# HAPPY = HUMAN: A FEELING OF BELONGING MODULATES THE "EXPRESSION-TO-MIND" EFFECT

Toshiki Saito
*Waseda University*
*Japan Society for the Promotion of Science*

Steven M. Almaraz
*East Los Angeles College*

Kurt Hugenberg
*Indiana University Bloomington*

Past research has demonstrated a link between facial expressions and mind perception, yet why expressions, especially happy expressions, influence mind attribution remains unclear. Conducting four studies, we addressed this issue. In Study 1, we investigated whether the valence or behavioral intention (i.e., approach or avoidance) implied by different emotions affected the minds ascribed to expressers. Happy (positive valence and approach intention) targets were ascribed more sophisticated minds than were targets displaying neutral, angry (negative-approach), or fearful (negative-avoidance) expressions, suggesting emotional valence was relevant to mind attribution but apparent behavioral intentions were not. We replicated this effect using both Black and White targets (Study 2) and another face database (Study 3). In Study 4, we conducted path analyses to examine attractiveness and expectations of social acceptance as potential mediators of the effect. Our findings suggest that signals of social acceptance are crucial to the effect emotional expressions have on mind perception.

*Keywords:* emotion, mind perception, facial expression, social acceptance, social cognition

Whereas humans have firsthand experience with their own cognitions and emotions, we must infer others' mental faculties and experiences. The extent to which we ascribe to others sophisticated versus simplistic mental faculties varies from perceiver to perceiver (Waytz, Cacioppo, & Epley, 2010), from target to target (Deska, Lloyd, & Hugenberg, 2018; Gray, Gray, & Wegner, 2007), and from situation to situation (Powers, Worsham, Freeman, Wheatley, & Heatherton, 2014). Understanding when and why people ascribe to others less than fully human minds is a critical part of social cognition because mind ascription affects numerous meaningful outcomes. For example, when people perceive others as fully human—capable of behaviors and experiences that differentiate humans from animals and machines (e.g., secondary emotions)—they feel more empathy and express greater intentions to help these individuals when in need (Cuddy, Rock, & Norton, 2007; see also Gray, Young, & Waytz, 2012). On the other hand, perceiving others as less than human has been linked to outgroup derogation, discrimination, and harmdoing (Kteily, Bruneau, Waytz, & Cotterill, 2015; Vaes, Paladino, Castelli, Leyenz, & Giovanazzi, 2003).

More recently, scholars have also begun investigating how the perception of even basic emotions, via others' facial expressions, can shape our understanding of others' mental faculties. In the first demonstration of this "expressions-to-mind" link, Bowling and Banissy (2017) had participants view a series of faces morphed between inanimate and animate (i.e., a doll-to-human morph continuum) and had participants decide where along the continuum the target seemed animate. Critically, targets were manipulated to appear either smiling or with neutral expressions. Participants perceived animacy (i.e., having a mind) earlier in the doll-to-human morph continuum for smiling faces. Faces needed less objective signal of humanness to be perceived as animate when smiling. In other words, happy faces seem more human.

However, exactly why happy faces seem more humanlike is unclear. One possibility is that this is a domain-general expression-to-mind effect. Thus, any expression may indicate that a target has an inner life, enhancing animacy. Casting doubt on this possibility, Krumhuber, Lai, Rosin, and Hugenberg (2019) recently found that faces displaying disgust were attributed less sophisticated minds than happy and neutral faces. However here, too, the results are subject to interpretation. Perhaps these results are due to the *approach versus avoidance signals* of the expressions. Some expressions, such as happiness and anger, indicate that a person is likely to approach the self, whereas other expressions, such as fear or disgust, suggest they are likely to avoid the self (Adams, Ambady, Macrae, & Kleck, 2006). Importantly, past work also finds that targets believed to be approaching the self are seen as more mentally sophisticated than those avoiding the self, arguably because perceivers upregulate mind perception in preparation for social interaction (Khalid, Deska, & Hugenberg, 2016). If true, perhaps any expression that implies approach regardless of valence, such as both happiness and anger, may upregulate mind perception and create the perception of mental sophistication in others.

Alternatively, perhaps happy expressions are specifically humanizing because *happy expressions signal social inclusion*, which itself influences mind ascription.

Whereas happiness is a signal of social inclusion and acceptance (DeWall, Maner, & Rouby, 2009), this is not necessarily true of other approach-oriented expressions, such as anger, which signals social rejection (to the point of potential interpersonal conflict). Critically, past work has found that motives for social inclusion are important in mind perception as well. For example, Powers and colleagues (2014) found that the motivation to socially connect influences ascriptions of animacy. Again using doll-to-human morph continua, lonely individuals lower their thresholds for animacy, observing targets as having humanlike inner lives with objectively lower levels of humanness. Extending this logic, perhaps happy expressions signal impending social inclusion, and thereby are seen as more animate and mindful.

