cultural preferences were not positively correlated with performance on bad-focal measures. The key difference, in our estimation, is that Study 5 included samples from two distinct national cultures. Detecting cultural impacts on implicit associations may be easier when including multiple national cultures rather than relying on variations in perceptions within a single national culture. Moreover, this suggests that insight into the cultural factors that shape bad associations may be limited to the general sense of which groups are preferred to others across cultures.

6.4. Additional studies

In addition to the five studies reported here, eight less informative studies are detailed in the online supplement (https://osf.io/u7tkp). We described Studies S1 and S2 in the Introduction. Studies S3-S5 (N's = 628, 239, 232) were lab replications of Studies 1–3, due to slightly higher dropout rates among participants in Studies 1–3 who completed bad-focal measures than those who completed good-focal measures. These studies replicated Studies 1–3's results on good-focal and bad-focal measures among dominant group members. Because it was harder to recruit samples of participants from non-dominant groups in the lab, Studies S3-S5 combined minorities for each social domain to maximize power, which still remained relatively low. Studies S3 and S5 replicated the results of racial and sexual non-dominant group members seen in Studies 1 and 3, but Study S4 did not replicate the pattern observed among non-dominant religious group members in Study 2.

Study S6 ($N = 8358$) tested the account that bad-focal measures reflect explicit system justification motives. White, Asian, Black and Hispanic participants completed Study 4's good-focal and bad-focal BIATs as well as measures of explicit system justification motives (e.g., Group-Based Dominance; Sidanius & Pratto, 1999). Contrary to the system justification account, good-focal measures were more related to explicit system justification than bad-focal measures. However, these findings are probably of limited informative value because self-report measures capture explicit system justification motives. If bad-focal measures are less related to self-endorsed attitudes in general, then they should also be less strongly correlated with explicit system justification motives. Possibly, the system justification motives that influence bad-focal measures are implicit rather than explicit. Therefore, Study S4 did not refute or confirm the perspective that negative associations are more sensitive to cultural learning or system justifying forces that may be consciously disavowed.

Finally, Studies S7 ($N = 231$) and S8 ($N = 215$) tested a methodological account of the present results. Because negative information and information about minority groups (which in our studies were the same as non-dominant groups) may be more salient, salience asymmetries could have impacted bad-focal BIAT performance (Rothermund & Wentura, 2004). Many of our bad-focal BIAT results among participants from non-dominant groups could be explained by a heuristic of "faster responses to minority groups and whatever is focal". Studies S7–S8 then tested whether bad-focal BIATs would always produce negative evaluations towards more salient, minority stimuli. Participants completed good and bad-focal BIATs measuring evaluations of Babies vs. Adults (Study S7) and Celebrities vs. Non-Celebrities (Study S8). Contrary to the salience asymmetry account, minority stimuli (babies and celebrities) were evaluated more positively on both measures.

6.5. Synthesis of key results

We meta-analyzed the results of Studies 1–4, and Studies S3–S6 (with preference for the dominant groups coded as positive scores). Dominant group members showed ingroup favoritism on good-focal measures ($M = 0.18$, $SD = 0.56$, $d = 0.32$), and more so on bad-focal measures ($M = 0.36$, $SD = 0.54$, $d = 0.66$). By social domains, White and straight participants showed more ingroup favoritism on bad-focal ($M_{\text{White}} = 0.36$, $SD_{\text{White}} = 0.54$, $d = 0.66$; $M_{\text{Straight}} = 0.47$, $SD_{\text{Straight}} = 0.46$, $d = 1.02$) than good-focal measures ($M_{\text{White}} = 0.16$, $SD_{\text{White}} = 0.55$, $d = 0.28$; $M_{\text{Straight}} = 0.30$, $SD_{\text{Straight}} = 0.46$, $d = 0.65$). This was reversed for Christian participants, who showed stronger ingroup favoritism on good-focal measures ($M = 0.53$, $SD = 0.45$, $d = 1.18$) than bad-focal measures ($M = 0.31$, $SD = 0.50$; $d = 0.62$). See Fig. 3 for a forest plot of BIAT D scores among dominant group members.

