Basic and Applied Social Psychology

When Social Perception Goes Wrong: Judging Targets' Behavior toward Gay Versus Straight People

Jin X. Goh, Mollie A. Ruben & Judith A. Hall

Published online: 07 Jan 2019.

Submit your article to this journal ⬈

Article views: 63

View Crossmark data ⬈

Check for updates

When Social Perception Goes Wrong: Judging Targets' Behavior toward Gay Versus Straight People

Jin X. Goh[a], Mollie A. Ruben[b], and Judith A. Hall [c]

[a]University of Washington; [b]University of Maine; [c]Northeastern University

ABSTRACT
Accurate social perception depends on many factors, including the extent to which perceivers hold correct beliefs about how behaviors reflect the characteristic being judged. In Study 1, target participants recorded videos introducing themselves to either a gay or straight student who was ostensibly in another room. Unbeknownst to the targets, the other student was illusory and not real. Analysis of the targets' videos revealed that they behaved more positively toward the gay than straight student. Two subsequent studies demonstrated that new perceivers were *below* chance in guessing the illusory student's sexual orientation from watching the male targets' behavior, presumably because they expected to see negative behavior toward the gay illusory student. The study documents processes whereby social perception can go awry.

In the large literature on accuracy of interpersonal perception, the emphasis has consistently been on demonstrating how accurate people can be in judging the meanings of other people's cues. Researchers have demonstrated people's ability to accurately perceive other people's personality, social attributes, emotional states, and thoughts, to name a few examples (Hall & Bernieri, 2001; Hall, Schmid Mast, & West, 2016). Researchers marvel at finding significant, and sometimes high, accuracy for judging even very degraded or subtle stimuli—for example, recognizing facial expressions of emotion shown for extremely short durations (Matsumoto et al., 2000) and guessing target individuals' religious or political affiliations from photographs that have no obviously identifying cues (Tskhay & Rule, 2013). Although a high accuracy rate is the norm, especially with standardized tests of accuracy that allow researchers to predetermine overall accuracy rates during test development (Hall, Andrzejewski, Murphy, Schmid Mast, & Feinstein, 2008), accuracy is harder to attain in some domains such as lie detection and judging sexual orientation, where stimuli are not preselected to be judgeable. Nonetheless, the narrative surrounding these research domains usually describes how accuracy is still significantly above chance on average (Bond & DePaulo, 2006; Tskhay & Rule, 2013).

Of course, there must be many kinds of interpersonal judgment on which perceivers *cannot* be accurate beyond the guessing level. Chance-level accuracy occurs either because the relevant information is not present in the stimuli to be judged or because perceivers do not have sufficient knowledge to apply to the judgment task (Funder, 1995; Hartwig & Bond, 2011). However, it is difficult to know how often these situations have arisen in research, because an accuracy task on which people could perform only at the guessing level would probably not find its way into publication.

The present article documents a third possibility that is also rarely reported: the situation where perceivers are *systematically wrong*; that is, their scores are *lower* than the guessing (chance) level. Anecdotally, many people have embarrassing stories about times when they were completely off base in judging someone else's identity, thoughts, or emotions. In some of these situations they probably were just guessing and guessed wrong, but in others they were likely applying an incorrect judgment principle—such as listeners often do when told the story of the gravely injured boy whose father just died in the same car crash that injured the boy. The trauma surgeon declares, "I cannot operate on this child—he is my son!" Listeners predictably think such a reply is impossible because they had just learned the boy's father was dead. Situations in which people are systematically inaccurate in a social inference occur because an initial premise is wrong—in this case,

people assume that trauma surgeons cannot be female or that a child cannot have two fathers. Such situations, however, are rarely reported in research. An example is a study on lie-detection accuracy in which accuracy was significantly below the chance level when perceivers were manipulated to try harder on the task, presumably due to the systematic derailment of otherwise correct judgment strategies (Porter, McCabe, Woodworth, & Peace, 2007).

