Who do we think of as good judges? Those who agree with us about us
Jinseok S. Chun a,∗, Daniel R. Ames a, Jose N. Uribe b, E. Tory Higgins a
a Columbia University, United States
b University of Michigan, United States

ARTICLE INFO

Article history:
Received 2 February 2016
Revised 12 July 2016
Accepted 27 July 2016
Available online 28 July 2016

Keywords:
Perception
Agreement
Self-verification
Shared reality

ABSTRACT

The present research considered what leads perceivers to evaluate someone as a good or poor judge of people. In general, we found a substantial role for agreement: perceivers evaluated another person as a good judge when he or she agreed with their perception of someone's characteristics. Importantly, the effect of agreement depended on who this "someone" was. We found that perceivers' evaluation of another individual as a good judge was more heavily shaped by agreement about their own characteristics than by agreement about a third-party target's characteristics. This effect emerged across a range of studies and research designs, including multi-rater evaluations among developing business professionals, experimentally controlled settings, and a survey in which US adults reported on existing relationships. Moderation analyses suggested that the effect of agreement was particularly strong in situations where the agreement could more effectively satisfy perceivers' motives to (a) feel relational connectedness and (b) verify the accuracy of their perception.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

While answers to the question of "who is a good judge" have been accumulating (e.g., Ambady, Hallahan, & Rosenthal, 1995; Bruner & Tagiuri, 1954; Letzring, 2008; Rees, Rothman, Lehavy, & Sanchez-Burks, 2013; Vazire, 2010), a related question appears to have received little attention: Who do people think is a good judge? In the present paper, we address this question. Our research focuses on good-judge evaluations, which we define as the extent to which a focal person is believed by someone else to accurately assess people's characteristics such as personalities, mental states, abilities, and attitudes. Research has found that such good-judge evaluations have meaningful real-world effects, including shaping behaviors surrounding relationship formation and development, as well as interpersonal exchange and collaboration (Ashford & Tsui, 1991; Hofmann, Lei, & Grant, 2009).

The extent to which people evaluate another person as a good judge likely has many sources. We focus here on how agreement in social perception shapes good-judge evaluations. Research has shown that people are motivated to (a) experience relational connectedness with others and (b) feel confident in their perception of the reality. Importantly, it has been demonstrated that reaching agreement with others in social perception can satisfy both of these motives (Echterhoff, Higgins, & Levine, 2009; Hardin & Higgins, 1996). We predict that the satisfaction of these two motives will in turn lead people to perceive another person who is agreeing with them (i.e., who is satisfying these motives) in a positive light, believing that the person is a competent judge of people. In short, we expect that people's good-judge evaluations will be positively affected by the extent to which the judge agrees with their perception of a target individual's characteristics.

Not all kinds of agreement are equal, however. Previous research has differentiated between the self and a third party as the bases of agreement and discussed their distinct implications (Blackman & Funder, 1998; Kenny & West, 2010). Accordingly, we distinguish between two different types of agreement in social perception: agreement-about-self and agreement-about-third-party (see Fig. 1). Agreement-about-self is the convergence between a perceiver's perception of him- or herself and the judge's perception of the perceiver (Swann, 2011). Agreement-about-third-party is the convergence between a perceiver's perception of a third-party target and the judge's perception of that same target (Echterhoff et al., 2009).

While we expect that both kinds of agreement in social perception often matter to good-judge evaluations, we predict that agreement-about-self will matter more than agreement-about-third-party with the following motivational reasons. First, feelings of connectedness that rise from agreement should be greater when the shared opinion is about a significant target (Echterhoff, Higgins, & Groll, 2005). Because the self is a target of ultimate significance for many people, the positive effect of agreement on good-judge evaluations may also be especially strong when the basis of the agreement is the self. Second, people tend to assume that they know themselves well (Pronin, Kruger, Savitsky, & Ross, 2001). Because they have stronger confidence in the knowledge about themselves compared with a third-party target,
people might react more strongly to agreement about themselves (Arkes, Boehm, & Xu, 1991). If another person disagrees with people's perception of themselves (whom they think they know very well), they may regard the judge's view as a challenge to their own competence as accurate perceivers, reacting negatively as a result. For these reasons, we expect the effect of agreement-about-self to be greater than agreement-about-third-party.

An alternative possibility is that people would actually put less weight on agreement-about-self when evaluating judges because they assume that others cannot really perceive their personal characteristics accurately due to their intricate and complex nature (Haslam, Bain, Douge, Lee, & Bastian, 2005). In other words, they might generally assume that other people cannot actually "know them" and, as such, discredit the validity of strong convergence between their self-view and a judge's perception of them. As a result, in evaluating good judges, people might discount the informational value of agreement-about-self, recognizing their own privileged perspective. Instead, they might place more emphasis on agreement-about-third-party because in such cases they and a judge might have potentially equal access to "the truth."

While acknowledging this alternative, we nonetheless expected that people would react more strongly to agreement-about-self than agreement-about-third party because self-perception is a central and principal reality they experience (Leary, 2007). Given its significance, people may want to feel connected to others and verify the accuracy of their perception particularly regarding this reality (Swann, 2011). This leads to our central prediction about which kind of agreement more heavily influences good-judge evaluations: We expected to find stronger overall links between good-judge evaluations and agreement about the self than agreement about a third party.

