



**HAL**  
open science

## Do Observer Ratings Validate Self-Reports of Intergroup Contact?: A Round-Robin Analysis

Miles Hewstone, Charles M. Judd, Melanie Sharp

► **To cite this version:**

Miles Hewstone, Charles M. Judd, Melanie Sharp. Do Observer Ratings Validate Self-Reports of Intergroup Contact?: A Round-Robin Analysis. *Journal of Experimental Social Psychology*, 2011, 10.1016/j.jesp.2010.12.014 . hal-00962265

**HAL Id: hal-00962265**

**<https://hal.science/hal-00962265>**

Submitted on 21 Mar 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Do Observer Ratings Validate Self-Reports of Intergroup Contact?: A Round-Robin Analysis

Miles Hewstone, Charles M. Judd, Melanie Sharp

PII: S0022-1031(10)00282-9  
DOI: doi: [10.1016/j.jesp.2010.12.014](https://doi.org/10.1016/j.jesp.2010.12.014)  
Reference: YJESP 2591

To appear in: *Journal of Experimental Social Psychology*

Received date: 14 October 2008  
Revised date: 11 December 2010



Please cite this article as: Hewstone, M., Judd, C.M. & Sharp, M., Do Observer Ratings Validate Self-Reports of Intergroup Contact?: A Round-Robin Analysis, *Journal of Experimental Social Psychology* (2010), doi: [10.1016/j.jesp.2010.12.014](https://doi.org/10.1016/j.jesp.2010.12.014)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Do Observer Ratings Validate Self-Reports of Intergroup Contact?:

A Round-Robin Analysis

Miles Hewstone (*University of Oxford*)

Charles M. Judd (*University of Colorado at Boulder*)

&

Melanie Sharp (*University of Oxford*)

December 9, 2010

Contact information:

Correspondence concerning this paper should be addressed to Miles Hewstone, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK. Email: miles.hewstone@psy.ox.ac.uk.

*Author note:* The authors acknowledge financial support from the Leverhulme Trust to Miles Hewstone and Charles Judd and the Economic and Social Research Council (UK) to Melanie Sharp, and they thank Diane Mackie and two anonymous reviewers for their constructive comments on earlier versions of this paper.

## Abstract

Two studies used a round-robin design to examine whether observers made consensual judgments of targets' degree and quality of intergroup contact, and whether these consensual judgments were correlated with the targets' own self ratings, and moderated by the observability of the contact. Study 1 revealed projection/assumed similarity, with participants rating others as similar to themselves to a large extent, but also yielded evidence for the validity of whites' self-reports of direct, but not extended, intergroup contact with Asians, even when controlling for extraversion and perceived attitudes. Study 2 replicated the main results, using both Asians and Gay men as outgroups, and showed that participants' ratings discriminated between the two discrete outgroups, with measures of contact and attitude being only meaningfully related within, but not between, outgroups. Overall, these findings help to validate self-report measures of direct intergroup contact.

Keywords: assumed similarity, bias, intergroup contact, observer ratings, projection, self-reports

Since Allport's (1954) pioneering research social psychologists have placed great faith in, and accumulated research evidence for, the reliable effect of intergroup contact in decreasing prejudice (for meta-analytic evidence, see Pettigrew & Tropp, 2006). The most impressive evidence surrounds Allport's 'key conditions' for direct, face-to-face contact (i.e., equal status, cooperation, common goals, institutional support) and the prejudice-reducing benefits of individuals forming direct friendships with outgroup members (for a review, see Turner, Hewstone, Voci, Paolini, & Christ, 2008). Further evidence in favour of the impact of intergroup contact comes from studies on 'extended' contact (Wright, Aron, McLaughlin-Volpe, & Ropp, 1997), showing that mere knowledge that other ingroup members have outgroup friends contributes, above and beyond direct contact, to prejudice reduction.

While the literature in support of the beneficial effects of contact is indeed impressive, one potentially serious threat to the validity of that work comes from the fact that the vast majority of the supportive research has assessed the crucial predictor variable, contact, using self-report measures. In Pettigrew and Tropp's (2006) meta-analysis of over 500 studies on intergroup contact, 81% of studies used self-report measures of contact. It is interesting to note that the magnitude of the effect of contact in studies using self-reports of contact ( $r = .210$ ) was significantly smaller than in studies using directly observed measures of contact ( $r = .246$ ); this could indicate that self-reports of contact are less accurate than direct measures of contact. In the vast majority of research on intergroup contact, the outcome measure has also been a self-report measure, typically of outgroup attitude. Our focus in this research is, however, on the potential limitations of self-report measures of contact, because several studies of intergroup contact have already used alternatives to self-report for the measurement of attitudes (e.g., implicit measures of attitude, Turner, Hewstone, & Voci, 2007), and response-time measures of attitude accessibility (e.g., Vonofakou, Hewstone, &

Voci, 2007). Thus doubts remain primarily about the self-report nature of measures of contact itself.

Notwithstanding this reliance on self-report measures of contact, there are well-known limitations to the validity of self-reports, two of which we wish to highlight here: *socially desirable responding (SDR)*, and *acquiescent responding* (Paulhus, 1991). Socially desirable responding is the response bias that has received the most research attention. Often participants tend to present themselves in an unrealistically positive light, either because they deceive themselves into believing that this is how they really are (self-deception; Paulhus & John, 1998), or because they wish to control how they appear to others (impression management; Barrick & Mount, 1996; Paulhus, 1984, 1986). Particularly when self-reports assess a socially sensitive issue, such as attitudes towards other races, results can be distorted (Sigall & Page, 1971). Although techniques exist to measure SDR (for a review, see Paulhus, 1991), some researchers recommend against partialing out SDR scores from the content measure of interest (Paulhus & Vazire, 2007), because they typically remove valid variance and may reduce the validity of the content measure (see Ones, Viswesvaran, & Reiss, 1996). Some researchers also argue that in many research contexts, and in research with student or volunteer samples, there is little reason to be concerned about contamination from socially desirable responding (Piedmont, McCrae, Riemann, & Angleitner, 2000).

Acquiescence (or, more rarely, its opposite) occurs when people tend to express agreement (or, more rarely, disagreement) with whatever they are asked (Dicken, 1963; Lentz, 1938). Paulhus and Vazire (2007) argue that the major research problem posed by this bias is that it exaggerates correlations between consistently-valenced items, and attenuates correlations between inconsistently-valenced items, and is best dealt with by having equal numbers of items on which high (or low) ratings indicate possession of a trait.

The disadvantages of self-reports have been reviewed extensively in social and personality psychology (e.g., Funder 1999; Funder & Colvin, 1997; John & Robins, 1993; Kolar, Funder, & Colvin, 1996; Paulhus & Vazire, 2007). Other disadvantages include their credibility when motives other than accuracy may shape self-reports (see Sedikides & Strube, 1995); these motives include, notably, consistency-seeking, self-enhancement, and self-presentation (Robins & John, 1997).

We see SDR as the major threat to the validity of self-report measures of contact, and this bias will likely be more pronounced for outgroups against whom prejudice is considered more socially unacceptable (see Crandall, Eshleman, & O'Brien, 2002). As a result, studies that use self-reports to examine the contact hypothesis may obtain inflated correlations in support of the effect of contact.

In response to these issues, researchers have developed alternatives to self-report measures of contact, and these have been used in 19% of the studies in Pettigrew and Tropp's (2006) meta-analysis. One proposed solution is to use observational measures of contact (e.g., Cummings & Lambert 1997; Fossett & Kiecolt, 1989; Giles, 1977; Hood & Morris 1997; Schofield & Sagar, 1977). Observational approaches to assessing contact have, however, commonly involved simply examining the percentage of non-ingroup members (Cummings & Lambert 1997; Fossett & Kiecolt, 1989; Giles, 1977) or the percentage of members of various minority groups (Hood & Morris 1997) in the respondents' area of residence. These methods, however, only measure *opportunity* for contact (Wagner, Hewstone, & Machleit, 1989), and do not ascertain whether or not that opportunity is taken up. Furthermore, these 'macro-level' observational measures of contact do not take the *quality* of the contact into account, and hence cannot assess whether the relevant conditions for beneficial contact (Allport, 1954) are met.