Assuming that relevant and judgeable information is present in the stimuli (Funder, 1995), systematically wrong responding could come about only if perceivers are operating on the basis of erroneous beliefs about how the characteristic being judged is reflected in the cues they see or hear. Inaccuracy actually involves the same processes that produce accuracy, as both fit within Brunswik's (1956) lens model. Specifically, perceivers apply a set of decision rules (called cue utilizations) based on knowledge, stereotypes, or expectations about how the characteristic being judged is revealed in the measured behaviors, and then the researcher determines whether these cue utilizations align with the behavior actually associated with that characteristic (cue validity). As an example, perceivers who rate targets who have cheerful and energetic voices as extraverted would be accurate in their cue utilization to the extent that extraverts actually have cheerful and energetic voices (Hirschmüller, Egloff, Nestler, & Back, 2013). But, inaccurate cue utilization—such as considering extraverts to have timid, soft voices—would contribute systematically to inaccuracy unless the ill effects of that erroneous belief were swamped by the salubrious effect of having many beliefs that are accurate. The lens model framework can be applied to many kinds of social judgments that could be accurate or inaccurate (Hall & Goh, 2017). It follows from the lens model logic that, given the presence of cue validities (i.e., the relevant information is available), systematic inaccuracy results from incorrect cue utilizations.

The current research uses a paradigm in which perceivers have to guess certain characteristics of an unseen individual based on another person's (target) behaviors. Judgments like this are not uncommon in everyday life. For instance, people might try to ascertain who is on the other end of a phone call when they can observe only the target person talking on the phone. Or people might study someone's conversational behavior to ascertain whether the person is talking to a friend or a stranger, or a superior or a subordinate. Two standardized tests of interpersonal judgment accuracy—the Interpersonal Perception Task (Costanzo & Archer, 1989, 1993) and its predecessor, the Social Interpretations Task (Archer & Akert, 1977)—used the phone call situation among their test items: The test taker sees a woman talking on the phone and has to guess whether she is talking with her mother, a female friend, or her boyfriend. In general, participants perform better than chance on these scenarios. In similar studies, perceivers could accurately discern whether people talking on the phone were speaking to a romantic partner or a friend (Farley, Hughes, & LaFayette, 2013; Montepare & Vega, 1988). In one study, however, cue utilizations were shown to be incorrect (Hall & Braunwald, 1981). Perceivers guessed the gender of a speaker's interaction partner from listening to the speaker's voice tone when addressing the partner. Perceivers applied the erroneous belief (cue utilization) that a woman speaking in a dominant tone was speaking to another woman when, in fact, women's dominant tone was directed toward men.

The current research also employed the unseen other paradigm. In Study 1, participants (targets) made a videotape introducing themselves to another student whom they thought was gay or straight and whom they expected to meet face-to-face in an upcoming interaction. This "other student" was actually illusory, and the interaction never took place. These videos were used as stimuli in Studies 2 and 3, which measured perceivers' accuracy for judging sexual orientation of the unseen other (illusory) student from watching the targets' behavior.

Study 1

Method

Participants

Two hundred undergraduates originally participated in exchange for partial course credit. Twenty-nine participants were excluded for failing manipulation checks (described next), camera malfunctions, being underage at the time of experiment without parental consent, not reporting gender, expressing discomfort, or not self-identifying as straight. There were 171 remaining participants (50.3% women; M_{age} = 19.11).

Procedure

After gaining consent, a female experimenter informed the participants that this was a study on first impressions and there was another student who would be completing the study alongside them, but ostensibly in the next room. Unbeknownst to the participants, this other student was not real. Participants

were told they would rate their first impressions based on limited information. First, participants were told that they would write a short autobiography that would supposedly be exchanged with the other student to make impression ratings. Second, they would make a video of themselves, which would also be exchanged for viewing and rating. Finally, participants were told that they would interact with the other student face-to-face, although this interaction never took place. This setup allowed us to manipulate the illusory student's gender and sexual orientation (see Vorauer & Turpie, 2004, for a similar procedure).

Participants were first given 5 min to write a short autobiography with the prompt that they could discuss their upbringing, hometown, path to college, possible majors and career paths, and any current romantic relationships. Afterward, the experimenter ostensibly exchanged the autobiographies, giving each participant a handwritten autobiography purportedly written by the other student that disclosed the other student's gender through stereotypical male and female names (Mike or Emily) and sexual orientation (has a boyfriend or girlfriend). For instance, participants paired with a gay male student received an autobiography ostensibly written by Mike who has a boyfriend. All information other than the gender and sexual orientation was exactly the same across conditions.