Beyond this basic effect, we also sought to examine when particular kinds of agreement might matter more or less. Earlier, we argued that agreement would affect good-judge evaluations because it gives individuals feelings of relational connectedness and confidence in their perception. We thus expected that the effect of agreement would depend on the extent to which these two motives can be satisfied by the agreement. First, we predicted that the effect of agreement might become stronger if the person who is agreeing with them is someone who is relationally close to them (Echterhoff et al., 2005) because he or she is in a position to satisfy their relational needs more effectively. In contrast, agreement might have a weaker effect when it is with a person they are not close to, because the person does not and cannot strongly satisfy their relational motives. Therefore, we expected that liking for the judge would increase the effect of agreement.

Second, we predicted that the effect of agreement-about-third-party would become stronger if people think that they have rich and accurate knowledge about that third-party target. When people think that they have fairly reliable appraisals of a third-party target (i.e., they are highly familiar with the target), they will have more confidence in their perception of him or her (Beaupré & Hess, 2006), and such confidence would make it especially important that the judge verifies their perception of the target. Therefore, the extent to which people think they know a third-party target (i.e., familiarity) might moderate the relationship between agreement-about-third-party and good-judge evaluations.

We believe that our account and results are relevant for several traditions of research. In particular, our work draws on and contributes to research on self-verification theory (Swann, 2011) and shared-reality theory (Echterhoff et al., 2009). Both self-verification and shared-reality phenomena are based on agreement in social perception. However, these lines of thinking have generally not been integrated in prior research: Self-verification scholarship has examined agreement-about-self whereas empirical work on shared-reality has focused mainly on agreement-about-third-party. The present research bridges these two areas, yielding empirical results and an account that shed light on the joint operation of self-verification and shared-reality.

1.1. Overview of studies

We conducted four studies testing the (relative) effects of agreement-about-self and agreement-about-third-party on good-judge evaluations. Study 1 used multi-rater evaluation data drawn from a large population of MBA students. Studies 2 and 3 were conducted in experimental settings, confirming the causality of the effects found in Study 1 and addressing alternative explanations. Study 4 used a survey format to gather adults' reports on existing acquaintances, testing our predictions about moderating conditions. We operationalized agreement as actual convergence in social perception (i.e., objective agreement) in Studies 1, 2, and 3 whereas it was measured as people's perception of convergence (i.e., subjective agreement) in Study 4. We report all measures, manipulations, and exclusions in these studies.

2. Study 1

Study 1 employed a large multi-rater evaluation dataset collected over a period of seven years in a Master of Business Administration
(MBA) program located on the East Coast of the United States. Participants completed the evaluations in the first two months of their MBA program. Accordingly, they were in relatively early stages of relationship development and were interested in learning about each other's characteristics, whereas the emotional bonds among them had not developed significantly. This situation provided an ideal context in which we could test whether people drew on their perception of agreement when they were evaluating each other in terms of how good judges they were.

2.1. Method

2.1.1. Sample

The data were collected from 4,894 MBA students. The profile of the MBA classes included in the dataset suggested that about 35% of them were females and their average age was 28 at the point of their participation.

2.1.2. Data

Participants (a) evaluated themselves in multiple domains of their characteristics, (b) invited peer students to evaluate them in the same domains, and (c) evaluated peer students who invited them in the same domains. The evaluation was done online and it took approximately two weeks. Importantly, participants did not have access to the contents of others' evaluations of them until the entire process ended. Each evaluation consisted of two parts. The first part captured Big Five personality domains (i.e., extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience). The second part captured competencies in various domains (e.g., teamwork abilities, decision-making skills), including good-judge evaluations. The specific details are discussed in the Measures section below. The original dataset consists of 4,894 self-evaluations and 26,386 peer evaluations (i.e., each participant received 5.4 peer evaluations on average). After removing incomplete responses, the final dataset had 8,302 cases for agreement-about-self and 18,079 cases for agreement-about-third-party.

2.1.3. Measures

Every item used a 7-point Likert scale. Perception in the domains of Big Five personality was measured using the Ten Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003) with a scale ranging from 1 (“Disagree strongly”) to 7 (“Agree strongly”). The TIPI measures a person’s perception of a target in the Big Five domains using two items for each domain.

Good-judge evaluations should reflect people's evaluation of a focal judge's ability to accurately perceive others’ characteristics in general. Here, “characteristics” can span various domains such as personality, mental states, competencies, and attitudes and perspectives. Therefore, good-judge evaluations were measured by four items that capture accurate judgment in these domains with a scale ranging from 1 (“Never”) to 7 (“Always”): “S/he misjudges people’s personality and character (reverse-coded)”; “S/he is good at sensing what other people are thinking and feeling”; “S/he is good at assessing other people’s strengths and weaknesses”; and “S/he is able to empathize and understand someone else’s perspective.” These four items demonstrated reasonable internal consistency (α = 0.80).