More ‘micro-level’ observational studies (e.g., Campbell, Kruskal, & Wallace, 1966; Clack, Dixon & Tredoux, 2005; McCauley, Plummer, Moskalenko, & Mordkoff, 2001; Schofield & Sagar, 1977) record *actual* contact in a certain situation, such as seating preferences in a cafeteria. As the situation (physical, temporal, political, historical etc) is known, this method permits the researcher to ascertain to some extent the presence of Allport’s conditions. However, this method of contact measurement is only of use with a view to examining the contact existing in general in a location or organization (e.g., for evaluating the success of a desegregation intervention), and has not been used as a measure that varies between individuals within these locations. Individual attitudes of those who are observed have also not been measured. Because these studies have not related individual-level measures of contact to individual-level measures of attitude, their ability to assess the effectiveness of contact is limited (indeed, on this criterion they would have been excluded from Pettigrew & Tropp’s, 2006, meta-analysis).

A more robust method for research on the contact hypothesis may be that used in experimental studies in which contact is manipulated and its causal effect examined (e.g., Wilder, 1984). While this approach provides strong evidence for the causal effects of contact, it cannot, however, examine the various nuanced forms that contact can and does take outside the laboratory, for which self-reports appear to be invaluable.

It is, thus, clear that proposed alternatives to self-reports of contact have problems of their own, and moreover one should not overlook the unique advantages of self-report measures. According to Paulhus and Vazire (2007), self-reports have numerous advantages including their ease of interpretability by researchers, the richness of information they provide, participants’ motivation to report, their influence on how people interact with the world (their ‘causal force’), and their practicality in terms of costs and time (see also Funder 1999; Funder & Colvin, 1997; John & Robins, 1993; Kolar et al., 1996).



Given, on the one hand, the reliance on self-report measures of contact in most research on contact and prejudice reduction and, on the other hand, doubts about the validity of self-reports (especially in the domain of prejudice), it is important that we try to validate self-reports of intergroup contact. We suggest doing so by comparing self-reports to observer-reports, that is, what others who know the target say about the target's intergroup contact. Self-reports and observer-reports are unlikely to share the same systematic biases. As a result, the correlation between the two should largely reflect true score variance rather than shared error variance, thus permitting validity assessments of the self-reports (Borkenau & Ostendorf, 1992; Piedmont et al., 2000). To take as an example the self-report bias of socially desirable responding, identified above as a potential threat to the validity of self-reports of contact, while such a bias may distort *self*-reports of sensitive issues involving intergroup contact, it is less likely that it would bias observer-ratings (Krahé, 2007).

The field of personality research has recently provided considerable evidence for the validity of observer-reports (Fiedler, Oltmanns, & Turkheimer, 2004; Oltmanns, Gleason, Klonsky, & Turkheimer, 2005; Vazire, 2006; Vazire & Mehl, 2008; Wagerman & Funder, 2007), hence we propose their usefulness in contact research too. Observer-reports, especially when they are from judges who actually know the target well, rather than merely observe them in a brief study, are frequently found to be at least as similar to objective assessments as are self-reports (Kolar et al., 1996; Vazire & Mehl, 2008). A number of studies have found, as one would expect, that length of acquaintance positively moderates self-other agreement (e.g., Biesanz, West & Millevoi, 2007; Connolly, Kavanagh, & Viswesvaran, 2007; Vazire, 2010; but cf. Bernieri, Zuckerman, Koestner, & Rosenthal, 2010), and even exposure time within interactions lasting just five minutes was associated with increased accuracy (Carney, Colvin, & Hall, 2007).

Despite the advantages of observer reports, we acknowledge that they also have their own limitations. For example, previous research (albeit with unacquainted observers) has demonstrated the phenomenon of ‘observer harshness’ (e.g., Campbell & Fehr, 1990; Gotlib & Meltzer, 1987), whereby observers/judges tend to be overly harsh in their ratings of targets. Hence we should not rush to conclude that social consensus implies accuracy, but rather we offer observer reports as a *complement* to the use of self-reports in the area of intergroup contact. We note, however, that there is some risk that close others are not only likely to be knowledgeable, but also may be similar and motivated to (mis-) represent their friends positively. They could thus be subject to the same biases as self-reports. We suggest, however, that this risk can be mitigated by using multiple observers (see below), and that Kenny’s (1994) ‘round robin’ design provides a unique methodology for using observer ratings to validate self-reports of contact.

Judgments made by single observers, although they may not share the same systematic biases as self-reports, will still, however, contain unreliability, no matter how well acquainted the observer is to the target. To overcome this limitation, multiple observers can be used, and if they show consensus in their judgments, then that part of their judgments that is consensual becomes the criterion for validating self-reports (Kenny, 1994). Kolar et al. (1996), for example, found that aggregated personality judgments of two acquaintances were more predictive of behavioral criteria than either self-judgments or judgments made by single acquaintances. The ‘round robin’ design (Kenny, 1994) employs groups of participants whose members act as targets as well as judges of each other person in their group. Kenny and Albright (1987) suggested that research should focus on employing participants as both judges and targets in this manner, and the method has been used to great effect in personality and social psychology research since (e.g., De Paulo, Kenny, Hoover, Webb, & Oliver, 1987; Kwan, Gosling, & John, 2008; Robins, Mendelsohn, Connell, & Kwan, 2004; Vazire, 2010).

Kenny and Albright (1987) also suggested that research should focus on when and how, rather than whether, observer-judgements agree with self-judgements (for examples see Flink & Park, 1991; Park & Flink, 1989) and our first study focused on a moderator of self-other agreement that we thought would be especially important for research on intergroup contact.

Vazire's (2010) self-other knowledge asymmetry (SOKA) model hypothesised, and found, that self-reports should be more accurate (as judged against a criterion) than other-reports for traits that are low in *observability*. We contend that the moderating factor of observability leads to the prediction that observer-reports will validate self-reports of direct, but not extended, contact. Thus, if I have a close friend who belongs to a different ethnic group from mine (direct contact), that form of contact is typically viewed by others, and my close friends are likely to know about this relationship (i.e., it has high observability). If, however, one of my same-race friends or a family member has a friend from a different race, this extended cross-group friendship may well not be known to my friends (i.e., it has low observability). We therefore hypothesised that self-other agreement would validate direct intergroup contact, but not extended contact.

To summarize, it is clear that the validity of self-reports of intergroup contact is an issue that needs to be systematically addressed, and the comparison of self- and observer-reports provides a useful research approach for its investigation. The primary goals of this research were: (1) to examine whether knowledgeable others ('raters') made consensual judgments of targets' type, degree and quality of intergroup contact; (2) whether these consensual judgments were correlated with the targets' own self-ratings; and (3) whether self-other agreement was moderated by the observability of contact. Accomplishing these goals would provide some evidence for the validity of self-reports of intergroup contact. Additionally, by measuring self- and observer-reports of the personality dimension of extraversion, we sought to demonstrate discriminant validity. Thus, we sought to rule out the possibility that when

observers were judging a target's intergroup contact they were not simply judging how outgoing the target was, a judgment that past research has suggested can be made consensually, accurately, and also quickly (Borkenau & Liebler, 1995; Borkenau et al., 2004; Carney et al., 2007; Funder, 1999; Levesque & Kenny, 1993). Finally, we were also interested in whether observers showed consensus in judgments of targets' outgroup attitudes that also agreed with self-reports, and whether reports of intergroup contact, be they self-reports or consensual observer judgments, were related to reports of intergroup attitudes, again be they self-reports or consensual observer judgments. Again, we sought to rule out an alternative explanation to that of observer-reports validating self-reports of contact, whereby observers were able to judge what the target's attitude was, and then infer from the target's attitude his/her level of outgroup contact. Our first study was designed to meet these goals.

### Study 1

Our first study investigated self-other agreement in reports of intergroup contact in a manner that could control for perceptions of both the target's own attitudes and his/her extraversion. To do this, we used a round-robin design, in which each member of four-person groups rated themselves and all other individuals. Judgments were made not only on contact, but also on outgroup attitudes, and extraversion.