Participants then completed questions concerning their first impression of the other (illusory) student, which included a list of demographic questions as manipulation check. They then made a video introducing themselves to the other student. They were instructed to talk about themselves and respond to any information from the other student's autobiography. Participants were allowed to talk for up to 5 min.

Nonverbal behavior coding

All 171 videos were shortened to the first 30 s for uniform coding across videos due to variation across participants' speaking time (Ambady, Hallahan, & Conner, 1999; Murphy, 2005; Murphy et al., 2015). All videos were muted to ensure that only visible nonverbal cues were being coded and that the coders were not influenced by the gender and sexual orientation conditions (in case participants mentioned either of these in their videos). Three trained female research assistants individually coded all of the video excerpts, and only variables with intercoder reliability of at least .60 (Cronbach's alpha) were used for analysis.

Coders rated their overall global impressions of the participants' nonverbal behaviors (positive/friendly, engaged/interested, relaxed/comfortable, and assertive/dominant) on 7-point scales. Intercoder reliabilities for each of these four impressions were acceptable (Cronbach's $\alpha = .60-.84$). Because these four items correlated positively with one another ($rs > .21$), they were averaged to form one *global positivity* score ($\alpha = .79$). Global impression ratings were used as the primary coding procedure because research on lie detection (Hartwig & Bond, 2011), physical pain (Ruben & Hall, 2016a), and interracial interactions (Dovidio & LaFrance, 2013) have demonstrated that global impressions are better predictors of actual intentions and feelings than more specific nonverbal cues. These impressions were selected based on previous research on behavioral outcomes during intergroup interactions, and these nonverbal impressions have been successfully used in prior research as indicators of positive nonverbal behaviors during intergroup interactions (e.g., Dasgupta, 2004; Dasgupta & Rivera, 2006; Goh & Hall, 2015; Mendes & Koslov, 2013).

Manipulation check

Participants were asked to fill out a demographic information sheet for the other student as a manipulation check. For gender, participants indicated whether they thought the other student was male, female, or other (with a fill-in option). For sexual orientation, participants chose from one of four options: gay, straight, bisexual, or other (with a fill-in option). Those who failed to identify the other student's gender and sexual orientation correctly according to their assignment were excluded from analysis. To be conservative, participants who said the other student was bisexual as opposed to gay were also excluded. A research assistant watched all of the videos to examine if participants appeared to take the task seriously and if they completed it. All participants completed the video-making task and did not question whether the other student was real or fake in their videos or afterward.

Results

The goal of Study 1 was to see how participants behaved toward an illusory student who was believed to be gay or straight. We used Cohen's *d* as our unit of effect size, with convention demarking small ($d = .20$), medium ($d = .50$), and large ($d = .80$) effect size.

Participants gave more positive nonverbal impressions when they believed that the student they were addressing was gay ($n = 84$, $M = 4.62$, $SD = .82$) compared to straight ($n = 87$, $M = 4.26$, $SD = .65$); this between-group difference had a medium effect size ($d = .49$). Of import, participants did not behave differently toward a gay male student ($n = 46$, $M = 4.61$, $SD = .83$) relative to a gay female student ($n = 38$, $M = 4.64$; $SD = .82$)—effect size was near zero ($d = -.04$), suggesting that the preferential treatment of gay student over straight student was not affected by the gay student's manipulated gender.[1]

Study 1 showed that participants behaved more positively when introducing themselves on video to a gay student relative to a straight student. There may be more than one explanation for this result; as we elaborate later, the college student targets may have been authentically signaling their positive attitudes, or they may have concealed different attitudes behind a friendly facade. What matters for the current research is the targets' actual behavior captured in the videos. Study 1 provided the necessary stimuli for Studies 2 and 3, in which perceivers viewed excerpts of these videos and made judgments of the unseen student's sexual orientation.