Finally, the specific happy-to-human effect observed by both Bowling and Banissy (2017) and Krumhuber and colleagues (2019) may be driven by *perceptions of facial attractiveness*. Past work has reliably found that smiling faces are more attractive than their frowning counterparts (Mueser, Grau, Sussman, & Rosen, 1984; Reis et al., 1990), and that attractive faces are often ascribed more humanlike sophistication than their less attractive counterparts (Alaei, Deska, Hugenberg, & Rule, 2021). Thus, perhaps the happy-to-human link is mediated by expression-driven changes in physical attractiveness.

Notably, his past research has relied extensively on human-to-non-human morph continua. This has been quite useful in clarifying the point at which morphed stimuli psychologically transition from animate to inanimate. For example, Bowling and Banissy (2017) employ human to doll morph continua. Similarly, Krumhuber and colleagues (2019) employ human-to-cartoon continua. However, it is unclear whether similar effects replicate with fully human stimuli, which is an important advance if we are to understand whether others' expressions affect mind ascription in vivo.

## THE PRESENT WORK

The present work had two goals. Our primary goal was to investigate the effects of emotional expression on mind attribution, and to elucidate the underlying mechanisms. Our secondary goal was to replicate past work showing effects of expressions on mind ascription with fully human images. We conducted four experiments to pursue these goals. In Studies 1–3, participants completed the Mind Ascription Scale (Kozak, Marsh, & Wegner, 2006), responding to images of human faces showing anger, happy, fear, and neutral expressions. This tested the approach versus avoidance hypothesis. These tests showed that happy expressions, but not approach expressions generally (i.e., anger expressions), facilitated the ascription of sophisticated humanlike mental faculties. Studies 2 and 3 replicated this basic effect across target race and gender. After establishing that happiness has a unique effect on perceptions of mental faculties in Studies 1–3, Study 4 investigated the psychological mechanisms by which emotional expression influences mind attribution. Specifically, Study 4 showed that the happy-to-human effect is mediated by beliefs that happy expressions signal acceptance, but not mediated by facial expression effects on physical attractiveness.

## STUDIES 1, 2, AND 3

Because the procedures and findings of Studies 1–3 are nearly identical, we present them together. Study 1 was designed to test the approach versus avoidance explanation of the happy-to-human effect. Participants viewed a series of White male faces taken from the Chicago Face Database (CFD; Ma, Correll, & Wittenbrink, 2015), displaying happiness, anger, fear, and neutral expressions. Participants rated each of the individuals on Kozak, Marsh, and Wegner's (2006) Mind Attribution Scale, a well-validated measure of mind perception. We predicted that happy targets would be ascribed more sophisticated humanlike minds relative to neutral expressions, replicating past work with human-to-inanimate morph continua. Of primary interest, however, was what would occur with angry expressions. If the previous happy-to-human link was caused by happy targets signaling approach, and approach signals lead perceivers to upregulate mind perception (Khalid et al., 2016), then angry expressions should similarly elicit mental sophistication. However, if the happy-to-human effect is caused instead by happiness signaling social acceptance (Powers et al., 2014) or by happiness improving physical attractiveness (Reis et al., 1990), targets displaying angry expressions should not be ascribed high levels of mental sophistication because they lack both signals of social acceptance and facial attractiveness.

Study 2 was an exact replication of Study 1, except it employed both White male and Black male faces taken from the CFD. Study 3 was an exact replication of Study 1, except that it employed White male and White female faces from the Pictures of Facial Affect database (Ekman, 1993). Because race and gender can affect mind ascription (e.g., Alaei et al., 2021; Cassidy et al., 2017; Goff, Eberhardt, Williams, and Jackson, 2008), we included these as additional factors in Studies 2 and 3 to confirm the generalizability of the emotion effect on mind ascription.

### METHODS

*Participants.* To calculate sample size for Study 1, we used Bowling and Banissy's (2017) effect size of $\eta_p^2 = .12$ for the main effect of emotion expression (i.e., happy vs. neutral) on animacy. A G*Power (Version 3.1.9.3) power analysis suggested 18 participants to achieve 95% power. We conservatively recruited 40 participants from Amazon's Mechanical Turk. Based on the effect size observed in the first study, we recalculated sample sizes for the subsequent two studies, recruiting 50 participants in Studies 2 and 3. All participants provided informed consent before participation. All data were retained in all studies, and all measures and manipulations are disclosed.

*Sensitivity Power Analysis.* Because we used multilevel generalized linear mixed model (GLMM) analyses, for which formulas of power analysis are not available (Eitan et al., 2018), we applied a simulation approach for post-hoc sensitivity power analysis by using the *powerSim* function in the *simr* package (Green & MacLeod, 2016). Assuming 80% power and an alpha of 0.05, the sensitivity power analyses

suggest that each sample size (Study 1: 40; Study 2: 50; Study 3: 50) would detect the minimum effects, $\beta = 0.17$ (Study 1), $\beta = 0.26$ (Study 2), and $\beta = 0.22$ (Study 3) for the key analyses.