The round robin design (Kenny, 1994) involved recruiting close acquaintances in groups of four to rate themselves and each other member of the group on judgmental dimensions of outgroup attitudes, type, quantity and quality of intergroup contact, and a variety of other measures (including, extraversion). This design provides a means of exploiting Kenny's (1994) Social Relations Model (SRM) to decompose judgment data into several components. As Kwan et al. (2008) explain, Kenny's SRM is analogous to a 2-way analysis of variance (ANOVA) design. The first main effect is termed, the 'rater (or

perceiver) effect', which may be understood conceptually as whether the rater generally rated the other group members as higher or lower than the grand mean (the average rating of all targets by all raters). A high (low) rater effect implies a general tendency to rate the targets above (below) average on a particular dimension (see De Paulo et al., 1987). The second main effect is termed, the 'target effect', which may be understood conceptually as whether the target is generally rated by the group members as higher (or lower) than the grand mean. There is also a third effect, which Kenny (1994) termed, the 'relationship effect.' Just as in the 2-way ANOVA, it can be thought of as the interaction between rater and target. It reflects the unique perspective that a given rater has of a particular target, and captures the manner in which each rater judges each specific target, in a way that cannot be explained by the general (group-level) rater and target effects.

Study 1 not only asked whether there was consensus in observers' judgment of a target's level of contact, and whether observers agreed with target's self-judgments, but it also considered observability of contact as a potential moderator of self-other agreement. We predicted that self-observer agreement would be greater in the case of direct, than extended, contact. We first conducted a pilot study, to obtain ratings from independent judges concerning the observability of the various aspects on which self and observers would later be asked to report, and then we conducted a study to test these hypotheses.

#### *Pilot Study*

We collected data on the relative observability of the personality, attitudes and behaviour of well-known others. For all these questions, and for the single outgroup in Study 1, and one of the two outgroups in Study 2, participants were informed: "For the following questions, please note that 'Asians' refers to people living in the UK who are of South-West Asian origin (Indian, Pakistani, or Bangladeshi), rather than East-Asian origin (e.g., Chinese, Japanese, Korean)." South-West Asians constitute the largest minority group in the United

Kingdom, where the term ‘Asian’ is widely used to refer only to them (as opposed to East Asians, as in the United States).

We asked a sample of respondents to rate the observability of their friends’ (1) direct contact (friendships with) Asians; (2) extended contact (ratings of contact between their friends’ friends and Asians); (3) attitudes towards Asians; and (4) extraversion. We hypothesised that extraversion, which has shown considerable observer-consensus in other work (e.g., Borkenau & Liebler, 1995; Borkenau et al., 2004; Funder, 1999), would be judged most observable. We also expected direct cross-group friendship to be judged quite observable, since when one sees a friend interacting with another, the ethnicity of that other is visible. We expected that attitudes would be judged less observable, as prejudice is a sensitive subject that may not be frequently discussed. Most importantly, we predicted that extended cross-group friendship would be judged least observable, as one would not always know about the friends of one’s friends, or be able to observe them interacting.

*Method.* Seventy-nine white students at a British university (17 male and 62 female, between the ages of 18 and 21) completed a questionnaire in exchange for course credit. They rated how observable (from 1, *unobservable*, to 7, *observable*) the following four items were: their friends’ (1) extraversion, (2) attitudes to Asians, (3) direct contact (friendships with) Asians; and (4) extended contact (ratings of contact between their friends’ friends and Asians). A preceding paragraph explained exactly what was meant by ‘observable’; participants were asked to consider how easily they could provide a judgment that would adequately reflect reality.

*Results and Discussion.* A one-way repeated measures analysis of variance (ANOVA), with four levels, yielded a significant main effect,  $F(3, 76) = 109.36, p < .001$ . All pairwise comparisons were significant,  $p < .01$  (with Bonferroni adjustment), except the difference between extraversion and direct cross-group friends. As expected extraversion ( $M = 5.62, SD$

= 0.85) and direct contact with Asians ( $M = 5.53$ ,  $SD = 1.24$ ), were rated most observable, followed by attitude to Asians ( $M = 4.72$ ,  $SD = 1.39$ ); also as expected, extended contact with Asians was rated significantly the least observable ( $M = 2.57$ ,  $SD = 1.33$ ). These results provide a basis for testing, in the main study, the hypothesis that observability will moderate the degree to which observer-reports validate self-reports of contact. On the basis of judged observability, we expect self-other agreement on extraversion, direct contact, and ethnic attitude, but much less agreement for extended contact.

### *Method*

#### *Participants*

One hundred and twenty-eight students at a British university (90 females, 37 males, one missing response; mean age 21 years) took part in the research. They comprised 32 groups of four (11 all-female; 1 all-male; 20 mixed), group members being acquainted with each other for at least four months (mean acquaintance time, 19.09 months,  $SD = 8.29$ ). Although we did not explicitly exclude Asian participants, we did not include their data in the study. One group had to be removed from the analysis due to missing data, and the fact that one of the group members was of Asian origin. The final sample was 31 groups, 124 participants (86 females, 37 males and one missing response).

#### *Procedure*

Thirty-two students fulfilled a research requirement by each recruiting three close friends and bringing the group of four people to the lab at a prearranged time. Recruits were required to be white English speakers who were not studying psychology. Group members were assigned a letter (A, B, C or D) and asked to remember or take note of each other's letter assignment, so that they could refer to each other by letter. Group members filled in the questionnaire either in separate cubicles or in different parts of a large room, without

conferring. Each group member first answered questions about him- or herself and then about each of the other three group members, in a 'round robin' design (Kenny, 1994).

### *Measures*

Participants completed a questionnaire consisting of measures of contact with Asians (quantity, quality, and direct and extended cross-group friendship), attitude towards Asians, and extraversion. Demographic questions were presented first with an introductory paragraph. The order of the contact measures and the attitude measures was counterbalanced (no order effects were detected).

We measured quantity of contact by two items (inter-item correlation,  $r = .81$ ): 'How often do you meet Asians in your everyday life at university?' and 'How often do you talk with people who are Asians?' Answers were given on seven-point scales labelled with 'Never', 'Rarely', 'Sometimes', 'Quite a bit', 'Often', 'Very often' and 'All the time'. Quality of contact was measured with two seven-point scales anchored by the adjectives 'cooperative – competitive' and 'superficial – intimate', presented following the stem question 'Are your interactions with Asians generally: . . .?'. With the 'cooperative – competitive' item reversed, higher scores on the two-item measure (inter-item correlation  $r = .31$ ,  $p < .001$ ) indicated higher quality of contact.

Direct cross-group friendship was measured with two items (inter-item correlation,  $r = .38$ ,  $p < .001$ ): 'About how many of your friends are Asian?' (with answers given on a seven-point scale labelled with 'None', 'A few', 'Less than half', 'About half', 'More than half', 'Most' and 'All'), and 'How often do you spend time with friends who are Asians?' (answers on the same seven-point scale as that presented with the contact quantity items).<sup>1</sup>

Extended cross-group friendship was measured with two items (inter-item correlation,  $r = .52$ ,  $p < .001$ ); 'About how many of your closest White friends have friends who are



Asian?’ and ‘About how many people in your immediate family have friends who are Asian?’ Answers were given on the previously described ‘None - All’ seven-point scales.

Outgroup attitudes were measured using a ‘feeling thermometer’ (participants had to report their feeling towards Asians in general on a 100-point scale, ranging from 0°, *cold*, to 100°, *warm*) and a set of five bi-polar adjectives on seven-point scales preceded by the instruction: ‘Based on your experience rate the extent to which you have each of the following feelings about Asians in general. (Please circle one number on each scale).’ Bi-polar adjectives were ‘*Negative - Positive*’, ‘*Friendly - Hostile*’, ‘*Suspicious - Trusting*’, ‘*Respect - Contempt*’ and ‘*Admiration - Disgust*’. These five adjectives were taken from Wright et al.’s (1997) six-item scale; we did not include the sixth item, ‘*Warm – Cold*’, as it was made redundant by the feeling thermometer. After transforming the feeling thermometer scores to put them on the same scale as the other attitude scores (by dividing all scores by 100 and multiplying by seven) and reversing ‘*Friendly - Hostile*’, ‘*Respect - Contempt*’ and ‘*Admiration - Disgust*’ items, higher scores on the seven-item measure (Cronbach’s  $\alpha = .90$ ) indicated more positive attitudes towards Asians.