Study 2

Method

Stimuli and coding

Of the 171 participants in Study 1, 137 consented to release their videos for future studies. From these 137 videos of participants (henceforth called *targets*), 40 videos of male targets and 40 videos of female targets were randomly selected with equal distribution of the other (illusory) student's gender and sexual orientation conditions. For instance, in the set of 40 male targets, 10 targets addressed a gay male student, 10 addressed a straight male student, 10 addressed a gay female student, and 10 addressed a straight female student. The first 5 s of each video was discarded because targets often mentioned their names and the gender or sexual orientation of the other student at the beginning. We then extracted the following 15 s for use in both Studies 2 and 3 to conserve time and avoid participant fatigue. Videos in Study 2 (and in

Study 3, next) were shown with the audio on. The order of video clips by the other student's sexual orientation and gender were intermixed.

Because these stimulus videos contained shorter excerpts than those coded in Study 1, and were shown with audio on rather than off, it was necessary to confirm that Study 1's finding of more positivity toward the gay than straight other student would still show up in the shorter clips. Two new independent coders rated each 15 s video muted (nonverbal channel), and another two independent coders rated the same videos with the audio on (full channel); all coders were blind to the sexual orientation conditions that the targets were in, and these coders were not involved in the coding for Study 1. All ratings were based on five global impressions using 7-point scales: smile a lot, positive/friendly, engaged/interested, relaxed/comfortable, and dominant/assertive. The five ratings were averaged to form a global positivity score for nonverbal channel ($\alpha = .90$) and full channel ($\alpha = .89$). Nonverbal and full channel correlated strongly and positively ($r = .75$).

Analysis showed that, as in Study 1, targets were rated as more positive nonverbally when speaking to the gay ($n = 40$, $M = 4.56$, $SD = 1.43$) than straight ($n = 40$, $M = 3.86$, $SD = 1.52$) other student, with a medium effect size ($d = .47$). For full-channel positivity, targets were also rated as behaving more positively toward the gay other student ($n = 40$, $M = 4.41$, $SD = 1.20$) than the straight other student ($n = 40$, $M = 3.80$, $SD = 1.39$), with the same magnitude as before ($d = .47$). Thus, even when shorter video clips were used, the greater positivity directed toward the gay than straight other student was still evident.

Participants and procedure

Fifty-seven new straight undergraduate students (77.2% women; $M_{age} = 19.17$) were then recruited for Study 2 as perceivers. They were run in groups of one to four. Each group was randomly assigned to watch videos of all male or all female targets. Perceivers watched the videos with the sound on (full channel), and the videos were displayed on a television. The experimenter explained the procedure of the stimuli-generation phase, including the fact that the people in the videos (targets) were introducing themselves to an unseen (illusory) student whom the targets believed was either gay or straight. Perceivers' task was to guess whether the target was addressing a gay or a straight student. Perceivers recorded their responses immediately after watching each video clip on paper,

---

[1]The difference between positive nonverbal impressions toward the manipulated male student ($n = 88$, $M = 4.36$, $SD = .78$) compared to the female student ($n = 83$, $M = 4.52$, $SD = .73$) was small ($d = -.21$). Furthermore, male participants ($n = 85$, $M = 4.40$, $SD = .77$) did not behave differently than female participants ($n = 86$, $M = 4.48$, $SD = .75$), with a small effect size ($d = -.11$).

which had the options of circling their guesses as straight or gay/lesbian for each target.

Calculation of accuracy scores

Each perceiver watched 40 videos of male targets or 40 videos of female targets, and each perceiver judged the sexual orientation (straight or gay/lesbian) of the student being addressed by the target after watching each video. Accuracy for judging sexual orientation was scored as 1 for correctly identifying gay or straight and 0 for incorrect guesses. Across all videos, each perceiver's mean accuracy score could range from 0 to 1, with .50 as chance (guessing level).

After Study 2 was completed, it was discovered that one of the female targets (talking to gay Emily) did not pass our conservative manipulation check (i.e., indicated Emily was bisexual instead of gay). For this reason, ratings of this particular target were excluded, and all accuracy scores for the female target condition were based on 39 videos rather than 40 videos.