The agreement scores were generated in the following manner. We first aggregated the values from the TIPI items (after reverse-coding) in each domain of the Big Five personality. Then we created the agreement-about-self and agreement-about-third-party scores: We (1) calculated the absolute discrepancy scores between a perceiver’s self-perception and a judge’s perception of the perceiver (for agreement-about-self) and between a perceiver’s and a judge’s perceptions of a same third-party target (for agreement-about-third-party) in each domain of the Big Five personality, (2) calculated the average discrepancy score across the five domains, and (3) reversed the score so that higher scores reflected greater agreement.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Good-judge evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement</td>
<td>-0.11 (0.02)</td>
</tr>
<tr>
<td>Raw evaluation: extraversion</td>
<td>-0.01 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: agreeableness</td>
<td>0.07 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: conscientiousness</td>
<td>0.07 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: neuroticism</td>
<td>-0.00 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: openness to experience</td>
<td>0.05 (0.01)</td>
</tr>
</tbody>
</table>

Note: Level-1 n = 8,302, level-2 n = 3,775. Values in parentheses are standard errors.

2.2. Results

Most judges in the dataset provided and received evaluations from multiple people. Therefore, we created multilevel datasets with the people who were evaluating the judges at level 1 and the judges at level 2 (thus taking into account the individual characteristics of different judges). Random-intercept models were run using the software program R with the lme4 and lmerTest packages (Bates et al., 2015; Kuznetsova, Brockhoff, & Christensen, 2013). The raw evaluations judges provided in the Big Five domains of the self and a third party were included as fixed-effect control variables.

Good-judge evaluations (M = 5.41, SD = 0.92) were positively related to both agreement-about-self (M = −1.19, SD = 0.49; see Table 1), γ = 0.11, p < 0.001, and agreement-about-third-party (M = −1.09, SD = 0.49; see Table 2), γ = 0.06, p < 0.001. Thus, agreement in general had a positive relation to good-judge evaluations.1 To test whether the effect of agreement-about-self was larger than the agreement-about-third-party, we built a model that examined how the effect of agreement depended on whether it was about the self or a third party. We created a dataset that contained (mean-centered) agreement as a single variable and a dummy variable denoting whether the agreement was about the self (−0.5) or a third party (0.5). In addition, we created an interaction term between the two variables. If the effect of agreement-about-self on good-judge evaluations is greater than the effect of agreement-about-third-party, the coefficient of this interaction term should be negative and statistically significant. In testing this relationship, the slopes of agreement, third-party dummy, and their interaction term were modeled as random effects along with the intercept. Again, the raw-evaluation variables were included as fixed-effect control variables.

As predicted, agreement had a significant positive effect on good-judge-evaluations, γ = 0.06, p < 0.001, and the interaction term of agreement and third-party dummy had a negative and significant effect, γ = −0.06, p = 0.002, indicating that agreement-about-self had a stronger relation to good-judge evaluations than agreement-about-third-party.

1 It is possible that judges’ individual-level characteristics and perceived participant-judge similarity had significant implications on how participants evaluated the judges. We therefore examined these possibilities. Specifically, we tested how extraversion, agreeableness, and conscientiousness of the judges (average perception they received from other people) were related to good-judge evaluations they received. The results suggested judges’ extraversion (γ = 0.05, p < 0.001), agreeableness (γ = 0.41, p < 0.001), and conscientiousness (γ = 0.11, p < 0.001) were significantly related to good-judge evaluations. More importantly, the positive relationships that good-judge evaluations had with agreement-about-self (γ = 0.10, p < 0.001) and agreement-about-third-party (γ = 0.04, p < 0.001) remained significant after controlling for these characteristics of judges.

We further examined whether the assumed similarity between participants and the judges (i.e., the similarity between how participants perceived themselves and the judges in the Big Five domains) was significantly related to the good-judge evaluations the judges received from participants. The assumed similarity between participants and the judges across the Big Five traits was positively related to good-judge evaluations (γ = 0.05, p < 0.001). More importantly, the positive relationships that good-judge evaluations had with agreement-about-self (γ = 0.07, p < 0.001) and agreement-about-third-party (γ = 0.04, p < 0.001) remained significant after controlling for the assumed similarity.

Table 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Good-judge evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw evaluation: extraversion</td>
<td>-0.01 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: agreeableness</td>
<td>0.07 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: conscientiousness</td>
<td>0.07 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: neuroticism</td>
<td>-0.00 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: openness to experience</td>
<td>0.05 (0.01)</td>
</tr>
</tbody>
</table>
Table 2
Multilevel modeling of the relationship between agreement-about-third-party and good-judge evaluations from Study 1.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Good-judge evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>γ (p-Value)</td>
</tr>
<tr>
<td>Agreement</td>
<td>0.06 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: extraversion</td>
<td>0.01 (0.00)</td>
</tr>
<tr>
<td>Raw evaluation: agreeableness</td>
<td>-0.01 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: conscientiousness</td>
<td>-0.00 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: neuroticism</td>
<td>0.00 (0.01)</td>
</tr>
<tr>
<td>Raw evaluation: openness to experience</td>
<td>0.00 (0.01)</td>
</tr>
</tbody>
</table>

Note: Level-1 n = 18,355, level-2 n = 3,732. Values in parentheses are standard errors.