Across members of non-dominant groups, good-focal measures revealed ingroup favoritism ($M = -0.10$, $SD = 0.55$, $d = -0.17$) and bad-focal measures revealed outgroup favoritism ($M = 0.14$,
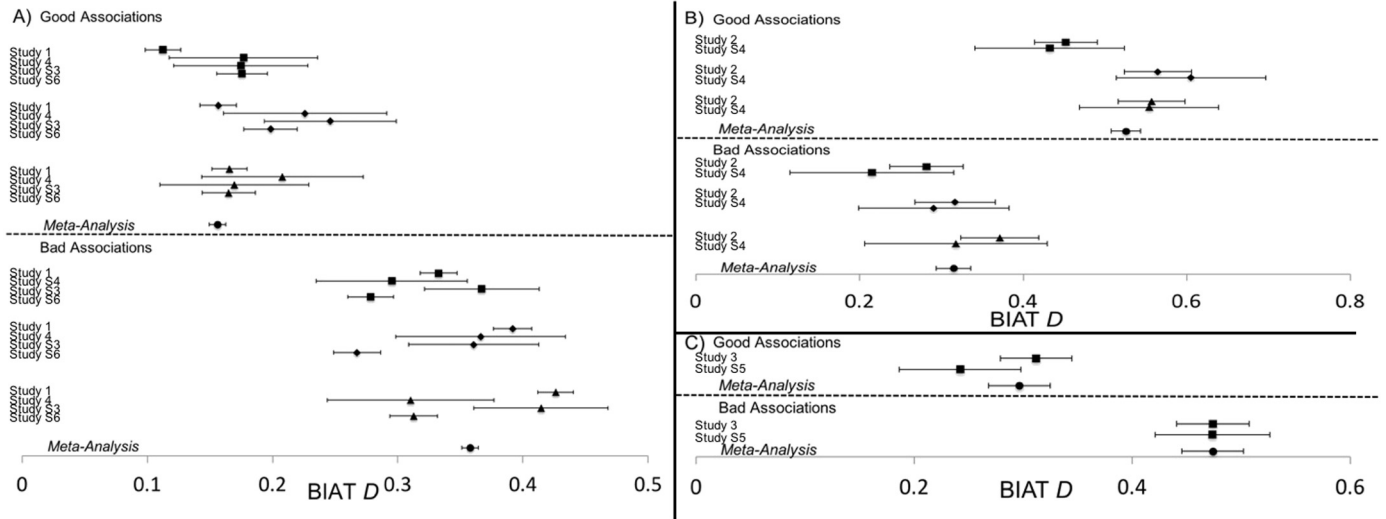
Fig. 3. Average BIAT D scores among White (Panel A), Christian (Panel B) and Straight (Panel C) participants. More positive scores indicate greater ingroup favoritism. ■ = White-Asian, Christianity-Judaism, and Straight-Gay associations. ◆ = White-Black and Christianity-Hinduism associations. ▲ = White-Hispanic and Christianity-Islam associations. Error bars indicate 95% confidence intervals.
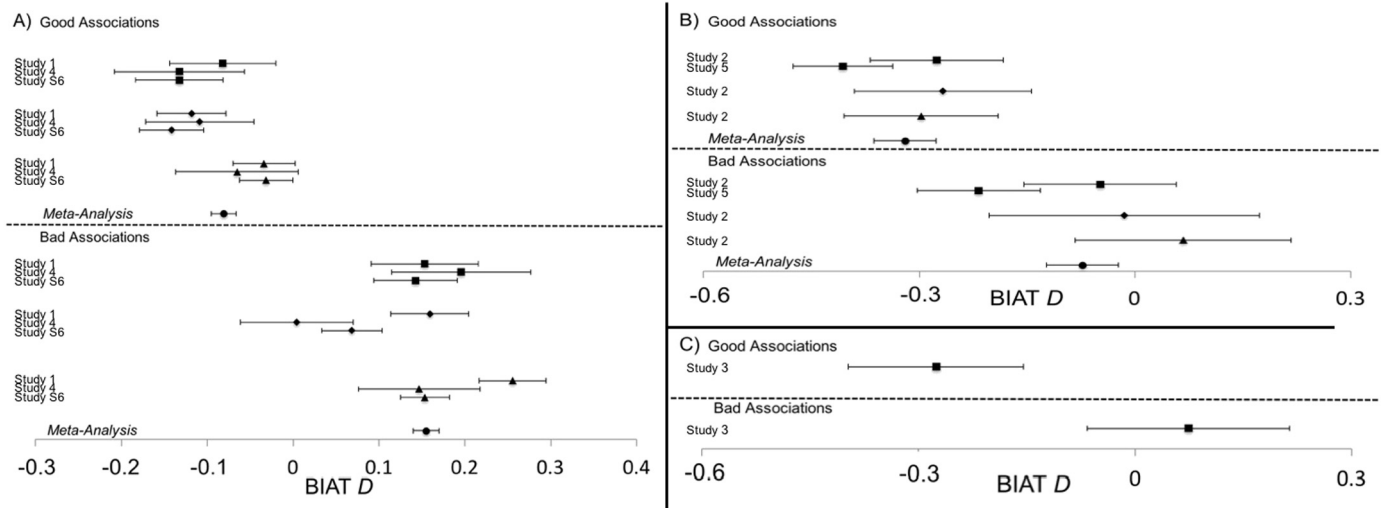