Results and discussion

We calculated effect size (d) by subtracting chance level (.50) from the mean of all the individual participants' mean accuracy scores and then dividing the difference by the standard deviation. A positive effect size indicates that perceivers were accurate above chance, whereas a negative effect size indicates that perceivers were below chance (i.e., inaccurate), and an effect size of zero means their accuracy was at the guessing level. Overall, perceivers ($n = 57$, $M = .49$, $SD = .07$) were slightly below chance ($d = -.14$). Inaccuracy was substantially more evident when perceivers were making judgments from male targets' behaviors specifically ($n = 30$, $M = .46$, $SD = .07$, $d = -.57$). Perceivers' accuracy in judging the illusory student's sexual orientation from female targets was slightly above chance ($n = 27$, $M = .51$, $SD = .06$, $d = .17$).[2]

This study examined accuracy in detecting the sexual orientation of the recipient of a communication based on short excerpts of the targets' behaviors. Perceivers were below chance in judging the unseen illusory student's sexual orientation when the targets (i.e., the speakers) were men. This showed that, at a rate exceeding chance, perceivers said the male targets were addressing a gay student when it was actually a straight student, and vice versa. To be systematically

wrong in this way, perceivers would first have detected the behavioral differences in the video clips, which they could do (because our coders could), and then made the wrong inference regarding the sexual orientation of the other student based on their assumption about how targets would speak to a gay versus straight student. In other words, perceivers made the assumption that male targets behaved more positively toward straight than gay people and they were wrong because, in fact, the difference went the other way. Study 3 was designed as a replication of Study 2.

Study 3

Method

Participants and procedure

Eighty-eight straight undergraduate students (75.0% women; $M_{age} = 18.45$) were recruited as perceivers. They were run in groups of one to four. As in Study 2, each group was randomly assigned to watch videos of male targets or female targets displayed on a television and perceivers recorded their responses on paper with the options to circle either straight or gay/lesbian as their guess for each target.[3]

Videos

Among the set of female target videos, one new target was selected to replace the target who failed the manipulation check (see Study 2 Method). The set of male target videos remained the same as in Study 2.

Results and discussion

Accuracy scores and effect sizes were calculated in the same manner as in Study 2. Overall, perceivers performed below chance in judging the sexual orientation of the other student ($n = 88$, $M = .48$, $SD = .07$, $d = -.29$). This overall inaccuracy was again primarily driven by those in the male target condition. When perceivers made judgments based on male targets' behaviors, they performed considerably below chance ($n = 45$, $M = .45$, $SD = .06$, $d = -.83$), replicating Study 2. Perceivers' accuracy for judging female

---

[2]Overall accuracy rate did not differ greatly between male perceivers ($n = 13$, $M = .49$, $SD = .08$) and female perceivers ($n = 44$, $M = .48$, $SD = .06$, $d = .15$).

[3]Study 3 included an additional manipulation that sought to increase accuracy through an instructional change for half of the perceivers. They were instructed that sometimes people act very positively even when they hold negative attitudes or anxiety toward certain social groups. This did not influence accuracy: Perceivers who received the instructional change ($n = 44$, $M = .48$, $SD = .08$) did not differ much from those that did not receive additional information ($n = 44$, $M = .47$, $SD = .07$, $d = .13$).

targets did not differ from chance ($n = 43$, $M = .50$, $SD = .08$, $d = .00$).[4]

Mini meta-analysis of Studies 2 and 3

We next conducted a "mini" meta-analysis to examine the mean effect sizes across Studies 2 and 3 (Goh, Hall, & Rosenthal, 2016). We first converted the effect size metric from d to r. These correlations were Fisher's z transformed for analysis and converted back to regular Pearson correlation as well as to Cohen's d for presentation. As earlier, a positive effect indicates accuracy above chance and a negative effect indicates accuracy below chance (i.e., inaccuracy). Overall, perceivers performed below chance with a small meta-analytic effect size ($N = 145$, M $r = -.11$, $d = -.22$). Perceivers were more inaccurate when viewing male targets specifically ($N = 75$, M $r = -.34$, $d = -.72$). Finally, perceivers' accuracy was close to chance when viewing female targets ($N = 70$, M $r = .03$, $d = .06$). Thus, across both studies, inaccuracy occurred in judging sexual orientation of the unseen other student when perceivers were observing male targets' behavior.