Extraversion was measured using the four-item extraversion measure (7-point bi-polar scales anchored at ‘*Restrained - Candid*’, ‘*Active - Passive*’, ‘*Silent - Talkative*’, ‘*Aloof - Gregarious*’) from the NEO-Five Factor Inventory (Costa & McCrae, 1988) as well as an additional single seven-point scale anchored at ‘*Extraverted*’ and ‘*Introverted*’. After reversing the ‘*Extraverted - Introverted*’ and ‘*Active - Passive*’ measures, the composite scale yielded a reliable index ( $\alpha = .83$ ) on which higher scores reflected higher levels of extraversion.

After these ‘self-ratings’, participants were presented with the following paragraph: ‘Now we would like to know about your friends. The rest of this booklet contains questions similar to those you just answered, but they will relate to the people who accompanied you to

the lab today -- one set of questions for each person.' Participants were then given the 'self-rating' questions again, three times, but this time the questions were reworded so that they asked the participant to rate each of the other group members in turn. Participant A received questions about B then C then D, Participant B received questions about A then C then D, and so forth. As the letters were assigned in an arbitrary fashion, this ensured that group members did not rate each other in any meaningful order, for instance rating their closest friend first.

### *Results and Discussion*

Participants rated themselves and the other targets in their group on the following composite variables: contact quantity, contact quality, cross-group friendship, extended friendship, outgroup attitude and extraversion. In Table 1 we present the mean rating of the self and the mean rating given to others in one's group on each of these variables. Using the group as the unit of analysis, we tested whether these mean ratings differed between self and the other targets. As indicated in Table 1, on average raters said they had more contact with Asians than they said other group members had, they said they had more cross-group friendships than others, they reported significantly higher extended friendships for themselves than they attributed to others, and they saw others as significantly more extraverted than they judged themselves to be.

We conducted further analyses of these rating data from each group using the Social Relations Model (SOREMO; Kenny, 1994, 1998; Kenny & La Voie, 1984; Malloy & Kenny, 1986) to analyze the round-robin data structure. The input for this program consists of matrices of ratings on each variable for each group. These matrices are structured, with each row indicating the person in the group who gave the ratings (i.e., the rater, or in SOREMO parlance the actor) and the columns indicating the person who is rated (the target, or what SOREMO refers to as the 'partner'). The diagonal elements in each matrix consist of the

rater's self-rating on the variable in question and are analyzed separately from the basic SOREMO analysis. In essence, for each such matrix, SOREMO partitions the variance in the rater by target ratings into three components: the rater effect, the target effect, and the relationship (or rater by target interaction) effect.

We conducted a SOREMO analysis for each of the rating variables on which we have just reported the means (contact quantity, contact quality, cross-group friendship, extended friendship, outgroup attitude, and extraversion). Table 2 presents the relative variance in the matrices of ratings attributable to rater, target, and relationship effects, computed across the 31 groups with complete data. As relative variances, these values sum to 1.00 for each rating dimension; thus each value indicates the average percentage of variance in each matrix of ratings attributable to each source. Also indicated is whether each component of variance differs significantly from zero, except in the case of the relationship variance component, which cannot be tested since it also includes error variance. These effects are tested across the 31 groups. All of the ratings that concern the outgroup (contact quantity and quality, cross-group friendship, extended friendship, and outgroup attitude) manifest considerable rater variance, thus suggesting large individual differences in ratings attributable to the rater. Importantly, however, each of these variables also shows evidence of reliable target variance, meaning that there are consensual differences among the targets in how they are rated, with some in each group consistently seen as higher on these variables than others. In other words, raters manifest some degree of agreement with each other on how they rate others in the group on these outgroup judgments. Interestingly, extraversion, the only rating that did not focus on the outgroup, shows considerable target variance, indeed more than rater variance. This replicates other work (e.g., Borkenau & Liebler, 1995; Borkenau et al., 2004; Funder, 1999) that has shown considerable consensus in how others are judged on the extraversion factor of the Big Five personality inventory.

Given that all participants rated themselves on these variables, in addition to rating the others in their group, SOREMO provides (and tests) the correlations between the self-ratings on each variable and both the rater and target effects. These self-correlations are given in Table 3. The first column of correlations in this table reports the extent to which raters who judge others as high or low on average on a judgmental dimension also judge themselves as high or low on that dimension. Unsurprisingly, in the case of all judgmental dimensions there are significant correlations between self and other ratings (e.g., the higher a rater rates him- or herself, the higher he or she rates others on average). This pattern of results is consistent with a projection or assumed similarity hypothesis, whereby a perceiver uses him/herself as an anchor, and estimates or judges others with reference to the self (additionally, people tend to, sometimes erroneously, impute their own knowledge to others; see Nickerson, 1999). We use the term ‘assumed similarity’, which in Kenny’s Social Relations Model is assessed by the correlation between self-perception and the rater effect.<sup>2</sup>

In terms of the motivation that guided this research, we are most interested in the correlations in the second column of Table 3, between self-ratings on each dimension and how one is judged by others. Having established that there exists significant, albeit modest, consensus within the groups in how targets are judged on the outgroup-relevant judgments, the question these correlations address is whether self-ratings on those judgmental dimensions are correlated with these target effects. Thus, for instance, is it the case that targets who report that they have more outgroup contact are also seen consensually by their friends as having more outgroup contact? If such correlations exist, then they would suggest the validity of the self-ratings, in so far as they agree with how the self is consensually rated by others.

As the correlations in the second column of Table 3 reveal, there are substantial and significant self – target effect correlations in the case of all of the rating variables, with the

exception of the extended friendship variable. In others words, if a rater says he or she has more outgroup contact, reports higher quality contact, reports more outgroup friends, or reports more positive outgroup attitudes, then other close friends tend to consensually agree with those self-ratings. On the other hand, there is no evidence that friends agree with self-ratings in terms of extended friendship, i.e., “my family and friends have outgroup friends.” This is consistent with our prediction, that observability of contact would moderate self-other agreement. We further confirmed this difference by comparing the self-target correlations for direct (contact quantity,  $r = .55$ ) and extended ( $r = -.06$ ) contact, which were significantly different, Chi-square (1) = 4.66,  $p < .05$  (testing the difference between correlated correlations, see Olkin & Finn, 1990). This result is not so surprising given that others would have to observe, or at least know about, the contact not of the target but of the friends and family of the target to have such knowledge. The correlations also reveal considerable self – other agreement in ratings of extraversion: the more one rates oneself as extraverted, the more other close friends tend to consensually agree in these assessments.

Before concluding, however, that these data on self-other agreement provide incontrovertible evidence of the validity of self-reports of direct contact, we should rule out two alternative explanations. First, given that a target’s self-rating of contact with the outgroup is substantially correlated with the consensual judgment of that target’s contact by his or her friends, one might wonder whether this correlation might be produced simply by highly correlated judgments of a target’s style of interrelating with others more generally, not specifically focused on outgroup members. In other words, given the substantial self – target effect correlation in extraversion, it could potentially be the case that the substantial self – target effect correlation in contact is entirely due to general appraisals by others (and by oneself) of extraversion generally, and not by appraisals more specifically of intergroup contact. To rule out this interpretation, we conducted further partial correlation analyses,

examining the partial correlation between self ratings and target effects for each outgroup variable, partialing out self ratings on general extraversion as well as consensual target effects on extraversion (i.e., how others raters judge a target). These partial correlations are shown in column 3 of Table 3. In the case of all of the outgroup rating variables, with the exception of the extended contact variable, the self – target effect correlations remained substantially positive and significant even when controlling for general appraisals of self-judged extraversion and friends' consensual appraisals of target extraversion. In other words, there is substantial agreement between how an individual judges his or her own contact and interactions with the outgroup and how others consensually see that person's outgroup contact and interactions, even when controlling for consensual agreement in generalized extraversion of the individual (both in his or her self-judgments and in how he or she is seen by others).