2.3. Discussion

Using multi-rater evaluation data, we found that both agreement-about-self and agreement-about-third-party had positive relationships with good-judge evaluations, and the former had a stronger relationship than the latter. The results (Table 1) also suggested that people thought of those as good judges who rated them highly in the domains of agreeableness, conscientiousness, and openness to experience. This implies that people generally considered those who perceived them favorably as good judges. However, given that the relationships between agreement and good-judge evaluations were observed controlling for the judges’ raw evaluations of the targets in the Big Five personality domains, whether people regarded someone as a good judge did not appear to be simply about how the judges evaluated the targets. It was also related to the similarity of perception, especially when the self was the target.

As we argued above, these findings suggest agreement may have led people to evaluate the other person to be a good or poor judge of people. However, given the correlational nature of the results, the other direction of causality (i.e., people first discerned who was a good judge and then altered their views of targets) could not be ruled out. Moreover, the effect sizes were relatively small. Finally, some alternative explanations exist for the findings. One is that the positive relationship between agreement and good-judge evaluations could have been caused by a confounding variable: interpersonal closeness. As two people become close, their perceptions might become more similar (McNulty & Swann, 1994) and, because they are close, they might view one another more positively (Festinger, Schachter, & Back, 1950). Another explanation regarding the difference between the two types of agreement concerns the availability of information (Wilson & Dunn, 2004): Agreement-about-third-party might matter as much as agreement-about-self but participants in Study 1 might not have had clear evidence regarding another person’s perception of a third person. When the information regarding agreement-about-third-party is not available to them, it is unlikely to affect their evaluations. These questions about causality, and the alternatives concerning closeness and evidence availability, shaped our approach to Study 2.

3. Study 2

In Study 2, we experimentally manipulated participants’ apparent levels of agreement with another previously unknown person. By capturing participants’ views of this judge, we attempted to rule out the effect of interpersonal closeness. We also made information about agreement clearly available to participants in order to establish that the difference between the effects of agreement-about-self and agreement-about-third-party was not caused by the availability of information.

3.1. Method

3.1.1. Sample

The target sample size was 120. We calculated it assuming a medium-to-large effect size ($d = 0.65$) with 0.05 significance level and 0.80 power. We recruited 118 participants from a university located on the East Coast in exchange for $12. The mean age of the participants was 23.7 ($SD = 5.7$) and 62% were females. Participants identified their race with the following frequencies: 40% Asian, 33% White, 11% Black, 5% Latino(a), and 8% two or more races. Five of them did not report their race.

3.1.2. Study design

Participants made a video-recording in which they recalled and described their interactions with a person they recently met. The recording was made under the premise that this video would be shared with another participant in the next room, although in reality no such person existed. The gender of this hypothetical fellow participant was matched to the gender of participants. Therefore, although “he or she” and “his or her” are used in the explanation below, participants actually heard only “he” and “his” or “she” and “her.”

After participants provided video accounts of their interactions with a recent acquaintance, they received information regarding the hypothetical fellow participant’s perception of them and the acquaintance they had described. In this information, we manipulated agreement-about-self and agreement-about-third-party to each participant. There were three conditions with different levels of agreement-about-self and agreement-about-third-party: low-low (low agreement), high-low (high agreement-about-self), and low-high (high agreement-about-third-party) conditions.

3.1.3. Procedure

Participants came to the lab individually and were escorted to one of two small rooms located in a larger room. Upon arrival, participants filled out a questionnaire containing the TIPI items capturing their self-perception, starting with “I think I am.” Once they finished the questionnaire, the experimenter came back to the room and said, “This experiment investigates interactional dynamics between two people. One participant will be asked to describe personal interactions he or she had with a new acquaintance, and this will be videotaped using an iPad. The other participant will then watch this video and share his or her impressions about the interactions.” Participants were given their unique participant number and were assigned to the role of a describer in an ostensibly random manner by a computer program. They then took a moment to decide on which interactions with which person to describe. Once they made the decisions, they filled out questionnaires that contained the TIPI items capturing their perception of the interaction partner they had identified. Participants were advised not to use the interaction partner’s real names but to use initials to refer to him or her. The TIPI questionnaire started with “I think ______ is” and participants wrote the initials in the blank. After the questionnaire, participants took three to 5 min to prepare the description about their interactions, and then used the iPad to record a three- to five-minute description of the interactions. When they were done, the experimenter came back to pick up the iPad. Then participants worked on a filler Remote Association Task (RAT) for 4 min. Once participants finished RAT, the experimenter came back with the questionnaires allegedly filled out by the fellow participant in the other room. As explained below, this worked as the manipulation. After the manipulation, participants filled out a final questionnaire capturing their perception of the other participant using the good-judge questions. Finally, participants were debriefed. During the debriefing session, participants were asked how their and the fellow participant’s perceptions were different from each other. After the experimenter checked that participants’ responses corresponded to the experimental conditions, he provided explanations regarding the research questions and the study procedures.