It is important to note that these results were not due to perceivers simply guessing "gay" less often than "straight" to match what they may have thought was the base rate in the student population (i.e., that there are fewer gay than straight people). We can assert this for two reasons. First, if perceivers were simply using perceived base rates (i.e., deliberately choosing "gay" less often than "straight" but in a random fashion), the same degree of inaccuracy would have occurred for judgments of female targets, but this did not happen. Second, we correlated the coded full-channel positivity in the 15-s clips used in Studies 2 and 3 with the pooled "gay" (= 1) versus "straight" (= 0) guesses from both studies. This correlation was substantially negative in the case of judgments of male targets ($r = -.45$, $N = 40$ targets). In other words, when male targets were rated as displaying more positivity, perceivers were more likely to guess that these targets were addressing a straight student (and less likely to guess that the student was gay). The corresponding correlation for judgments of female targets was very small ($r = -.10$, $N = 40$ targets). These correlations demonstrate that guessing "gay" versus "straight" was tied to the cues displayed in the video clips.

---

[4]Accuracy rate did not differ by perceivers' gender. Male perceivers ($n = 22$, $M = .48$, $SD = .07$) did not differ greatly from female perceivers ($n = 66$, $M = .47$, $SD = .08$, $d = .13$).

General discussion

This research examined circumstances under which social perception can go awry. In Study 1, participants (targets) introduced themselves in a video that they thought would be shown to a student they believed was in another room and whom they believed was gay or straight. Video coding found that the targets behaved more positively toward the gay than straight student. This finding provided the basis for testing whether perceivers would be accurate in guessing the other (unseen) student's sexual orientation from watching excerpts of these videos. The answer would depend on the judgment processes used by perceivers, which would presumably be guided by their stereotypes and knowledge of data on negative attitudes about homosexuality in the population, and perhaps by their own possibly negative attitudes about (or discomfort with) homosexuality. If, for any of these reasons, they assumed that a gay student would be addressed with more negativity than a straight student, this would logically mislead their judgments and produce below-chance accuracy.

Studies 2 and 3 both found exactly this result: inaccuracy (i.e., guessing below chance), but only when perceivers' judgments were based on viewing male targets' (not female targets') behavior. Men generally hold more antigay prejudice than women (Herek & McLemore, 2013), and perceivers may have used this knowledge to assume that the male targets would be conflicted (uncomfortable, anxious, or even hostile) when introducing themselves to a gay student, and they would exhibit more negativity toward the gay student as a result. Given that targets in our stimulus set actually conveyed more *positive* behaviors toward the gay than straight student, such a belief would have led perceivers to make incorrect judgments about male targets.

It may be surprising that Study 1 found more positive behavior directed toward a gay person considering that people generally express more nonverbal negativity toward someone from a different social group (Stephan, 2014; Toosi, Babbitt, Ambady, & Sommers, 2012). However, as societal demand to be nonprejudiced increases, overt expression of negativity toward minorities is discouraged (Dasgupta, 2004; Mendes & Koslov, 2013; Norton, Dunn, Carney, & Ariely, 2012; Plant & Devine, 1998). Furthermore, the targets were recruited from a major private university in one of the most lesbian, gay, bisexual, and transgender friendly cities in the United States. Nonetheless, it is interesting that the perceivers, who were from the very same population as the original

targets whose behavior they judged (and arguably would share the same knowledge, attitudes, norms, and stereotypes with participants in Study 1), still misattributed their behavior and were inaccurate with respect to male targets in particular. It appears that perceivers' stereotype-based assumptions outweighed any awareness of their peers' likely *actual* behaviors towards gay and straight people.

Inaccurate perception can hold implications for intergroup relations, and it informs us about the nature of interpersonal social perception as well. In the field of interpersonal accuracy, perceivers are often accurate (or at least above chance) in judging the qualities or characteristics of visible or unseen targets (Hall et al., 2016). It is extremely rare for research to report *in*accuracy. However, human beings are not perfect interpreters of social information. When perceivers are provided false evidence or when they hold a wrong theory about information given to them, they would, predictably, be inaccurate in their judgment.

The present findings can be placed within Funder's (1995, 2012) realistic accuracy model (RAM), which offers four preconditions for making accurate judgments of others. The RAM argues that cues relevant to the aspect of the target that is being judged must be available for judges (i.e., perceivers) to perceive, and then judges must detect and correctly use (i.e., interpret) these cues. The RAM was developed for accurate perception of personality traits but has been applied in many other judgment domains (e.g., Ruben & Hall, 2016b) and is a suitable framework for understanding the current results.