A second, alternative explanation for our findings is the possibility that raters may have used knowledge about the target's outgroup attitudes in order to infer the amount of contact the target might have. In other words, rather than actually *know* a target's amount of contact, one might have a general feeling about their outgroup attitudes and *infer* levels of contact from that. Further analysis, however, finds little evidence for this alternative hypothesis. Column 4 of Table 3 reports partial correlations between self contact variables and target contact, while controlling for target's perceived outgroup attitude and target's self-rated attitude; correlations remain significant for two out of three measures of direct contact (contact quantity, and direct cross-group friendship, but not for contact quality). These partial correlations suggest, however, that raters may, to some extent, rely on the target's perceived outgroup attitudes to rate his/her *quality* of intergroup contact, which is a more subjective and less observable measure. Along with the findings when controlling for extraversion, these findings lend further credence to our claim that the presence of significant target

variance in these outgroup judgments and their significant correlations with self-ratings on these dimensions suggest that such self-ratings show considerable validity, especially for more objective measures of direct contact. Over and above any generalized appraisals of how outgoing someone is and is seen to be, or what one's outgroup attitudes are or are seen to be, the more contact someone reports with outgroup members, the more others who are close friends consensually agree in that assessment.

### Study 2

The results of Study 1 provide clear support for the use of self-reports in research on direct intergroup contact. However, since these are the first such results reported, we conducted a second study in which we sought to replicate the principal results of the first study and additionally provide a further demonstration of the discriminant validity of self-other agreement, by showing agreement when contact and attitude items referred to the same outgroup ('matched'), but not when they referred to different outgroups ('unmatched'). Study 2 closely followed Study 1, with three main additions and changes.

First, we sought to obtain further evidence of discriminant validity. In Study 1, we saw that there was discriminant validity between the Asian contact variable and general extraversion. That is, while others agreed with self-judgments of both contact and extraversion, it was not the case that the agreement on contact was due to agreement on extraversion. In Study 2, we examined an additional and more refined issue of discriminant validity, namely whether participants' ratings discriminate between two discrete outgroups. Specifically, are self-reports and target judgments of contact with (and attitudes towards) Asians distinct from self-reports and target judgments of contact with (and attitudes towards) Gay men? We expect that if self-reports of contact show validity, they should be more highly correlated with target judgments of contact for the same outgroup than between different

outgroups. We would have greater faith in our proposed method of validating self-reports of contact if they showed discriminant validity in this way.

Second, Study 1 included a variety of measures of outgroup contact, and, as predicted, we had found no evidence for self-other agreement on the relatively unobservable measure of extended contact. In Study 2 we therefore focused exclusively on direct cross-group friendships as the measure of contact, and asked about this form of contact at university and in participants' hometowns to obtain a more reliable measure of direct contact.

Third, we counterbalanced the order of self- and observer-reports. In Study 1 we always had participants complete self-reports first, because we thought that this would be easier for them. The limitation of this fixed order, however, is that it may have inflated the extent to which participants projected their ratings of others from the self. The extent to which people perceive others as similar to themselves increases when they focus on themselves during interaction with another (Cartwright, 1957), and completing a self-report first would arguably cause participants to focus on themselves while completing ratings of others (Schwarz & Sudman, 1992). If Study 2 detects no order effect, and if correlations are still high between self-ratings and own ratings of others, irrespective of order, this would be more likely to indicate actual similarity between rater and target, rather than assumed similarity.

Finally, we removed the 'extraversion' measure, which was no longer necessary as Study 1 had compared self-other agreement on contact measures with this baseline, and confirmed that evidence of self-other agreement on direct contact was not due to perceived extraversion.

### *Method*

#### *Participants*

One hundred and thirty-six students at a British university (88 females and 48 males; mean age 20 years) took part in the research. They comprised 34 groups of four (14 all-



female; 6 all-male; 14 mixed), group members being acquainted with each other for at least six months (mean acquaintance length of 20 months,  $SD = 9.54$ ). Although we did not explicitly exclude Asian participants, there were none; and because we asked participants to bring heterosexual friends, there were no homosexual participants.

### *Procedure*

We followed the same procedure as in Study 1.

### *Measures*

The same questionnaire was used as in Study 1, with some changes and additions. The contact items were included twice, once with reference to contact at university, and once for contact in the participant's hometown. To keep the questionnaire of manageable length, questions about attitudes and contact with Asians included only a feeling thermometer, and a measure of direct cross-group friendship, at home and at university, and for each outgroup (Asians, as in Study 1; and Gays, with 'homosexual men' specified). In fact, the two contact measures were highly correlated; for Asians,  $r(32) = .76, p < .01$  for rater effects, and  $r(32) = .51, p < .01$  for target effects, and for Gays:  $r(32) = .62, p < .01$  for rater effects, and  $r(32) = .48, p < .01$  for target effects (degrees of freedom were computed in both cases as the number of effective groups minus 2 = 32). We therefore collapsed over contact location in subsequent analyses.

### *Results and Discussion*

No order effects were detected as a result of the counterbalancing of the contact and attitude measures, nor of the self- and other-ratings. For all four measures of interest, we also compared the correlations involving target and self, for participants who answered self-first ( $r_s =$ : Asian contact: .39; Asian attitude: .44; Gay contact: .56; Gay attitude: .25) and self-last

( $r_s =$ : Asian contact: .34; Asian attitude: .47; Gay contact: .50; Gay attitude: .22); differences were negligible.

We start the main analysis again by comparing mean self-ratings with how others are perceived on average. These means and tests of the differences between them are reported in Table 4. There was no difference in how respondents judged the amount of contact (as cross-group friends) they themselves had with Asians, compared with the amount of contact others were judged to have had with Asians. It is hard to compare these results directly with Study 1, because the measure of contact has changed. In Study 1 self was judged to have more Asian friends than others were, and to have more contact in general; however, there was no rated difference in the quality of contact experienced by self and others. As in Study 1, there was no difference in how respondents rated their own warmth towards Asians, compared with the amount of warmth others were judged to feel towards Asians. On the other hand, respondents' self-reported contact with Gays was substantially lower than the amount of contact that was attributed to others. That is, respondents generally said that others had substantially more contact with Gays than they indicated for themselves. At the same time, interestingly, raters rated their own warmth towards Gays as significantly higher than the warmth that they attributed to others.

The relative variances attributable to raters, targets, and relationship are shown in Table 5. As in Study 1, there was considerable rater variance, or individual differences in the way in which others are rated, particularly for the warmth thermometer ratings. In other words, some raters consistently rated targets higher than did other raters, and this was more true on the thermometer ratings than on the ratings of contact. Importantly, however, on all variables there was also significant variance attributable to the target, meaning that there was consensus across raters within groups as to who was higher on a particular variable and who was lower. Target variance in the case of contact and liking for Gays was clearly more

substantial than in the case of target variance for Asians. Thus, there was considerable consensus within these groups of four close friends about who has more contact with Gays and who feels warmer towards them.

Given that there was reliable target variance for all of the contact and liking variables, both for Asian and Gay outgroups, we next investigated whether one's own reports of contact and liking were correlated with how others perceived oneself. That is, if I see myself as having high contact with members of an outgroup, do others in my group of friends see me as having relatively high contact? The self-correlations, both with the rater effect and the target effect, are reported in Table 6.

All of the self-rater correlations in the first column of Table 6 are large and significant. As in Study 1, this is consistent with a large assumed similarity or projection effect: If I judge myself as high on liking or contact for one of the outgroups, then I judge others in my group of friends as high on liking or contact for that outgroup. The self-target correlations (column two) are also positive and most are significant. In the case of contact, both self-target correlations are significant. Observers' ratings of a target's contact with both outgroups agree with self-ratings of contact with those groups. For liking, the self-target correlation is substantial and significant for the Asian thermometer variable, whereas it is smaller and non-significant for the Gay thermometer variable; the more a person reports feeling warmly towards Asians, the more others concur in this judgment, but this is less true in the case of expressed warmth of feeling towards Gays. As in Study 1, we computed partial correlations to determine whether judges were using what they judged targets' attitudes to be in order to infer the amount of contact targets might have. As shown in column 3 of Table 6, the resulting partial correlation was significant in the case of contact with Gays (.45) but became non-significant in the case of Asians (.20).

Turning next to our additional and more refined investigation of discriminant validity, we consider whether self-reports and target judgments of contact with one outgroup (e.g., Asians) are the same as those for another, quite different outgroup (e.g., Gays). In Table 7 we present the correlations among the target effects and self-ratings for four different variables: contact with Asians, Asian thermometer, contact with Gays, and Gay thermometer. The target effects represent how a target is judged consensually by others in the group. The matrix of correlations is divided into three sections. Correlations that lie above the boxed set are correlations of target effects with each other. Correlations to the right of the boxed set are correlations of self-ratings with each other. Correlations inside the box are correlations of target effects with self ratings.