Fig. 4. Average BIAT D scores among racial (Panel A), religious (Panel B) and sexual (Panel C) non-dominant group members. More positive scores indicate greater positivity for the dominant group. ■ = Asian, American Jewish, and Gay participants. ◆ = Black and Hindu participants. ▲ = Hispanic and Muslim participants. Error bars indicate 95% confidence intervals.

$SD = 0.54$, $d = 0.26$). Splitting by social domains, racial and sexual minorities displayed ingroup favoritism on good-focal measures ($M_{Race} = -0.08$, $SD_{Race} = 0.55$, $d = -0.14$; $M_{Gay} = -0.25$, $SD_{Gay} = 0.38$, $d = -0.66$) and outgroup favoritism on bad-focal measures ($M_{Race} = 0.15$, $SD_{Race} = 0.55$, $d = 0.28$; $M_{Gay} = 0.07$, $SD_{Gay} = 0.45$, $d = 0.16$). Religious minorities showed ingroup favoritism on good-focal measures ($M_{Religion} = -0.30$, $SD_{Religion} = 0.46$, $d = -0.65$) and similarly, but more weakly, on bad-focal measures ($M_{Religion} = -0.09$, $SD_{Religion} = 0.52$, $d = -0.17$). See Fig. 4 for a forest plot of BIAT D scores among non-dominant group members for studies having sufficient power.

Using weighted correlations, good-focal measures ($r = 0.210$, 95% CI [0.201, 0.219]) were more strongly correlated with explicit preferences than bad-focal measures ($r = 0.149$, 95% CI [0.140, 0.158]). This difference was present when looking within racial associations (Good $r = 0.196$, 95% CI [0.187, 0.205]; Bad $r = 0.147$, 95% CI [0.138, 0.156]), religious associations (Good $r = 0.332$, 95% CI [0.299, 0.364]; Bad $r = 0.134$, 95% CI [0.096, 0.172]) and sexual associations (Good $r = 0.400$, 95% CI [0.350, 0.447]; Bad $r = 0.230$, 95% CI [0.174,

0.284]). See Fig. 5 for a forest plot of correlations of explicit evaluations with good-focal and bad-focal measures.