*Stimuli.* In Study 1, we used 20 White male targets from the CFD (Ma, Correll, & Wittenbrink, 2015) expressing four emotions: anger (negative and approaching), happy (positive and approaching), fear (negative and distancing), and neutral, for a total of 80 stimuli. In Study 2, we added 20 Black male targets from the CFD, for a total of 160 stimuli (80 Black; 80 White). In Studies 1 and 2, we used only male faces because extensive past work has linked facial expressions to gender (e.g., Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007; Hess, Blairy, & Kleck, 2000). The attractiveness of the targets in Studies 1 and 2 was controlled using the subjective attractiveness ratings from the face database norming data (Ma et al., 2015). Specifically, we selected targets that were not perceived as especially attractive (more than 4 in a 1–ert scale, 1 = Not at all attractive; 7 = Extremely attractive) and unattractive faces (less than 2.5 on the same scale). Then, the attractiveness of the chosen stimulus faces ranged 2.70 to 3.85 ($M = 3.13$, $SD = 0.61$). Importantly, the attractiveness of White and Black targets was not significantly different, $t(38) = 0.01$, $p = .99$. In Study 3, we used four White male targets and four White female targets from the Pictures of Facial Affect database displaying the same emotions as Studies 1 and 2 (32 total stimuli; Ekman, 1993).[1] This provided a replication with both male and female stimuli, and provided an additional stimulus set with extensively validated expressions. Regarding the targets in Study 3, we did not attempt to control attractiveness because the number of stimuli was very limited; there were 11 targets who expressed four emotions (six females and five males). To balance the gender of the stimuli, we chose eight targets (four from each gender) for stimulus faces in Study 3. The stimuli were resized to 770 × 541 pixels (Studies 1 and 2) and to 384 × 570 pixels (Study 3) and placed onto a white background.

*Procedure.* In Study 1, participants were informed they would rate individuals on their personalities and abilities at zero acquaintance. Participants then saw 20 White male faces, varying within-subjects in emotional expressions (happy, anger, fear, neutral; five faces per expression). The identity-emotional expression pairing for each face was counterbalanced across participants, such that participants were equally likely to see a given identity displaying happiness, fear, anger, or a neutral expression, but no participant saw the same target with multiple expressions. Participants rated each face using the Mind Attribution Scale (Kozak et al., 2006), our measure of mind ascription. Each face was displayed until the participant completed all ratings.

The Mind Attribution Scale is a 10-item scale in which participants rate targets on their emotion (e.g., "This person has complex feelings"), intention (e.g., "This person is capable of doing things on purpose"), and cognition (e.g., "This person can

---

1. Stimulus codes for the exact stimuli employed are available in the online supplemental material (https://osf.io/93e25/?view_only=0e48caf17ae14ffe96b8cd18bef839a4).

engage in a great deal of thought") on 7-point Likert scales (1 = strongly disagree; 7 = strongly agree). Higher scores indicate the ascription of more sophisticated minds. Reliability for the summed items was high across Studies 1–4 (Cronbach's α = .98, .97, .95, .98). Faces were presented and rated one at a time in a random order. Finally, participants provided demographic information, and were thanked and debriefed.

The procedures for Studies 2 and 3 were identical, except as noted. In Study 2, participants rated 20 White and 20 Black male faces varying in emotional expressions (happy, anger, fear, neutral; 5 each per race-expression paring). In Study 3, participants rated 4 female and 4 male faces (happy, anger, fear, neutral; 2 faces per emotion condition).

*Analysis Plan.* Data were analyzed using GLMM the *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) packages in R software (R core Team, 2019). The models predicted ratings of the Mind Attribution Scale, with a fixed effect for emotional expressions (four levels: anger, happy, fear, neutral) and random intercepts for both stimuli and participants. In Study 2, the model also included both a fixed effect for race (two levels: Black, White) and an interaction between race and emotional expressions in addition to the fixed effect for emotional expressions. Furthermore, we conducted other GLMM analyses which included three covariates (i.e., the age, gender, and ethnicity of participants) in Studies 1–3 because each of these could influence the perception of facial expressions. Finally, all effects replicate closely using traditional repeated-measures ANOVAs (see Supplemental Figure S1 and Tables S1–3).

## RESULTS: STUDY 1

Consistent with past work, happy faces ($M$ = 5.31, $SD$ = 1.22) were attributed more sophisticated minds than neutral faces were ($M$ = 4.97, $SD$ = 1.32), $B$ = 0.31, $SE$ = 0.06, 95% CI [0.19, 0.42], $t(72.87)$ = 5.01, $p < .001$. However, neither angry faces ($M$ = 4.86, $SD$ = 1.30), $B$ = 0.06, $SE$ = 0.06, 95% CI [−0.06, 0.18], $t(72.87)$ = 0.97, $p$ = .34, nor fearful faces ($M$ = 4.81, $SD$ = 1.38), $B$ = 0.01, $SE$ = 0.06, 95% CI [−0.11, 0.12], $t(72.87)$ = 0.08, $p$ = .93, significantly differed from neutral expressions on mind attribution (see Figure 1). These results were unchanged when including covariates (Table S4 in the Supplementary Materials).