Looking first at the correlations between the target effects (above the box), it is clear that if one is consensually judged as having high contact with a group, then one is consensually judged to have positive attitudes towards that same group ( $r = .82$  for Asians;  $r = .67$  for Gays). On the other hand, all other correlations among these target effects are non-significant, and three of the four of them are negative. In other words, individuals who are consensually seen as having high contact and positive attitudes towards one of the two groups are *not* seen as having high contact and positive attitudes towards the other group.

If we look at the correlations among the self-judgments (to the right of the box), the correlation between reported contact towards the two groups is positive and significant, but quite low ( $r = .23$ ). On the other hand, the correlation between the two thermometer measures is very large and highly significant ( $r = .67$ ). Thus, liking towards the two outgroups is redundant in the self-ratings, in a way that it is not in the target effects. This difference is surely in part due to the large rater effects on these thermometer measures, noted above: if I give high thermometer ratings for myself towards one outgroup, I give high thermometer ratings for myself for other outgroups, and also for others in general.

Looking at the correlations between the target effects, on the one hand, and the self-ratings on the other (i.e., those in the box), the correlations on the diagonal are those reported earlier – are self-judgments on a variable congruent with how targets are consensually judged by others on that variable? Importantly, these self-target correlations are typically much larger than the correlations of self and target when the outgroup is mismatched. For instance, others consensually agree with targets in the target's self-rating of contact with Asians ( $r = .36$ ) and they consensually agree with targets in the target's self-rating of contact with Gays ( $r = .53$ ). But when we mismatch the outgroups between self-ratings and target effects, the correlations are not significant (and are negative in one of the two cases). That is, a person's self-rating of contact with Asians is not reliably correlated with how observers judge that target's contact with Gays ( $r = -.08$ ). And one's self-rating of contact with Gays is not reliably correlated with how observers judge one's contact with Asians ( $r = .10$ ). So the self-ratings agree with the target effects only when the same outgroup is referenced.

Thus, overall, Study 2 confirms the finding that direct contact ratings are valid, as indicated by self-other agreement. Moreover, at least in the case of one of the outgroups, that was novel to this study, this agreement is again not due to inferences based on perceived attitudes. Given that the resulting partial correlations were significant for the Asian outgroup in Study 1, the overall picture is that self-other agreement is in all likelihood not due simply to inferences based on perceived attitudes.

Finally, this study confirmed that there is discriminant validity to our results as a function of the specific outgroup examined: self-other correlations were consistently higher when the ratings were on matching outgroups rather than mismatching ones. This finding suggests that agreement of others with self on contact is specific to an outgroup, and does not simply reflect some generalized appraisal of who is more or less in touch with outgroups in general.

### General Discussion

These two studies constitute an original contribution to the research literature on intergroup contact by demonstrating for the first time the validity of self-report measures of intergroup contact. Using a round-robin design, with four-person groups of friends, we were able to show, as predicted, significant agreement between participants' self-reports and observers' ratings for direct, but not extended, contact. We also demonstrated the discriminant validity of our findings by completely ruling out the possibility that observers' ratings of another person's contact were based on the target's extraversion, and partially ruling out the possibility that observers' ratings of the target's contact were based on their ability to judge his or her ethnic attitudes, and then infer contact from this. Moreover, Study 2 also showed that participants' ratings discriminated between two discrete outgroups. In providing evidence for the validity of self-reports of direct contact, these studies remove a source of criticism of many of the studies comprising the large database of research on intergroup contact. We discuss the conclusions and implications of this research in terms of the utility and limitations of self-reports of contact, and their discriminant validity.

#### *Utility and Limitations of Self-reports of Contact*

Replicating previous research in the judgment of personality traits (e.g., for measures of extraversion; see Borkenau & Liebler, 1995; Borkenau et al., 2004; Funder, 1999), we found in both studies that raters made consensual judgments of targets' degree and quality of direct intergroup contact, and that these consensual judgments were correlated with the targets' own self ratings. In Study 1 we also replicated the results of previous research using a measure of extraversion; extraversion appears to be the most visible, easily-rated attribute, as it manifested the greatest target variance, the smallest assumed similarity, and the highest self-target correlation.

We also reported compelling evidence of reliable target variance, meaning that there were consensual differences among the targets in how they were rated by observers. Thus raters manifest some degree of agreement with each other on how they rate others in the group. As expected, extraversion showed considerable target variance, but target effects were significant, albeit smaller, for all measures of contact, especially cross-group friendship and contact quality, and for outgroup attitude.

Substantial self – target effect correlations substantiated these findings, evident on all rating variables except extended friendship. Thus, when a rater reported that he or she had more direct outgroup contact, higher quality contact, more outgroup friends, or more positive outgroup attitudes, then other close friends tended to consensually agree with those self-ratings. As predicted, however, self-other agreement was moderated by the observability of different types of contact. The self-target effect correlation was not significant for extended contact (Study 1), whereby others would have to observe, or at least have knowledge of, the contact not of the target but of the friends and family of the target to be in receipt of this information. These findings confirm the potential value, but also limitations, of observer reports of contact as a tool for validating self-reports of contact; validity was confirmed for direct, but not extended, contact.

Notwithstanding the general support we provide for self-reports of intergroup contact, there remain two potential limitations with this method of assessing contact, namely socially desirable responding, and assumed similarity. We found only partial evidence of socially desirable response bias. In Study 1, raters reported having significantly more contact with Asians, and more direct and extended cross-group friendship than their observers in their group. In Study 2 respondents reported more positive attitudes towards Gay men, but actually reported lower contact with Gays than they thought their friends had. These results do not suggest a strong influence of social desirability; nonetheless given that people have a

tendency to rate themselves more favourably than they rate others (Dunning, 2005), this remains a threat to the validity of self-reports of intergroup contact.<sup>3</sup> There is ample room for future research using observer-reports to validate self-reports of contact, and some of this should try to demonstrate whether participants actually under-estimate others' contact rather than (or as well as) over-estimating their own. This could be done, as in much personality research, by using an objective behavioral criterion for more sensitive tests of accuracy.

Turning to assumed similarity, we found that participants did indeed rate others as similar to themselves to a large extent. In Study 1, for all judgmental dimensions there were significant correlations between self – other ratings. It could be argued that this assumed similarity weakens observer-reports as a method for evaluating self-reports. However, as SOREMO calculates the variance due solely to the target and unaffected by the rater, and as this variance was found to be significantly different from zero, we argue that assumed similarity effects do not question our conclusions. In any case, assumed similarity may have been overestimated, as people tend to be friends with similar others (e.g., Daniels & Plomin, 1985; Kiesner, Maass, Cadinu & Vallese, 2003), and ratings of others are therefore necessarily correlated with self-ratings.

To reduce any impact of self-anchoring we counterbalanced the order of self- and other-ratings in Study 2, but we found no effect of this order variable, either on means or on self-target correlations. Assumed similarity occurs most frequently when a person focuses on themselves during contact with another (Cartwright, 1957), and answering questions about oneself first would be likely to increase this effect (Schwarz & Sudman, 1992). We therefore suggest that this similarity between ratings was due more to actual similarity between the friends than to assumed similarity. In Studies 1 and 2, self-rater correlations were large and significant, consistent with a large assumed similarity and/or scale usage effect (i.e., the judges use the scale differently both in judging themselves and others).



*Discriminant validity*

We also presented two types of evidence for discriminant validity. First, in Study 1, we measured self-reports and observer-reports of extraversion, and demonstrated that when observers judged a target's intergroup contact they were not simply judging how outgoing the target was, a judgment that past research has suggested can be made consensually, quickly and accurately (e.g., Borkenau et al., 2004; Carney et al., 2007; Levesque & Kenny, 1993). Of course, it is plausible that more extraverted people (people who like being around others in general) would be judged higher on ratings of contact with outgroup members (in particular). We were able to exclude this alternative explanation for the agreement between targets' self-ratings of contact and consensual judgment by their friends in terms of targets' levels of extraversion. Correlations between self ratings and target effects remained reliable when partialing out self-ratings on general extraversion as well as consensual target effects on extraversion. Analogously, it might be argued that apparent self-other agreement on contact is achieved via perceptions of others' racial attitudes, and inferring from them targets' likely levels of contact. In supplementary analyses we controlled for perceived attitudes, but here our evidence was not as unequivocal. In Study 1, partial correlations between self contact variables and target contact remained significant for the two more objective measures of direct contact (contact quantity, and direct cross-group friendship), but not for the more subjective (less observable) measure of contact quality. It may be that raters do, to some extent, infer the more subjective *quality* of a target's outgroup contact based on his/her perceived outgroup attitudes. In Study 2 we found that the same partial correlations (controlling for outgroup attitude) remained significant in the case of contact with Gays, but became non-significant in the case of Asians. Thus, overall, there is evidence that observer reports validate self-reports of direct contact, and that this does not in general rely on inferring outgroup attitudes.