Figs. 3–5 provide a clear image of the consistencies and inconsistencies across studies. Among non-dominant group members, good-focal measures always indicated more ingroup favoritism than bad-focal measures, though bad-focal measures varied in the extent to which they indicated less ingroup favoritism, no reliable preference (like in religious or sexual associations), or reliable outgroup favoritism (like in racial associations). Among dominant group members, good-focal and bad-focal measures always indicated ingroup preference, though bad-focal measures did so more strongly for White and straight participants, whereas the opposite occurred for Christian participants. These figures provide summaries of average estimates across studies, but participants also showed potentially meaningful variation in displaying implicit ingroup or outgroup favoritism. The online supplement provides histograms and distribution statistics, as well as the percentage of participants displaying ingroup or outgroup favoritism across studies for both dominant and non-dominant group members.

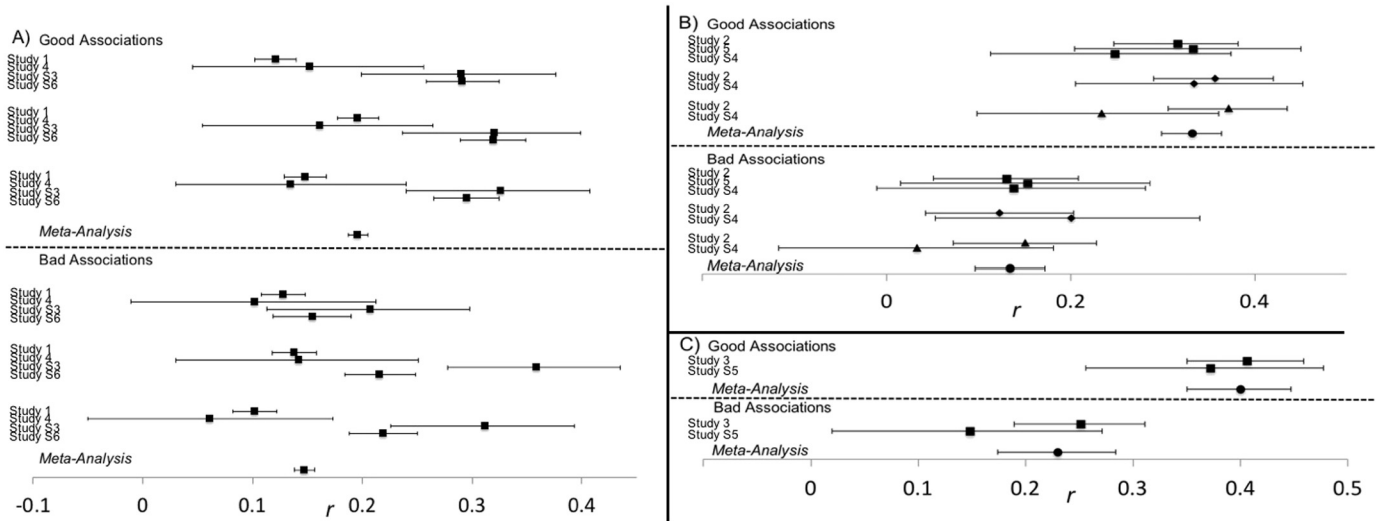Finally, good-focal measures were more related to explicit

Fig. 5. Implicit-explicit correlations in racial (Panel A), religious (Panel B) and sexual (Panel C) associations. ■ = White-Asian, Christianity-Judaism, and Straight-Gay associations. ◆ = White-Black and Christianity-Hinduism associations. ▲ = White-Hispanic and Christianity-Islam associations. Error bars indicate 95% confidence intervals on the correlation.

preferences in each social domain, but this difference was much smaller in racial compared to religious or sexual attitudes. While such results are compatible with the accounts provided here, the variability in how these results manifest across non-dominant and dominant group members in different social domains leaves room for future investigations into possible moderators of these effects.