Because angry faces are a negative-approach emotion (Adams et al., 2006), they are of particular interest in an approach versus avoidance explanation of this expression effect. Thus, we compared angry faces to both happy and fearful faces on mind ascription. Notably, angry faces were judged as having significantly less sophisticated minds than happy faces were, $B$ = −0.25, $SE$ = 0.06, 95% CI [−0.37, −0.13], $t(72.87)$ = −4.04, $p < .001$, but did not significantly differ from fearful faces, $B$ = 0.05, $SE$ = 0.06, 95% CI [−0.06, 0.17], $t(72.87)$ = 0.88, $p$ = .38. This indicated that not all approaching expressions upregulate mind ascription equally.
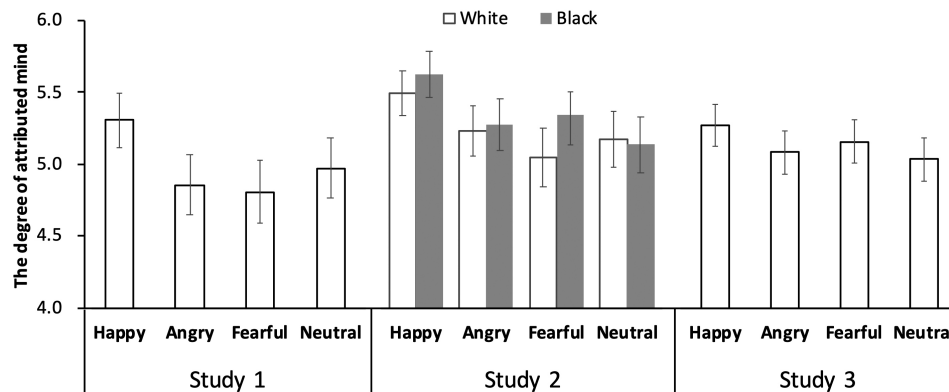
FIGURE 1. Mean attributed mind toward an expressor of each emotion in Studies 1 through 3. Happy faces were attributed more mind than other expressions. Error bars represent standard errors.

## RESULTS: STUDY 2

In Study 2, happy faces ($M = 5.52$, $SD = 1.12$) were again attributed more sophisticated minds than neutral faces were ($M = 5.13$, $SD = 1.33$), $B = 0.29$, $SE = 0.08$, 95% CI [0.14, 0.45], $t(150.71) = 3.65$, $p < .001$. Here again, neither angry ($M = 5.23$, $SD = 1.22$), $B = 0.02$, $SE = 0.08$, 95% CI [−0.13, 0.18], $t(150.71) = 0.28$, $p = .78$, nor fearful ($M = 5.17$, $SD = 1.29$) faces differed from neutral expressions, $B = −0.13$, $SE = 0.08$, 95% CI [−0.29, 0.02], $t(150.71) = −1.67$, $p = .10$. These results were unchanged when including covariates (Table S5 in the Supplementary Materials).

Angry faces again differed significantly from happy faces, $B = −0.27$, $SE = 0.08$, 95% CI [−0.43, −0.12], $t(150.71) = −3.38$, $p < .001$, but not from fearful faces, $B = 0.16$, $SE = 0.08$, 95% CI [0.00, 0.31], $t(150.71) = 1.94$, $p = .054$, replicating the unique effect of happy expressions.

Notably, these differences occurred for both Black and White target expressions. The only significant difference across race was an unexpected target race difference in the fearful expression condition, $B = 0.32$, $SE = 0.11$, 95% CI [0.10, 0.54], $t(150.75) = 2.81$, $p = .006$, such that Black fearful faces were attributed more mental sophistication ($M = 5.32$, $SD = 1.14$) than White fearful faces ($M = 5.02$, $SD = 1.42$).

## RESULTS: STUDY 3

Replicating the previous studies, Study 3 found that happy faces ($M = 5.34$, $SD = 1.08$) were attributed more mental sophistication than neutral faces were ($M = 5.09$, $SD = 1.11$), $B = 0.26$, $SE = 0.10$, 95% CI [0.08, 0.44], $t(24.77) = 2.71$, $p = .012$, whereas neither angry faces ($M = 5.14$, $SD = 1.11$), $B = 0.06$, $SE = 0.10$, 95% CI [−0.12, 0.24], $t(24.77) = 0.61$, $p = .55$, nor fearful faces ($M = 5.22$, $SD = 1.11$) were

significantly different from neutral expressions, $B = 0.13$, $SE = 0.10$, 95% CI [−0.05, 0.31], $t(24.77) = 1.38$, $p = .18$. These results were again unchanged when including covariates (Table S6 in Supplementary Materials).