Second, in Study 2, we examined an additional and more refined issue of discriminant validity, namely whether participants' ratings discriminate between two discrete outgroups, Asians and Gays. Specifically, we asked, are self-reports and target judgments of contact with Asians and contact with Gays the same? We compared the correlations among the target effects and self-ratings for four different variables, based on contact and attitude measures for the Asian outgroup, and contact and attitude measures for the Gay outgroup. As should be the case, if measures have discriminant validity, contact and attitude measures were correlated within, but not between, outgroups. Thus, if an individual is seen consensually as having high contact with and positive attitudes towards one of the two groups, this does *not* mean that he or she is seen as having high contact with and positive attitudes towards the other group. Further evidence of discriminant validity comes from the correlations between target effects and self-ratings. These correlations assess, for example, the congruence between self-judgments on a variable and how one is consensually judged by others on that variable (e.g., if I report high contact with Asians, do others also report that I have high contact with Asians?). These correlations are significant only when there is a match between the outgroup for the two ratings, but not when there is a mismatch. For example, self-rated contact with Asians (Gays) is *not* significantly correlated with own target effect on contact with Gays (Asians).

To conclude, the results of these two studies confirm our innovative attempt to validate self-reports of intergroup contact by observer reports. We have reported largely consistent results across two studies, and two outgroups, showing that observer reports validate self-reports of direct contact, which we had earlier shown to be rated as quite observable. Extended contact, in contrast, was rated reliably less observable, and self-reports of extended contact could not be validated with observer-reports. Our research also yielded two types of evidence for discriminant validity, showing that consensus across observers about targets is

not simply based on judged extraversion, nor generally on inferred outgroup attitudes. Finally, measures of contact and attitude are only meaningfully related within, but not between, different outgroups. Overall, our research makes novel empirical and theoretical contributions. Empirically, we demonstrated the utility of observer reports, used in the context of a round-robin design, to validate self-reports of direct contact, and we provided two types of evidence of discriminant validity. Theoretically, we investigated not simply whether observer-reports validated self-reports, but when they do so, and we showed that observability of the contact was a key moderator. In sum, these findings increase the confidence with which conclusions can be drawn from the large literature on intergroup contact that relies on self-report measures of direct contact.

ACCEPTED

## References

- Allport, G. (1954). *The nature of prejudice*. Reading: Addison-Wesley.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261-272.
- Becker, G., & Bakal, D. (1970). Subject anonymity and motivational distortion in self-report data. *Journal of Clinical Psychology, 26*, 207–209.
- Bernieri, F. J., Zuckerman, M., Koestner, R., & Rosenthal, R. (1994). Measuring person perception accuracy: Another look at self-other agreement. *Personality and Social Psychology Bulletin, 20*, 367-378.
- Biesanz, J. C., West, S. G., & Graziano, W. G. (1998). Moderators of self-other agreement: Reconsidering temporal stability in personality. *Journal of Personality and Social Psychology, 75*, 467–477.
- Biesanz, J. C., West, S. G. & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self–other agreement in judgments of personality. *Journal of Personality and Social Psychology, 92*, 119-135.
- Biesanz, J. C., & West, S. G. (2000). Personality coherence: Moderating self-other profile agreement and profile consensus. *Journal of Personality and Social Psychology, 79*, 425–437.
- Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality, 63*, 1-25.
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behaviour as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*, 599-614.

- Borkenau, P., & Ostendorf, F. (1992). Social desirability as moderator and suppressor variable. *European Journal of Personality*, *6*, 199-214.
- Boruch, R. F., & Cecil, J. S. (1979). *Assuring the confidentiality of social research data*. Philadelphia, PA : University of Pennsylvania Press.
- Bradburn, N.M. (1983). Response effects. In P. Rossi, J. Wright, & A. Anderson (Eds.), *Handbook of survey research* (pp. 289-328). New York: Academic Press.
- Campbell, D. T., Kruskal, W., & Wallace, W. (1966). Seating aggregation as an index of attitude. *Sociometry*, *29*, 1-15.
- Campbell, J. D., & Fehr, B. (1990). Self-esteem and perceptions of conveyed impressions: Is negative affectivity associated with greater realism? *Journal of Personality and Social Psychology*, *58*, 122-133.
- Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, *41*, 1054-1072.
- Cartwright, D. (1957). Social psychology and group processes. *Annual Review of Psychology*, *8*, 211-236.
- Clack, B., Dixon, J. A. & Tredoux, C. (2005). Eating together apart: Patterns of segregation in a multiethnic cafeteria. *Journal of Community and Applied Social Psychology*, *15*, 1-16 .
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, *15*, 110-117.
- Costa, P. T., Jr., & McCrae, R. R. (1988). Personality in adulthood: A 6-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology*, *54*, 853-863.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82, 359-378.

Cummings, S. & Lambert, T. (1997). Anti-Hispanic and anti-Asian sentiments among African Americans, *Social Science Quarterly*, 78, 338-353.

Daniels, D., & Plomin, R. (1985). Differential experiences of siblings in the same family. *Developmental Psychology*, 21, 747-760.

De Paulo, B. M., Kenny, D. A., Hoover, C. W., Webb, W., & Oliver, P. V. (1987). Accuracy of person perception: Do people know what kinds of impressions they convey? *Journal of Personality and Social Psychology*, 52, 303–315.

Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological Measurement*, 23, 699-720.

Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.

Fiedler, E. R., Oltmanns, T. F., & Turkheimer, E. (2004). Traits associated with personality disorders and adjustment to military life: Predictive validity of self and peer reports. *Military Medicine*, 169, 207–211.

Flink, C., & Park, B. (1991). Increasing consensus in trait judgments through outcome dependency. *Journal of Experimental Social Psychology*, 27, 453-467.

Fossett, M.A. & Kiecolt, K.J. (1989). The relative size of minority populations and white racial attitudes. *Social Science Quarterly*, 70, 820-835.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

Funder, D. C., & Colvin, C. R. (1997). Congruence of others' and self-judgments of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 617-648). San Diego, CA: Academic Press.

Giles, M. W. (1977). Percent black and racial hostility: an old assumption reexamined, *Social Science Quarterly*, *58*, 412-417.

Gotlib, I. H., & Meltzer, S. J. (1987). Depression and the perception of social skill in dyadic interaction. *Cognitive Therapy and Research*, *11*, 41-54.

Hood, M.V., & Morris, I.L. (1997). Amigo o enemigo?: Context, attitudes, and Anglo public opinion toward immigration. *Social Science Quarterly*, *78*, 309-323.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*, 521-551.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford.

Kenny, D.A. (1998), *SOREMO Version V.2*. Unpublished manuscript, University of Connecticut, Storrs, CT.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, *102*, 390-402.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. (Vol. 18, pp. 142-182). Orlando, FL: Academic Press.

Kiesner, J., Maass, A., Cadinu, M., & Vallese, I. (2003). Risk factors for ethnic prejudice during early adolescence. *Social Development*, *12*, 288-308.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64*, 311-337.

Krahé, B. (2007). Aggression. In M. Hewstone, W. Stroebe & K. Jonas (Eds). *Introduction to social psychology: A European perspective* (4th Edition). Oxford, UK: Blackwell.

Kwan, V. S. Y., Gosling, S. D., & John, O. P. (2008). Anthropomorphism as a special case of social perception: A cross-species social relations model analysis of humans and dogs. *Social Cognition, 26*, 129-142.

Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality. *Psychological Bulletin, 35*, 659.