7. General discussion

Good-focal and bad-focal indirect measures revealed distinct components of social cognition. Across race, religion and sexual orientation, indirect measures that focused on positive valence showed effects consistent with social identity theories (Tajfel & Turner, 1979) and the totalitarian ego (Greenwald, 1980) - one's group is better than the others. When attitudes were measured indirectly with a focus on positive valence, members of non-dominant groups showed ingroup favoritism, often at levels comparable to dominant group members. These results are not typical observations in measures of implicit attitudes (e.g., Nosek et al., 2007; Rudman et al., 2002); they were made possible by a methodological innovation distinguishing a focus on positive valence from a focus on negative valence.

Results more compatible with previous findings emerged when measuring social cognition with indirect measures that focused on negative valence. Measures that focused on negative valence found effects partly consistent with system justification (Jost et al., 2004; Jost & Banaji, 1994) and cultural learning perspectives concerning the development of implicit attitudes (Dasgupta, 2013), with more favoritism towards the culturally dominant group. When focusing on negative valence, non-dominant group members showed either lower levels of ingroup favoritism, no reliable preference between their own and dominant groups, or outgroup favoritism. Simultaneous ingroup and outgroup favoritism in good and bad associations among social minorities indicates complexity in social cognition that has not been observed previously. This provides evidence that there are distinct components of implicit social evaluation for the same attitude objects (e.g., De Liver, van der Pligt, & Wigboldus, 2007; Zayas & Shoda, 2015).

7.1. The origins of the discrepancy

Why do indirect measures that focus on positive valence show results more compatible with the effect of group identity on group preferences, whereas indirect measures that focus on negative valence

show results more compatible with cultural standards? This pattern of results might reflect the influence of *distinct sources* of information, with each source affecting one measure to a larger extent than the other. We speculate that the positivity-focused indirect measures reflected the positive information both dominant and non-dominant group members derive from their ingroup identity because positivity-focused measures are more sensitive to positive information, as shown in Study S2. Conversely, negativity-focused measures reflected the negative information both dominant and non-dominant groups members learn to associate with the non-dominant groups as a result of cultural messages that emphasize the negative characteristics of such groups (e.g., Tukachinsky et al., 2015; Verkuyten & Thijs, 2002). Moreover, because dominant group identities are rarely identified explicitly (Hartmann et al., 2009; Pratto & Stewart, 2012), cultural preference for the dominant group over the non-dominant groups might be based mostly on negative information about the non-dominant groups rather than on positive information about the dominant group.

Consistent with this perspective, in Study 5 we found that cultural context (U.S. vs. Israel) had a larger effect on the negativity-focused indirect measure than on the positivity-focused indirect measure. Jewish participants showed less ingroup-favoritism on the negativity-focused measure in a Christian-dominant culture than in a Jewish-dominant culture. These data demonstrate more dramatic evidence of the impact of social environment and cultural values on implicit evaluations than prior work (e.g., Jost et al., 2004; Newheiser & Olson, 2012; Nosek et al., 2002).

The distinction between person and culture in implicit social cognition has engendered substantial theoretical innovation and debate (Banaji, 2001; Gawronski, Peters, & LeBel, 2008; Karpinski & Hilton, 2001; Nosek & Hansen, 2008a, 2008b; Olson & Fazio, 2004). Person and culture are clearly "mutually constitutive" (Fiske, Kitayama, Markus, & Nisbett, 1998) in that personal attitudes form within a culture, and cultural preferences emerge from the collective actions of individuals. Moreover, there is little doubt that the source of these good and bad associations is the individual's memory (Nosek & Hansen, 2008a). Nonetheless, it is appealing to consider whether the origins of these associations are distinguishable.

Similarly, more recent work has posited that implicit biases might reflect situational, rather than individual, psychological processes; that is, performance on indirect measures may tell us less about the individuals who provided the data and more about the contexts that the data came from (Payne, Vuletich, & Lundberg, 2017). Based on our

results, it is possible that the bad-focal measures in our research capture situational effects (e.g., media messages) better than good-focal measures, whereas good-focal measures are more sensitive to dispositional effects.