In our stimuli, the relevant cues were displayed and were available (or else the coders would not have observed them), and clearly the perceivers detected them (or else they could not have been inaccurate). So something went wrong at the cue utilization stage. Specifically, perceivers understood which cues were positive, but they misunderstood what the positive cues meant to the targets. Perceivers thus exercised good cue utilization for the lower order inference about the positivity of cues but fell down on the higher order inference about how positivity reflected the targets' intentions, assuming the positivity was a manifestation of attitudes that were more favorable to straight than gay students.

In any social context where targets have attitudes and motives vis-à-vis communication partners, perceivers may make erroneous assumptions, even to the point of being systematically inaccurate. What we learn from the present analysis is that cue utilization involves a hierarchy of judgments that likely become increasingly error prone as one moves from simpler judgments to more complex ones. In the field of personality judgment, a relatively simple judgment might be what emotion a face showing lowered brows and a tight mouth conveys, whereas the more complex judgment would be how the identified emotion contributes to a correct assessment of someone's personality (Hall, Gunnery, Letzring, Carney, & Colvin, 2017). Or, in the present study, the first judgment is about the affective tone of the cues and the second, more complex judgment is about what attitudes or intentions motivate the target to send those cues. If the second judgment rests mainly on guesswork, we might expect accuracy to hover around the chance level; but if the second judgment rests on a shared stereotype that happens to be wrong, accuracy will dip below chance as in the present studies. It follows that if the shared stereotype is correct, accuracy would follow, other things being equal (Hall & Goh, 2017).

We hope that the present studies will encourage researchers to consider inaccuracy more critically and willingly (Hall & Goh, 2017). To fully comprehend the extent or limits of accurate social perception, research is needed to understand the conditions under which perceivers are accurate as well as inaccurate and what types of information and interventions can alter this inaccuracy. In the real world, inaccurate judgments of an unseen person's sexual orientation (e.g., being present during a friend or colleague's phone conversation) could lead one to treat the unseen person differently once they meet them as first impressions of others have strong and permeating influences on our future attitudes and behaviors (Ambady & Skowronski, 2008). These inaccurate judgments could also lead one to treat the seen person differently. For example, if one believed their friend or colleague was being less warm over the phone because the other person was a sexual minority, we may like that person less. For a characteristic as sensitive and personally relevant as sexual orientation, factors permitting erroneous judgment should be taken seriously as potential sources of confusion, pain, or frustration.

References

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology, 77*, 538–547.

Ambady, N., & Skowronski, J. J. (Eds.). (2008). *First impressions.* New York: Guilford Press.

Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation.

Journal of Personality and Social Psychology, 35, 443–449. doi:10.1037/0022-3514.35.6.443

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10, 214–234.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13, 225–245. doi:10.1007/BF00990295

Costanzo, M., & Archer, D. (1993). Interpersonal Perception Task-15. Center for Media and Independent Learning, University of California Extension.

Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, 17, 143–169. doi:10.1023/B:SORE.0000027407.70241.15

Dasgupta, N., & Rivera, L. M. (2006). From autonomic anti-gay prejudice to behavior: The moderating role of conscious beliefs about gender and behavioral control. *Journal of Personality and Social Psychology*, 91, 268–280. doi:10.1037/0022-3514.91.2.268

Dovidio, J. F., & LaFrance, M. (2013). Race, ethnicity, and nonverbal behavior. In J. A. Hall & M. L. Knapp (Eds.), *Nonverbal communication* (pp. 671–695). Berlin: deGruyter Mouton.

Farley, S. D., Hughes, S. M., & LaFayette, J. N. (2013). People will know we are in love: Evidence of differences between vocal samples directed toward lovers and friends. *Journal of Nonverbal Behavior*, 37, 123–138. doi:10.1007/s10919-013-0151-3

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670. doi:10.1037/0033-295X.102.4.652

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21, 177–182. doi:10.1177/0963721412445309

Goh, J. X., & Hall, J. A. (2015). Nonverbal and verbal expressions of men's sexism in mixed-gender interactions. *Sex Roles*, 72(5-6), 252–261. doi:10.1007/s11199-015-0451-7

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, 10, 535–549. doi:10.1111/spc3.12267

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Schmid Mast, M., & Feinstein, B. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42, 1476–1489. doi:10.1016/j.jrp.2008.06.013

Hall, J. A., & Bernieri, F. J. (Eds.). (2001). *Interpersonal sensitivity: Theory and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hall, J. A., & Braunwald, K. G. (1981). Gender cues in conversations. *Journal of Personality and Social Psychology*, 40, 99–110. doi:10.1037/0022-3514.40.1.99

Hall, J. A., & Goh, J. X. (2017). Studying stereotype accuracy from an integrative social-personality perspective. *Social & Personality Psychology Compass*, 11, 1–10.