3.1.4. Measures

Every survey item used a 7-point Likert scale ranging from 1 (“not at all”) to 7 (“very much”). As explained above, the TIPI items (Gosling et al., 2003) were used to capture participants’ perception of themselves.
and their interaction partner. Good-judge questions about the other participant were measured using the identical items to those in Study 1 (α = 0.79).²

3.1.5. Manipulation
As explained above, participants filled out questionnaires capturing their perception of themselves and their interaction partner using the TIPI items before they described their interactions. Using these questionnaires, the experimenter created two fake questionnaires (which were allegedly filled out by the other participant) to manipulate the levels of agreement-about-self and agreement-about-third-party. The fake questionnaires started with “I think ________ is.” The experimenter first put “Participant N” (participant number of the participant) and the initials of the interaction partner in “________” of the two questionnaires respectively. Participants’ answers on the TIPI items were then entered into a spreadsheet that generated specific values of the fake questionnaires. The low agreement sheet was set to generate the values with total 21 points of absolute discrepancy from the input values across the 10 questions, whereas the high agreement sheet was set to generate the values with total 7 points of absolute discrepancy. While the total discrepancy was fixed within each condition, the specific discrepancies were randomly distributed across 10 items. The experimenter then copied these values to the fake questionnaires.

Along with the fake questionnaires, participants received the original TIPI questionnaires they had filled out (regarding the self and interaction partner) and were told that the other participant had watched the video and shared his or her impressions about the interactions. Participants were asked to copy the answers from the fake questionnaires (ostensibly completed by their fellow participant) onto the original questionnaires they filled out using a pen with a different color. This copying ensured that the information regarding agreement-about-self and agreement-about-third-party was clearly available to participants. Whether the questionnaire on the self or the interaction partner was put on the left side of participants was counterbalanced and did not affect the results reported below.

3.2. Results
Good-judge evaluations from participants (M = 3.94, SD = 0.99) were submitted to a one-way ANOVA with the three experimental conditions. The model was statistically significant, F(2,114) = 7.53, p < 0.001, adjusted R² = 0.10. The results from planned contrast analyses suggested that the average good-judge evaluations from the high agreement-about-self (M = 4.37, SD = 0.98, n = 40) and high agreement-about-third-party (M = 3.91; SD = 0.82, n = 36) conditions were higher than the low agreement condition (M = 3.56, SD = 0.98, n = 42), t(114) = 3.21, p < 0.001, d = 0.62. Thus, as expected, there was an overall positive effect of agreement on good-judge evaluations. Moreover, the good-judge evaluations from the high agreement-about-self condition were higher than the high agreement-about-third-party condition, t(114) = 2.11, p = 0.037, d = 0.39 (see Fig. 2).

Thus, as expected, agreement-about-self had a stronger positive effect on good-judge evaluations than agreement-about-third-party. These relationships held when the hypothetical participant’s raw evaluations were included as control variables.³

³ A list of the variables that were measured in Study 2 can be found in Supplemental materials.
³ These results demonstrated how the degree of discrepancies (or agreement) between participants’ and the hypothetical participant’s perceptions affected good-judge evaluations. In exploratory analyses, we further examined whether the distribution of discrepancies across the five domains of personality affected good-judge evaluations. Specifically, we computed the standard deviations of the discrepancies (across the five domains) in perceptions of (a) the self and (b) a third-party target. Neither of the two indices significantly affected good-judge evaluations (t ≤ 1.2, ps > 0.25). Thus, the effects of agreement-about-self and agreement-about-third-party on good-judge evaluations appeared to be driven by the degree of discrepancies, not their distribution.

3.3. Discussion
The results from Study 2 were consistent with Study 1. A high level of agreement between two people’s social perception had a positive effect on good-judge evaluations. Moreover, as expected, the strength of this effect depended on the basis of agreement: The effect of agreement-about-self was stronger than agreement-about-third-party.

The experimental procedures of Study 2 precluded any possibility of direct social interactions between participants and the non-existent person whom they evaluated as a judge. Therefore, the positive effect of agreement cannot be explained by the interpersonal closeness between them. Study 2 also made information about agreement clearly available to participants. As explained above, during debriefing all participants reported that they were aware of how different their perception and the fellow participant’s perception were regarding their and the interaction partners’ characteristics. As a result, the pattern of findings revealed in Study 2 cannot be explained by the availability of information.

4. Study 3
Study 2 replicated the findings from Study 1, while addressing potential confounding effects of interpersonal closeness and information availability. In Study 3, we sought additional experimental evidence for the impact of agreement while addressing two key points. The first had to do with the nature and source of the information received for perception. The second had to with the domains of perception.

First, in Study 2 there was a difference in the roles played by participants and the third parties. That is, participants reported on their own interactions with the third parties. In other words, participants described themselves and the third parties. It is possible that this asymmetry (i.e., participants described themselves whereas third parties did not) influenced the effects of agreement on good-judge evaluations, potentially inflating the effect of agreement-about-third-party. Specifically, a judge with high agreement-about-third-party might have been viewed as not only having similar perception as participants on the third parties’ characters but also empathizing with participants’ experiences and perspectives. Study 3 addressed this potential confound by having participants and the (hypothetical) judge receive the same amount of information regarding the third party’s characteristics from the third party him- or herself.