Finally, angry faces differed significantly from happy faces, $B = −0.20$, $SE = 0.10$, 95% CI [−0.38, −0.02], $t(150.71) = −3.38$, $p < .001$, but not from fearful faces, $B = −0.07$, $SE = 0.10$, 95% CI [−0.26, 0.11], $t(24.77) = −0.77$, $p = .45$.

In addition to the above analysis, we included the gender of the stimulus face as a factor in the model given that previous work has identified that the gender of stimulus face also influences mind attribution. However, inconsistent with the previous studies, we did not find an effect of target gender on mind, $B = −0.02$, $SE = 0.14$, 95% CI [−0.11, 0.12], $t(20.90) = 0.18$, $p = .86$. Furthermore, there was no interaction between target gender and emotional expressions, nor did we find significant simple effects of stimulus gender on mind attribution across angry, happy, and fearful faces: $B$s = −0.13, 0.04, −0.05, $SE$s = 0.20, 0.20, 0.20, 95% CIs [−0.48, 0.23], [−0.32, 0.39], [−0.41, 0.30], $t$s(20.90) = −0.63, 0.18, −0.27, $p$s = .54, .86, .79.

## DISCUSSION

Across Studies 1–3, we replicated the happy-to-human link observed in past work (Bowling & Banissy, 2017; Krumhuber et al., ____), but extended this in three important ways. First, we demonstrated that the ____ results replicate with real human faces. Thus, this is not merely an effect of where in a human-to-inanimate morph continuum faces transition from animate to inanimate. Instead, even among fully human faces, the expression of happiness in particular leads targets to seem more mentally sophisticated. Second, this happy-to-human link appears to occur generally across target gender and race. Notably, we failed to observe a main effect of race or gender on mind ascription, as has been reported in related work on dehumanization (e.g., Alaei et al., 2021; Goff et al., 2008). However, this result might reflect social desirability because participants explicitly answered the questionnaire in the current study. Third, and most importantly, we also found that this happy-to-human effect does not extend to the approach-oriented emotion of anger. Specifically, we observed broadly consistent results that happy faces were attributed more mental sophistication than neutral expressions, an effect not observed for anger.

This is important because it indicates that the happy-to-human link is not easily explained by the approach orientation of the face. If the approach orientation of the expressions caused the perception of mental sophistication, we would expect targets displaying angry (negative/approach) expressions to be ascribed more sophisticated minds than targets displaying fearful (negative/avoidance) expressions. In actuality, we found consistent evidence that only happy expressions increase mind ascription.

Thus, whereas past research suggested that others' approach orientation toward the self may be an important cue in upregulating mind perception (Khalid et al., 2016), our findings are inconsistent with this explanation. Instead, the present results suggest either a social inclusion signaling or an attractiveness explanation. Study 4 was designed to competitively test these two hypotheses.

## STUDY 4

In Study 4, we investigated whether the *signal of social acceptance* or the *increased attractiveness* created by happy faces mediates the previously established happy-to-human effect. Both hypotheses seemed plausible a priori. In support of the social acceptance hypothesis, past work has shown both that happy expressions signal acceptance (DeWall et al., 2009) and that acceptance motives can enhance perceptions of faces' humanness (Powers et al., 2014). In support of the attractiveness hypothesis, past work has shown both that smiling faces are more attractive (Reis et al., 1990) and that attractive faces are ascribed more humanlike sophistication (Alaei et al., 2021).

In Study 4, participants completed the same procedure as in Study 1, with two exceptions: Participants also rated both the attractiveness of each target and whether each target seemed likely to socially accept the self. Of interest was whether structural equation modeling (SEM) would reveal that attractiveness and/or social acceptance mediate the happy-to-human effect. Given that both target attractiveness and facial signals of belongingness are typically quickly processed, affectively based experiences and deliberative mental sophistication judgments of faces have been shown to be influenced by such cues (e.g., Deska et al., 2018), we believe using facial attractiveness and social acceptance signals as mediators of the expression-to-humanness effect allows us to test the most plausible causal chains.

METHODS

*Participants.* Study 4 required a larger sample size than the previous studies because we sought to conduct SEM analysis. Per Jackson (2003), an *N* (number of cases):*q* (number of parameters estimated) ratio of 10:1 is considered adequate. In our hypothesized model, there were eight estimated parameters, suggesting a minimum of 80 participants. We conservatively targeted approximately 120 participants. In actuality, 117 participants completed this study through MTurk. All participants provided written informed consent to participate in this study before their participation.

*Procedure.* The procedure and stimuli were identical to Study 1, except as follows. Participants rated each target for attractiveness ("This person is attractive"), and the degree of social acceptance of the self ("This person would accept you") using a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree) before rating targets using the Mind Attribution Scale. Similar single-item measures have been employed in previous research measuring facial attractiveness (e.g., Ma et al., 2015) and perceptions of others' motives from faces (e.g., Khalid et al., 2016). Finally, participants also completed the Need to Belong (NTB) scale (Leary, Kelly, Cottrell, & Schreindorfer, 2013) for exploratory purposes. This scale did not yield meaningful results (see Figure S2 in supplementary material) and is not discussed further.