Hall, J. A., Gunnery, S. D., Letzring, T., Carney, D. R., & Colvin, C. R. (2017). Accuracy of judging affect and accuracy of judging personality: How and when are they related? *Journal of Personality*, 85, 583–592.

Hall, J. A., Schmid Mast, M., & West, T. V. (Eds.). (2016). *The social psychology of perceiving others accurately*. Cambridge, UK: Cambridge University Press.

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, 137, 643–659.

Herek, G. M., & McLemore, K. A. (2013). Sexual prejudice. *Annual Review of Psychology*, 64, 309–333.

Hirschmüller, S., Egloff, B., Nestler, S., & Back, M. D. (2013). The dual lens model: A comprehensive framework for understanding self–other agreement of personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 104, 335–353. doi:10.1037/a0030383

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., … Goh, A. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior*, 24, 179–209. doi:10.1023/A:1006668120583

Mendes, W. B., & Koslov, K. (2013). Brittle smiles: Positive biases toward stigmatized and outgroup targets. *Journal of Experimental Psychology: General*, 142, 923–933. doi:10.1037/a0029663

Montepare, J. M., & Vega, C. (1988). Women's vocal reactions to intimate and casual male friends. *Personality &Amp; Social Psychology Bulletin*, 14, 103–113.

Murphy, N. A. (2005). Using thin slices for behavioral coding. *Journal of Nonverbal Behavior*, 29, 235–246. doi:10.1007/s10919-005-7722-x

Murphy, N. A., Hall, J. A., Schmid Mast, M., Ruben, M. A., Frauendorfer, D., Blanch-Hartigan, D., … Nguyen, L. (2015). Reliability and validity of nonverbal thin slices in social interactions. *Personality and Social Psychology Bulletin*, 41, 199–213. doi:10.1177/0146167214559902

Norton, M. I., Dunn, E. W., Carney, D. R., & Ariely, D. (2012). The persuasive "power" of stigma? *Organizational Behavior and Human Decision Processes*, 117, 261–268. doi:10.1016/j.obhdp.2011.08.002

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811–832. doi:10.1037//0022-3514.75.3.811

Porter, S., McCabe, S., Woodworth, M., Peace, K. A. (2007). 'Genius is 1% inspiration and 99% perspiration' … or is it? An investigation of the impact of motivation and feedback on deception detection. *Legal and Criminological Psychology*, 12, 297–309. doi:10.1348/135532506X143958

Ruben, M. A., & Hall, J. A. (2016). A lens model approach to the communication of pain. *Health Communication*, 31, 934–945.

Ruben, M. A., & Hall, J. A. (2016). Healthcare providers' nonverbal behavior can lead patients to show their pain

more accurately: An analogue study. *Journal of Nonverbal Behavior*, 40, 221–234. doi:10.1007/s10919-016-0230-3

Stephan, W. G. (2014). Intergroup anxiety: Theory, research. *Personality and Social Psychology Review*, 18, 239–255. doi:10.1177/1088868314530518

Toosi, N. R., Babbitt, L. G., Ambady, N., & Sommers, S. R. (2012). Dyadic interracial interactions: A meta-analysis. *Psychological Bulletin*, 138, 1–27. doi:10.1037/a0025767

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A. *Personality and Social Psychology Review*, 17, 72–86. doi:10.1177/1088868312461308

Vorauer, J. D., & Turpie, C. A. (2004). Disruptive effects of vigilance on dominant group members' treatment of outgroup members: Choking versus shining under pressure. *Journal of Personality and Social Psychology*, 87, 384–399. doi:10.1037/0022-3514.87.3.384