Study 3 also used a more value-neutral domain of individual traits: people’s attitudes toward risk-taking behaviors. Using the Big Five personality traits in Studies 1 and 2 was effective in capturing people’s comprehensive perception of someone’s characteristics. However, some domains of the Big Five traits are evaluative in a sense that higher or lower ratings in those domains were considered to reflect more
favorable perception (particularly, agreeableness and conscientiousness). Accordingly, Study 3 employed people’s risk-taking attitudes as a less evaluative domain of person perception.

4.1. Method

4.1.1. Sample

The study was conducted across two consecutive days and the target sample size was 180. It was calculated based on the same assumptions as Study 2 and predicting 33% attrition rate on the second day. We recruited 180 US online participants via Amazon.com’s Mechanical Turk platform. Participants were paid one and two dollars for their participation in the first and the second days. The mean age of the participants was 34.4 (SD = 10.4) and 49% of them were females. The participants identified their race with the following frequencies: 73% White, 8% Asian, 7% Latino(a), 4% Black, and 6% two or more races. One participant did not report his or her race.

4.1.2. Study design

The study employed a two-stage design. On Day 1, participants described a recent challenge that they faced and read another person’s (i.e., third party’s) description of his or her challenge. The third party’s description was in fact a hypothetical story created for the present research. Participants also rated how likely they and the hypothetical participant were likely to engage in different types of risk-taking behaviors. On Day 2, participants were informed how a different participant (i.e., the judge) who had read both their and the third party’s challenges rated their likelihood to engage in the same types of risk-taking behaviors. Similar to Study 2, we manipulated the two types of agreement to each participant using this information. As a final step, participants rated the judge using the good-judge items. Again, Study 2, we manipulated the two types of agreement to each participant using this information. As a final step, participants rated their likelihood to engage in different types of risk-taking behaviors. This manipulation was counterbalanced. Participants were presented with the information on how Participant J rated the likelihood that they and Chris would engage in the six types of risk-taking behaviors. This information was presented along with the original ratings that participants themselves provided so that participants could observe any discrepancies. In order to strengthen the effect of our manipulation, participants were asked to look through the ratings and identify the behavior that led to the greatest discrepancy between their perception and Participant J’s perception (of themselves and Chris). They were further requested to report the actual degree of the greatest discrepancy. The order in which participants were presented with the ratings regarding themselves and Chris was counterbalanced.

Participant J’s responses were generated by a similar procedure as Study 2. Participants’ answers on the six DOSTPERT items were entered into a spreadsheet that generated specific values of the (hypothetical) responses of Participant J. The low agreement sheet was set to generate the values with total 12 points of absolute discrepancy from the input values across the questions, whereas the high agreement sheet was set to generate values with total four points of absolute discrepancy. As in Study 2, the total discrepancy was fixed within each condition and the specific discrepancies were randomly distributed across the six items.

These ratings were plotted on charts so participants could understand how their responses were similar to or different from those of Participant J. One chart was for agreement-about-self and the other was for agreement-about-third-party. As noted, the order in which charts were presented was counterbalanced.

4.1.3. Procedure

After participants agreed to participate in the study, they first reported their likelihood to engage in six types of risk-taking behaviors from the DOSTPERT scale (Blais & Weber, 2006). They were then asked to think about a challenge they recently encountered, either from their work life or personal life. They were asked to provide the specific details of their recent challenge (e.g., what happened, how they felt, and how they dealt with it) using about 10 sentences. To preserve anonymity, they were asked to use pseudonyms.

After participants described the challenges they recently experienced, they were informed that they would be asked to read someone else’s description of his or her challenge. After 5 s, they were presented with a description that “Participant 108” (code name Chris; the third party) provided regarding his or her recent challenge. The challenge involved a male boss who mistakenly overlooked the report Chris submitted and was getting angry with Chris as a result. Chris’ challenge was designed to be ambiguous regarding his or her risk propensity. After participants read the description of Chris’ challenge, they provided a short description of their impression of Chris. They then rated the likelihood that Chris would engage in the six types of risk-taking behaviors.

On Day 2, participants were told that a different participant, Participant J (the judge), had read both their challenges and the challenge reported by Chris. They were further informed that Participant J had reported impressions of them and Chris based on the challenges they described. On the next page, participants were presented with the information on how Participant J rated the likelihood that they and Chris would engage in the six types of risk-taking behaviors. As described below, these ratings served as the manipulation. Finally, participants rated Participant J using the good-judge items (as in Study 2).