Levesque, M. J., & Kenny, D. A. (1993). Accuracy of behavioural predictions at zero-acquaintance: A social relations analysis. *Journal of Personality and Social Psychology, 65*, 1178-1187.

Malloy, T. E., & Kenny, D. A. (1986). The social relations model: An integrative method for the study of personality. *Journal of Personality, 54*, 101-124.

McCauley, C., Plummer, M., Moskaleiko, S., & Mordkoff, J. T. (2001). The exposure index: A measure of intergroup contact. *Peace and Conflict: Journal of Peace Psychology, 7*, 321-336.

Norman, W.T., & Goldberg, L.R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology 4*, 681-691.

Olkin, I, & Finn, J. (1990). Testing correlated correlations. *Psychological Bulletin, 108*, 3330-3333.



Oltmanns, T. F., Gleason, M. E. J., Klonsky, E. D., & Turkheimer, E. (2005). Meta-perception for pathological personality traits: Do we know when others think that we are difficult? *Consciousness and Cognition: An International Journal*, *14*, 739–751.

Ones, D. S., Viswesvaran, C., & Reiss A. D. (1996). The role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660-679.

Park, B., & Flink, C. (1989). A social relations analysis of agreement in liking judgments. *Journal of Personality and Social Psychology*, *56*, 506-518.

Paulhus, D. L. (1984). Two-component models of socially desirable responding . *Journal of Personality and Social Psychology*, *46*, 598-609.

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire* (pp. 143-165). New York: Springer-Verlag.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & Wrightsman, L. S. (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*, 1025-1060.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York, NY: Guilford Press.

Pettigrew, T., & Tropp, L. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751-783.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78*, 582-593.

Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 649-679). San Diego, CA: Academic Press.

Schofield, J. W. (1979). The impact of positively structured contact on intergroup behavior. *Social Psychology Quarterly, 42*, 280-284.

Schofield, J. W., & Sagar, H. A. (1977). Peer interaction patterns in an integrated middle school. *Sociometry, 40*, 130-138.

Schwarz, N., & Sudman, S. (Eds.). (1992). *Context effects in social and psychological research*. New York: Springer Verlag.

Sedikides, C., & Strube, M. J. (1995). The multiply motivated self. *Personality and Social Psychology Bulletin, 21*, 1330-1335.

Sigall, F.L., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology, 18*, 247-255.

Turner, R. N., Hewstone, M., & Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology, 93*, 369-388.

Turner, R. N., Hewstone, M., Voci, A., Paolini, S., and Christ, O. (2007). Reducing prejudice via direct and extended cross-group friendship. In W. Stroebe and M. Hewstone (Eds.), *European review of social psychology* (Vol. 18, pp. 212-255). Hove, E. Sussex: Psychology Press.

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality, 40*, 472-481.

- Vazire, S. (2010). Who knows what about a person? The Self-Other Knowledge Asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281-300.
- Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*, 1202-1216.
- Vonofakou, C. (2006). *The role of intergroup anxiety in intergroup contact*. Unpublished Doctoral Dissertation, University of Oxford.
- Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41*, 221-229.
- Wagner, U., Hewstone, M., & Machleit, U. (1989). Contact and prejudice between Germans and Turks: A correlational study. *Human Relations, 42*, 561-574.
- Wilder, D. A. (1984). Intergroup contact: The typical member and the exception to the rule. *Journal of Experimental Social Psychology, 20*, 177-194.
- Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology, 73*, 73-90.

## Endnotes

1. The inter-item correlations for both quality of contact and direct cross-group friendship were rather low; we sought to improve this by forming a composite of the two variables, but its reliability was also quite low ( $\alpha = .56$ ). We therefore retained the three distinct measures of direct contact (quality, quantity, and direct friends). It should be noted that in the case of each of the measures with rather low inter-item correlations (quality of contact and direct cross-group friendship), the effects reported in the SOREMO analysis below look the same if each of the two items used to assess each construct is treated as an individual item. There is reliable target variance in each item, and the self-target correlations are significant.
2. We note that an alternative, psychologically less-interesting account for these data could be proposed in terms of scale-usage, i.e., that judges have individual differences in how they define the endpoints of the judgment scale, and these influence how they judge themselves and others. Given the extensive literature on projection effects in research of this type, however, we suggest that projection is the more likely account of these data.
3. Although we did not include a measure of SDR in this research, we have collected other data that did measure both direct and extended contact, and SDR (using the 40-item Biased Inventory of Socially Desirable Responding, BIDR, which consists of two SD constructs: impression management, IM, and self-deceptive enhancement, SDE; Paulhus, 1991). In a sample of white students ( $N=131$ ), correlations were all small and non-significant between both subscales of the BIDR and self-reported direct and extended contact with Asians (see Vonofakou, 2006, Study 5). Thus, overall, we do not think socially desirable responding is likely to have had a substantive impact on our data.

Table 1

*Mean Ratings of Self and Others (Study 1)*

Variable	Self Mean	Other Mean	<i>t</i> -test
Contact quantity	4.38	4.18	2.62*
Contact quality	5.25	5.18	.95
Cross-group friendship	3.15	2.95	3.05*
Extended friendship	4.31	3.63	7.25*
Outgroup attitude	5.31	5.36	-1.03
Extraversion	4.75	4.91	-2.45*

\* Significant mean difference,  $p < .05$

Table 2

*Relative Variance Partitioning – SOREMO analysis (Study 1)*

Variable	Rater	Target	Relationship
Contact quantity	.68*	.09*	.23
Contact quality	.50*	.16*	.34
Cross-group friendship	.47*	.22*	.30
Extended friendship	.79*	.04*	.17
Outgroup attitude	.61*	.18*	.21
Extraversion	.06*	.57*	.37

\* Variance component significantly different from zero,  $p < .05$

Table 3

*Self – Rater and Self – Target Correlations (Study 1)*

Variable	Rater	Target	Target (controlling for extraversion)	Target (controlling for attitude)
Contact quantity	.76*	.55*	.36*	.39*
Contact quality	.78*	.31*	.22*	-.04
Cross-group friendship	.40*	.53*	.38*	.33*
Extended friendship	.75*	-.06	-.06	-.11
Outgroup attitude	.79*	.28*	.23*	--
Extraversion	.63*	.72*	--	--

- Correlation significantly different from zero,  $p < .05$

Table 4

*Mean ratings of Self and Others (Study 2)*

Variable	Self Mean	Other Mean	<i>t</i> -test
Contact with Asians	1.46	1.47	.10
Thermometer towards Asians	78.13	77.96	.20
Contact with Gays	1.34	2.31	23.18*
Thermometer towards Gays	81.74	78.34	3.38*

\* Significant mean difference,  $p < .05$

ACCEPTED



Table 5

*Relative Variance Partitioning – SOREMO analysis (Study 2)*

Variable	Rater	Target	Relationship
Contact with Asians	.54*	.13*	.33
Thermometer towards Asians	.69*	.07*	.24
Contact with Gays	.27*	.41*	.32
Thermometer towards Gays	.54*	.17*	.29

\* Variance component significantly different from zero,  $p < .05$

ACCEPTED

Table 6

*Self-Rater and Self-Target Correlations (Study 2)*

Variable	Rater	Target	Target (controlling for attitude)
Contact with Asians	.73*	.36*	.20
Thermometer towards Asians	.91*	.46*	--
Contact with Gays	.62*	.53*	.45*
Thermometer towards Gays	.81*	.23	--

\* Correlation significantly different from zero,  $p < .05$

Table 7

*Correlations Among Target Effects and Self Ratings – Asian and Gay Contact and Liking**(Study 2)*

		Target Effects				Self Ratings		
		Asian Contact	Asian Therm.	Gay Contact	Gay Therm.	Asian Contact	Asian Therm.	Gay Contact
Target Effects	Asian Contact							
	Asian Therm.	.82*						
	Gay Contact	-.11	-.22					
	Gay Therm.	-.14	.20	.67*				
Self ratings	Asian Contact	.36*	.20	-.08	.11			
	Asian Therm.	.38*	.46*	-.34*	.00	.21*		
	Gay Contact	.10	.10	.53*	.30*	.23*	-.07	
	Gay Therm.	.12	.08	-.06	.23	.10	.67*	.15

\* Correlation significantly different from zero,  $p < .05$