*Statistical Analysis.* We used the *sem* function of the *Lavaan* package (Rosseel, 2012) in R to analyze the data. We established a hypothesized model consisting of an independent variable (emotional expression) and two parallel mediators (attractiveness and acceptance ratings for each stimulus by each participant). The dependent variable was the Mind Ascription Scale. Then, we compared the fit of multiple alternative models to our hypothesized model. We also tested two alternative models consisting of two indirect effects identical to the hypothesized model except the pathways of variables were in serial, rather than in parallel. In the first alternative model, the paths were specified sequentially from emotional expression to attractiveness, from attractiveness to acceptance, and acceptance to humanness. In the second alternative model, the paths were specified sequentially from emotional expression to acceptance, from acceptance to attractiveness, and from attractiveness to humanness.

We employed several fit indexes to test model fit, including chi-square test, comparative fit index (CFI), Tukey–Lewis index (TLI), and root mean square residual (RMSEA). We also used Akaike information criterion (AIC) and Bayesian information criterion (BIC) for choosing the best-fitting model. We calculated the 95% confidence interval of the indirect effects based on a Monte Carlo test of mediation using the *moteCarloMed* function of *semTools* (Jorgensen et al., 2018) in R.

RESULTS AND DISCUSSION

We tested the effects of acceptance and attractiveness on mind attribution via SEM (see Figure 2). The hypothesized model provided a good fit to the data (CFI = 1.00, TLI = 1.00, RMSEA = 0.00, AIC = 2203.71, BIC = 2234.81). Although happy faces were rated as more attractive than other expressions ($\beta$ = 0.507, *p* = .001, 95% CI [0.195, 0.839]), the effect of happy faces on mind attribution was not mediated by attractiveness ($\beta$ = 0.041, *p* = .239, 95% CI [−0.023, 0.121]). However, acceptance did mediate the effect of happy faces on mind attribution ($\beta$ = 0.231, *p* < .001, 95% CI [0.102, 0.387]). We also compared the fitting of the hypothesized model to the alternative models. As seen in Table 1, our hypothesized model showed the best fit, compared to the alternative models.

These results provide preliminary evidence as to why happy faces are more readily ascribed sophisticated minds, supporting the *signal of social acceptance* but not the *increased attractiveness* hypothesis. Put simply, people may attribute more sophisticated minds to happy faces because those observing such faces anticipate acceptance. This is conceptually consistent with past research linking the perception of animacy to social motives (Powers et al., 2014).

**GENERAL DISCUSSION**

The present work competitively tested three possible explanations for the previously established "happiness-to-human" link (Bowling & Banissy, 2017; Krumhuber et al., 2019). Studies 1–3 found that happy expressions, but neither anger nor fear expressions, led to the ascription of more sophisticated minds. Critically, this
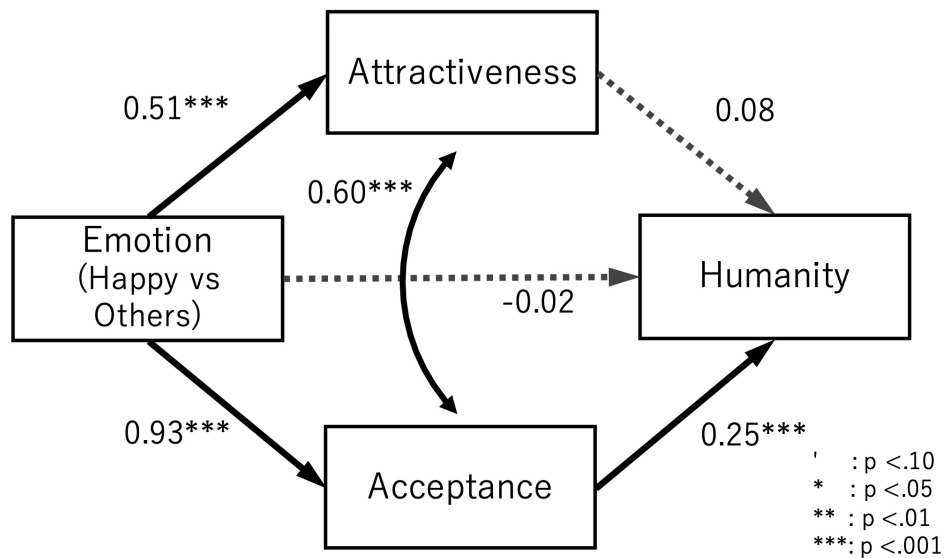
FIGURE 2. Results of structural equation model analysis in Study 4. Solid paths represent significant effects. Dashed lines represent nonsignificant effects.

rules strongly against the happiness-to-human effect as purely the product of an approach-oriented signal. Although past work has implicated approach signals in mind ascription (Khalid et al., 2016), given that anger is a clear approach-oriented emotion but does not lead to an upregulation of mind ascription, our findings are inconsistent with this explanation. Study 4 competitively tested both a happiness-as-signaling-acceptance hypothesis and a happiness-driven attractiveness hypothesis. The results showed that expectations of acceptance, but not happiness-driven attractiveness, had a significant indirect effect on mind attribution.