4.1.4. Measures

Items from the DOSTPERT scale were used to capture participants’ perception of themselves and the third party (i.e., Chris). Participants answered six items using a 7-point Likert scale ranging from 1 (“extremely unlikely”) to 7 (“extremely likely”). The items included “Admitting that your (his or her) tastes are different from those of a friend,” “Disagreeing with an authority figure on a major issue,” “Choosing a career that you (he or she) truly enjoy(s) over a more secure one,” “Speaking your (his or her) mind about a controversial issue in a meeting with people you (he or she) don’t (doesn’t) know well,” “Walking home alone at night in an unsafe area of town,” and “Moving to a city far away from your (his or her) family and/or friends.” Good-judge evaluations were measured using the same approach as in Studies 1 and 2 (α = 0.84).

4.1.5. Manipulation

Participants were presented with the information on how Participant J rated the likelihood that they and Chris would engage in the six types of risk-taking behaviors. This information was presented along with the original ratings that participants themselves provided so that participants could observe any discrepancies. In order to strengthen the effect of our manipulation, participants were asked to look through the ratings and identify the behavior that led to the greatest discrepancy between their perception and Participant J’s perception (of themselves and Chris). They were further requested to report the actual degree of the greatest discrepancy. The order in which participants were presented with the ratings regarding themselves and Chris was counterbalanced.

Participant J’s responses were generated by a similar procedure as Study 2. Participants’ answers on the six DOSTPERT items were entered into a spreadsheet that generated specific values of the (hypothetical) responses of Participant J. The low agreement sheet was set to generate the values with total 12 points of absolute discrepancy from the input values across the questions, whereas the high agreement sheet was set to generate values with total four points of absolute discrepancy. As in Study 2, the total discrepancy was fixed within each condition and the specific discrepancies were randomly distributed across the six items.

These ratings were plotted on charts so participants could understand how their responses were similar to or different from those of Participant J. One chart was for agreement-about-self and the other was for agreement-about-third-party. As noted, the order in which charts were presented was counterbalanced.

4.2. Results

Among 180 people who took the study on Day 1, 157 of them (87%) continued to participate in Day 2. Good-judge evaluations from participants (M = 4.19, SD = 0.95) were submitted to a one-way ANOVA with the three experimental conditions for agreement. The model was statistically significant, F(2,154) = 11.89, p < 0.001, adjusted R² = 0.12. As expected, the results from planned contrast analyses suggested that the average good-judge evaluations from the high agreement-about-self (M = 4.69, SD = 0.90, n = 51) and high agreement-about-third-party (M = 3.98, SD = 0.92, n = 51) conditions were higher than the low agreement condition (M = 3.93, SD = 0.85, n = 55), t(154) = 2.70, p = 0.008, d = 0.44. Thus, there was an overall positive effect of agreement on good-judge evaluations. However, it should be noted that good-judge evaluations from the high agreement-about-third-party condition were not much higher than the low agreement condition. More importantly, as predicted, good-judge evaluations from the high agreement-about-self condition were higher than the high agreement-about-third-party condition, t(154) = 4.06, p < 0.001, d = 0.78 (see Fig. 3). Thus, agreement-about-self had a more positive effect on good-judge evaluations than agreement-about-third-party.
4.3. Discussion

The results of Study 3 replicated the findings from the previous two studies. Importantly, in Study 3 we attempted to minimize a potential confound that might have arisen from the fact participants in Study 2 were the source of information about a third-party target. In Study 3, by having participants and a (hypothetical) third-party target describe themselves (through sharing a story about a recent challenge), we attempted to reduce the asymmetry in the roles that participants and the third-party target played. Moreover, in Study 3 our predictions were supported in the domain of risk-taking propensity, a relatively neutral domain of social perception.

Interestingly, high agreement-about-third-party did not seem to have a strong positive effect on good-judge evaluations in Study 3. This might be because, as speculated above, the high agreement-about-third-party was achieved by the third party's description of him-or herself (instead of participants' description of him or her). Alternatively, it might be also explained by the fact that the third party was a novel person to participants, with limited familiarity and/or personal significance. This might have reduced the effect of agreement-about-third-party by restricting its ability to satisfy participants' relational and epistemic motives.

5. Study 4

The previous three studies provided converging evidence that agreement, in general, was positively associated with good-judge evaluations and that, more specifically, the strength of this relationship was bigger for agreement-about-self than agreement-about-third-party. With these results in hand, we sought to examine mechanisms and moderators: When might agreement about the self, or agreement about a third party, matter more or less to good-judge evaluations?

As noted in the Introduction, we predicted moderation effects for liking and familiarity because of their relevance to relational and epistemic motives, respectively. In order to examine these relationships, we asked online adult participants to report on two existing, real-world acquaintances who also knew one another. We measured the two types of agreement, liking, familiarity, and good-judge evaluations for each of the acquaintances participants reported on.

5.1. Method

5.1.1. Sample

We recruited 417 US online participants via Amazon.com's Mechanical Turk platform. Participants were paid 60 cents. The mean age of the participants was 33.0 (SD = 10.6) and 41% of them were females. The participants identified their race with the following frequencies: 77% White, 7% Asian, 6% Black, 4% Latino(a), and 5% two or more races.