However, as suggested by the results of our attitude formation study (Study S2), the reason for our findings is potentially not that bad-focal measures are more sensitive to situational effects than good-focal measures. Rather, in the context of attitudes towards social groups, cultural messages (which happen to be situational factors) are the main factor that induces associations of non-dominant groups with negative valence, whereas positive self-regard is the main factor that induces associations between one's own group and positive valence.

7.2. Construct validation

At this point of discovery, there is only initial evidence that good-focal measures specifically assess associations with positive concepts and show greater sensitivity to positive information about attitude objects, while bad-focal measures specifically assess associations with negative concepts and show greater sensitivity to negative information. There are some reasons to believe that each measure is more related to positive or negative associations. The two measures have appealing face validity in focusing on only one of the two categories. Also, Sriram and Greenwald (2009) observed faster responses to stimuli assigned to the focal category, indicating that participants selectively attended to focal categories while completing the task. Our data provide additional evidence for convergent and discriminant validity for good-focal and bad-focal indirect measures. For instance, when measured together, the BIATs had an average correlation of $r = 0.30$, indicating that the two measures are related but distinct.

Some of our results are consistent with the notion that bad-focal measures assess the same construct as good-focal measures but simply suffer from lower construct validity (e.g., weaker correlations with explicit preferences). However, there is also evidence that good-focal and bad-focal measures assess distinct constructs, and performance on good-focal measures is just more related to self-endorsed attitudes. First, the measures showed comparable internal reliability (Median good-focal BIAT $\alpha = 0.81$, Median bad-focal BIAT $\alpha = 0.75$), and the online supplement shows the same conclusions are reached using Revelle's omega (McNeish, 2017). Second, Study 5 revealed a positive and reliable correlation between perceived cultural preferences and bad-focal measures, though a similar correlation was not present in Study 4. Finally, in analyses of racial and sexual associations, the bad-focal BIAT produced qualitatively different responses compared to a good-focal BIAT among participants from non-dominant groups (significant outgroup favoritism), an outcome inconsistent with a low construct validity argument. That is, an unreliable measure fails to capture the signal; it does not capture the reverse of the signal.

7.3. Limitations

These results represent a preliminary investigation of an intriguing asymmetry among non-dominant and dominant group members on indirect measures focused on positive versus negative valence. We hope this work generates further investigations into complexities of implicit social cognition, the impact of person versus culture on implicit attitudes, and seeks to address several of the limitations of these current studies.

For one, as evident in Figs. 3–5, our results were quite variable, and while our conclusions can highlight broad distinctions between indirect measures of positive versus negative valence, they cannot yet fully account for the observed variation across social dimensions. Future work should look to better explain the reasons behind this variability, particularly the variation in performance on bad-focal measures among participants from non-dominant groups and why, unlike their American counterparts, dominant group members from Israel in Study 5 showed

no difference between good and bad-focal measures.

Similarly, one notable inconsistency among dominant groups was that Christian participants had higher levels of ingroup favoritism on good-focal than bad-focal measures, whereas the opposite was true for White and straight participants. This discrepancy may speak to variation in the relation between good-focal and bad-focal measures among dominant group members regarding identities that are considered predetermined (like race) versus identities considered to be personally chosen (like religion). Follow-up work could examine whether similar reversals exist in other social identities that are viewed as more chosen (e.g., professions), or even explore whether effects differ among participants who believe that membership in the same social category is either predetermined versus self-selected (e.g., sexual orientation). Relatedly, this discrepancy may be explained by variations in the level of ingroup favoritism shown across social dimensions; it is notable that Christian participants in Study 2 showed much higher levels of implicit ingroup favoritism on good-focal measures than White and straight participants in Studies 1, 3 and 4 (Christian average $d = 1.15$, White average $d = 0.52$, Straight $d = 0.66$). Future research could likewise explore whether similar reversals arise in topics known to elicit greater implicit favoritism (e.g., preferences between one's own and other countries; Nosek, 2005).