These studies build on past work in important ways. First, as previously noted, the present work extends the past findings showing that happiness affects animacy judgments in animate-to-inanimate morph continua to fully human faces. This has important implications for how perceiving others' facial affect feeds downstream into important social inferences. Indeed, the perception that others have sophisticated versus simplistic mental states (i.e., who is seen as having fully human mental faculties) has implications for meaningful outcomes ranging from who is helped when in need (Cuddy et al., 2007), to who is chosen for a job (Sim, Almaraz, & Hugenberg, 2021), to responses about who is saved or sacrificed in an emergency (Alaei et al., 2021). That facial affect, and happiness specifically, may play a key role in such judgments is an important advance.

Further, the present work also extends past work demonstrating that social acceptance motives play a key role in judging others' inner states. As noted above, lonely individuals are more likely than socially connected individuals to perceive ambiguous morphed faces as animate (Powers et al., 2014). Recent research has

**TABLE 1. Fit Statistics for the Hypothesized and Alternative Models**

| Model | $\chi^2$ | CFI | TLI | RMSEA | AIC | BIC |
|---|---|---|---|---|---|---|
| Hypothsized model | $\chi^2(6) = 106.966$ *** | 1.000 | 1.000 | 0.000 | 2203.710 | 2234.808 |
| Alternative model 1 | $\chi^2(6) = 106.966$ *** | 0.848 | 0.545 | 0.181 | 2217.008 | 2241.195 |
| Alternative model 2 | $\chi^2(6) = 106.966$ *** | 0.803 | 0.410 | 0.206 | 2221.570 | 2245.757 |

***$p < .001$.

extended similar effects to the anthropomorphism of nonhuman agents as well (e.g., Shin & Kim, 2018), finding that lonelier individuals are more likely to impute sophisticated minds into nonhuman agents (e.g., animals, computers), relative to socially connected individuals. The present work finds conceptually similar effects in a very different experimental context. Here, the stimuli were faces of *actual humans*, rather than morphed dolls or computers, and the social acceptance signaled by targets' happy expressions led to increased perceptions of sophisticated minds. We believe these conceptually similar results across experimental contexts speak to the potency of social acceptance motives in mind perception. Further, using real faces is a boon to the ecological validity of the present research. Whereas research using morphed faces can be a strong test of theory, the present research employs stimuli that people may actually encounter in vivo, and who may actually benefit (or suffer) from being seen as sophisticated (or simplistic).

Our finding that angry expressions did not lead to ascriptions of more sophisticated minds is important as well. Past research has suggested that others' approach orientation toward the self (i.e., direct eye gaze) was an important cue in upregulating mind perception (Khalid et al., 2016). However, in this research, we found that approach-oriented signals do not always upregulate mind ascription. Although angry faces signal approach, they also signal hostility toward the self and activate avoidance tendencies in observers (e.g., Seidel et al., 2010). Thus, the signal of hostility might disrupt mind ascription by the signal of approach orientation because it is adaptive for observers to avoid interaction with hostile targets entirely. This is quite distinct from the acceptance signaling approach of happy expressions. Further studies will be needed to examine what facial signals are prioritized during mind ascription and how approach and affiliative motives jointly or distinctly trigger mind ascription.

Finally, the limitations of the current work may provide fruitful avenues for future research. First, although we observed differences across target race for fearful expressions in Study 2, we did not predict this pattern. If this finding is replicable, further elaboration will be needed to understand why fearful Black faces were perceived as having more sophisticated minds than fearful White faces were. Second, although negative basic emotions (i.e., fear, anger) did not increase mind ascription relative to neutral expressions, there is a possibility that negative complex or secondary emotion expressions, such as embarrassment, may have

different effects, especially given that secondary emotions are considered uniquely human. Thus, further study investigating various expressions, including secondary emotions, will expand our understanding of how facial expressions influence mind attribution.

## CONCLUSION

In the current work, we provide a test of three potential explanations for why happy faces are judged as harboring sophisticated minds. We did not find evidence that approach intentionality or attractiveness effects of happy faces lead to the upregulation of mind perception. Instead, our findings suggest happy faces are perceived as more mentally sophisticated in part because they signal social acceptance to perceivers.

## REFERENCES

Adams, R. B., Ambady, N., Macrae, C. N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation and Emotion*, *30*(2), 177–186.

Alaei, R., Deska, J. C., Hugenberg, K., & Rule, N. O. (2021). People attribute humanness to men and women differently based on their facial appearance. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org.10.1037/pspi0000364

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Bowling, N. C., & Banissy, M. J. (2017). Emotion expression modulates perception of animacy from faces. *Journal of Experimental Social Psychology*, *71*, 83–95.