In addition, we are missing clear evidence that bad-focal measures are directly impacted by cultural messages. Exposure to real-world cultural messages or values may be particularly difficult to measure given that they are unlikely to be open to introspection (Nosek & Hansen, 2008a, 2008b), and exposure to such existing cultural messages may be challenging to manipulate directly in a laboratory context given their pervasiveness in everyday life. We attempted to address this issue in Study 5 by "manipulating" culture through comparing American vs. Israeli Jews, but these data are ultimately correlational. Progress on this claim may come from more controlled studies using implicit evaluations of novel, artificial groups (e.g., Hussak & Cimpian, 2015), where cultural information could be manipulated directly.

Moreover, these studies lack clear evidence that good-focal measures are directly related to ingroup identity concerns or motivations. While the need for positive self-regard is widespread (Sedikides et al., 2003), and many people use their group identity as a source of positivity (Brewer, 1999), the current results do not causally show that changes in group identity motivation create changes on good-focal indirect measures. Manipulations of group identity motivations may be difficult to produce reliably with the groups used in this investigation, but one potential avenue could be through threats or affirmation to one's self-esteem (e.g., Correll & Park, 2005; Fein & Spencer, 1997). From this perspective, a self-esteem threat may increase group identity concerns and heighten ingroup favoritism shown on good-focal measures, while the experience of affirmation may decrease group identity concerns and dampen ingroup favoritism shown on good-focal measures. Such work will clarify the potential relationship between group identity motivations and good versus bad focal indirect measures.

Finally, this line of research may benefit from alternative statistical approaches. This project used only methodological variants of the IAT, the BIAT and MC-IAT, in an attempt to distinguish between constructs that may have been measured simultaneously in the IAT. Our analyses used the most common methods of analyzing BIAT and MC-IAT data (i.e., D scores), but it is possible that alternative methods for analyzing data from indirect measures may further decompose these constructs. For example, future work could apply Quad model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) analyses to see if the pattern of results found using D scores is likewise reflected in the association activation parameters produced by Quad model analyses.

8. Conclusion

Distinct theories articulate a need for positive self-regard and either a need to maintain the status quo or a general susceptibility to cultural

messages about group standing or value. The present results suggest that social evaluation may represent these occasionally conflicting motives, and that measurement innovation of assessing good and bad associations separately may help to detect them. The next step is to better understand how these two components of implicit evaluation are formed, and how they independently or interactively shape social judgment and behavior.

Open practices

This article earned the Open Materials and Open Data badges for transparent research practices. Materials and data for all experiments are available at https://osf.io/u7tkp/.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2018.08.007.

References

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The rules of implicit evaluation by race, religion, and age. *Psychological Science, 25*(9), 1804–1815.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. RoedigerIII, J. S. Nairne, I. E. Neath, & A. M. Surprenant (Eds.). *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Blodorn, A., O'Brien, L. T., Cheryan, S., & Vick, S. B. (2016). Understanding perceptions of racism in the aftermath of Hurricane Katrina: The roles of system and group justification. *Social Justice Research,* 1–20.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues, 55*(3), 429–444.

Brewer, M. B., & Gardner, W. (1996). Who is this "We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71*(1), 83–93.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The Quad model of implicit task performance. *Journal of Personality and Social Psychology, 89*(4), 469–487.

Correll, J., & Park, B. (2005). A model of the ingroup as a social resource. *Personality and Social Psychology Review, 9*(4), 341–359.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Dasgupta, N. (2013). Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and the self-concept. In P. G. Devine, & E. A. Plant (Vol. Eds.), *Advances in experimental social psychology. 47. Advances in experimental social psychology* (pp. 233–279). UK: Academic Press.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on automatic gender stereotyping. *Journal of Experimental Social Psychology, 40,* 642–658.