Cassidy, B. S., Krendl, A. C., Stanko, K. A., Rydell, R. J., Young, S. G., & Hugenberg, K. (2017). Configural face processing impacts race disparities in humanization and trust. *Journal of Experimental Social Psychology*, *73*, 111–124.

Cuddy, A. J. C., Rock, M. S., & Norton, M. I. (2007). Aid in the aftermath of Hurricane Katrina: Inferences of secondary emotions and intergroup helping. *Group Processes & Intergroup Relations: GPIR*, *10*(1), 107–118.

Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2018). Facing humanness: Facial width-to-height ratio predicts ascriptions of humanity. *Journal of Personality and Social Psychology*, *114*(1), 75–94. https://doi.org/10.1037/pspi0000110

DeWall, C. N., Maner, J. K., & Rouby, D. A. (2009). Social exclusion and early-stage interpersonal perception: Selective attention to signs of acceptance. *Journal of Personality and Social Psychology*, *96*(4), 729.

Eitan, O., Viganola, D., Inbar, Y., Dreber, A., Johannesson, M., Pfeiffer, T., . . . Uhlmann, E. L. (2018). Is research in social psychology politically biased? Systematic empirical tests and a forecasting survey to address the controversy. *Journal of Experimental Social Psychology*, *79*, 188–199.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384–392.

Goff, P. B., Eberhardt, J. L., Williams, M. J., & Jackson, M C. (2008). Not yet human: Implicit knowledge, historical dehumanization, and contemporary consequences. *Journal of Personality and Social Psychology, 94*(2), 292–306.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124. https://doi.org/10.1080/1047840X.2012 .651387

Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10 .1111/2041-210X.12504

Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior*, *24*(4), 265–283.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling*, *10*(1), 128–141.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., Rosseel, Y., Miller, P., Quick, C., & Garnier-Villarreal, M. (2018). semTools: Useful tools for structural equation modeling. R Package Version 0. 5.

Khalid, S., Deska, J. C., & Hugenberg, K. (2016). The eyes are the windows to the mind. *Personality and Social Psychology Bulletin*, *42*(12), 1666–1677. https://doi .org/10.1177/0146167216669124

Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology*, *90*(4), 543–555.

Krumhuber, E. G., Lai, Y.-K., Rosin, P. L., & Hugenberg, K. (2019). When facial expressions do and do not signal minds: The role of face inversion, expression dynamism, and emotion type. *Emotion*, *19*(4), 746–750.

Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology*, *109*(5), 901–931. https://doi.org/10.1037/pspp0000048

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct validity of the need to belong scale: Mapping the nomological network. *Journal of Personality Assessment*, *95*(6), 610–624. https://doi.org/10.1080/00223891 .2013.819511

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Mueser, K. T., Grau, B. W., Sussman, S., & Rosen, A. J. (1984). You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology*, *46*(2), 469–478. https://doi.org/10 .1037/0022-3514.46.2.469

Powers, K. E., Worsham, A. L., Freeman, J. B., Wheatley, T., & Heatherton, T. F. (2014). Social connection modulates perceptions of animacy. *Psychological Science*, *25*(10), 1943–1948. https://doi.org /10.1177/0956797614547706

R Core Team. (2019). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project .org/

Reis, H. T., Wilson, I. M., Monestere, C., Bernstein, S., Clark, K., Seidl, E., . . . Radoane, K. (1990). What is smiling is beautiful and good. *European Journal of Social Psychology*, *20*(3), 259–267. https://doi .org/10.1002/ejsp.2420200307

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2). https://doi. org/10.18637/jss.v048.i02

Seidel, E. M., Habel, U., Finkelmeyer, A., Schneider, F., Gur, R. C., & Derntl, B. (2010). Implicit and explicit behavioral tendencies in male and female depression. *Psychiatry Research*, *177*(1-2), 124–130. https://doi.org/10.1016/j. psychres.2010.02.001

Shin, H. I., & Kim, J. (2018). My computer is more thoughtful than you: Loneliness, anthropomorphism and dehumanization. *Current Psychology*, *39*(2), 445–453. https://doi.org/10.1007/s12144-018 -9975-7

Sim, M., Almaraz, S. M., & Hugenberg, K. (2021). Bodies and minds: Heavierweight targets are de-mentalized as lacking in mental agency. *Personality and Social Psychology Bulletin*. Advance

online publication. https://doi.org/10 .1177/01461672211039981

Vaes, J., Paladino, M. P., Castelli, L., Leyens, J.-P., & Giovanazzi, A. (2003). On the behavioral consequences of infrahumanization: The implicit role of uniquely human emotions in intergroup relations. *Journal of Personality and Social Psychology*, *85*(6), 1016–1034. https://doi.org/10.1037/0022-3514.85.6.1016

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336