De Liver, Y., van der Pligt, J., & Wigboldus, D. (2007). Positive and negative associations underlying ambivalent attitudes. *Journal of Experimental Social Psychology, 43*(2), 319–326.

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology, 73*(1), 31–44.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron, 1,* 3–32.

Fiske, A., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Vol. Eds.), *The handbook of social psychology*(4th ed.). *vol. 2. The handbook of social psychology* (pp. 915–981). San Francisco: McGraw-Hill.

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or extra-personal? Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass, 2*(2), 1002–1023.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2014). Measures of implicit attitudes may conceal differences in implicit associations: The case of antiaging bias. *Social Psychological and Personality Science, 5*(3), 271–278.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35*(7), 603–618.

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,* 1464–1480.

Greenwald, A. G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist, 69*(7), 669–684.

Hartmann, D., Gerteis, J., & Croll, P. R. (2009). An empirical assessment of whiteness theory: Hidden from how many? *Social Problems, 56*(3), 403–424.

Hussak, L. J., & Cimpian, A. (2015). An early-emerging explanatory heuristic promotes support for the status quo. *Journal of Personality and Social Psychology, 109*(5), 739–752.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33*(1), 1–27.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25*(6), 881–919.

Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology, 38*(6), 586–602.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*(5), 774–788.

Levin, S., & Sidanius, J. (1999). Social dominance and social identity in the United States and Israel: Ingroup favoritism or outgroup derogation? *Political Psychology, 20*(1), 99–126.

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods.* https://doi.org/10.1037/met0000144 (Advance online publication).

Newheiser, A. K., & Olson, K. R. (2012). White and Black American children's implicit intergroup bias. *Journal of Experimental Social Psychology, 48*(1), 264–270.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134,* 565–584.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*(1), 101–115.

Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J. R., & Greenwald, A. G. (2014). Understanding and using the Brief Implicit Association Test: Recommended scoring procedures. *PLoS One, 9*(12), e110938.

Nosek, B. A., & Hansen, J. J. (2008a). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion, 22,* 553–594.

Nosek, B. A., & Hansen, J. J. (2008b). Personalizing the Implicit Association Test increases explicit evaluation of the target concepts. *European Journal of Psychological Assessment, 25,* 226–236.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., ... Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*(1), 36–88.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*(5), 653–667.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry, 28*(4), 233–248.

Pratto, F., & Stewart, A. L. (2012). Group dominance and the half-blindness of privilege. *Journal of Social Issues, 68*(1), 28–45.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology. General, 133*(2), 139–165.

Rudman, L. A., Feinberg, J., & Fairchild, K. (2002). Minority members' implicit attitudes: Automatic ingroup bias as a function of group status. *Social Cognition, 20*(4), 294–320.

Schwartz, M. B., Vartanian, L. R., Nosek, B. A., & Brownell, K. D. (2006). The influence of one's own body weight on implicit and explicit anti-fat bias. *Obesity, 14*(3), 440–447.

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology, 84*(1), 60–79.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression.* Cambridge, UK: Cambridge University Press.

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology, 56*(4), 283–294.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 8*(2), 245–251.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations, 33,* 33–47.

Tukachinsky, R., Mastro, D., & Yarchi, M. (2015). Documenting portrayals of race/ethnicity on primetime television over a 20-year span and their association with national-level racial/ethnic attitudes. *Journal of Social Issues, 71*(1), 17–38.

Uhlmann, E., Dasgupta, N., Elgueta, A., Greenwald, A. G., & Swanson, J. (2002). Subgroup prejudice based on skin color among Hispanics in the United States and Latin America. *Social Cognition, 20*(3), 198–226.

Verkuyten, M., & Thijs, J. (2002). Racist victimization among children in the Netherlands: The effect of ethnic group and school. *Ethnic and Racial Studies, 25*(2), 310–331.

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.