5.1.2. Procedures

Participants were asked to think of two people they knew and who knew each other (we refer to these individuals as other1 and other2). Participants entered the initials of those people in blank boxes. These initials were used to refer to the individuals in subsequent questions in the survey. Participants then answered questions measuring their perception of agreement-about-self with one of the two people (agreement with other1 about the self; agreement with other2 about the self) and agreement-about-third-party (agreement with other1 about other2; agreement with other2 about other1). Then they answered questions on liking, familiarity, and good-judge evaluations regarding each of the identified people. Finally, participants answered demographic questions. We counterbalanced the order of various questions, including whether participants first answered the agreement questions for the first or second person they came up with, whether they first answered questions for agreement-about-self or agreement-about-third-party (separately for other1 and other2), and whether they first answered the good-judge questions for the first or second person they identified. Counterbalancing across these factors resulted in 16 different versions of the survey.

5.1.3. Measures

Every question used a 7-point Likert scale, ranging from 1 (“not at all”) to 7 (“very much”). Agreement-about-self was measured with the item, “To what extent do (other1/other2’s initials) see you as you see yourself? That is, to what extent does he or she have the same view of your character and personality as you have for yourself?” Agreement-about-third-party was measured with the item, “To what extent does (other1/other2’s initials) see (other2/other1’s initials) as you see (other2/other1’s initials)? That is, to what extent does he or she have the same view of (other2/other1’s initials)’s character and personality as you have for (other2/other1’s initials)?” Liking and familiarity were measured by the following questions, respectively, “How much do you like (other1/other2’s initials)?” and “How well do you know (other1/other2’s initials)?” Good-judge evaluations (α = 0.81) were measured using the same items as in Studies 1 and 2.

5.2. Results

Because each participant answered the agreement and good-judge questions for two other people, we constructed a multilevel dataset with the people participants identified at level 1 and participants themselves at level 2. Analyses were conducted using R with the lme4 and lmerTest packages (Bates et al., 2015; Kuznetsova et al., 2013). We first tested whether the results replicated our prior findings using a random-intercept model. As expected, when included in the same model predicting good-judge evaluations (M = 4.83, SD = 1.15), both agreement-about-self (M = 5.08, SD = 1.42), r = 0.36, p < 0.001, and agreement-about-third-party (M = 4.67, SD = 1.56), r = 0.09, p < 0.001, had positive relationships with good-judge evaluations. As predicted, the Pearson correlation between agreement-about-self and good-judge evaluations (r = 0.48) was significantly stronger than the correlation between agreement-about-third-party and good-judge evaluations (r = 0.28), Hotelling’s t = 5.86, p < 0.001. None of the counterbalancing factors significantly affected these patterns.

Before running moderation analyses, the agreement and potential moderator variables were mean-centered. As we predicted, the relationship between agreement-about-self and good-judge evaluations was moderated by liking for the judges (M = 6.01, SD = 1.24) such that the positive relationship between agreement-about-self and

![Fig. 3. Good-judge evaluations from Study 3. Note. Error bars represent standard error.](image-url)
good-judge evaluations was stronger when the perceivers had greater liking for the judges, $\gamma = 0.03, p = 0.038$. A simple slope analysis demonstrated that the relationship between agreement-about-self and good-judge evaluations was stronger when liking for the judge was high (slope $= 0.28, p < 0.001$) than when it was low (slope $= 0.19, p < 0.001$; see Fig. 4). Interestingly, liking for the judges did not significantly moderate the relationship between agreement-about-third-party and good-judge evaluations, $\gamma = 0.02, p = 0.198$. Finally, as expected, the relationship between agreement-about-third-party and good-judge evaluations was moderated by familiarity with the third party ($M = 5.82, SD = 1.26$) such that the positive relationship between agreement-about-third-party and good-judge evaluations was stronger when perceivers reported that they knew the third party very well, $\gamma = 0.04, p = 0.023$. A simple slope analysis demonstrated that the relationship between agreement-about-third-party and good-judge evaluations was stronger when familiarity with the third-party target was high (slope $= 0.25, p < 0.001$) than when it was low (slope $= 0.15, p < 0.001$; see Fig. 5).

5.3. Discussion

As in our first three studies, both agreement-about-self and agreement-about-third-party had positive relationships with good-judge evaluations—and agreement-about-self showed a stronger relationship than agreement-about-third party. Study 4 tested two moderation predictions, yielding evidence that supported both of them. As predicted, we found that people were especially sensitive to agreement-about-self when they were evaluating a judge they liked. We also found that people were especially sensitive to agreement-about-third-party when the third party was someone whom they thought they knew well. Liking for the judge did not demonstrate a significant moderation when the basis of the agreement was a third-party target. This pattern suggests that people might care more about sharing reality with someone that they liked, and the particular reality they wanted to share was the reality of who they were.

6. General discussion

In the present research, we examined the question of what leads perceivers to regard someone else as a good judge of people. Many factors, undoubtedly, play a role. We focused here on the role of agreement in perception. As we expected, people thought of those who agreed with them as good judges, especially when the agreement was about their own characteristics. In Study 1, we found these patterns in the context of relatively new real-world relationships. Studies 2 and 3 replicated the findings from Study 1 in experimentally controlled settings, addressing potential confounds and issues of internal validity. In Study 4, we found that the relationship between agreement and good-judge evaluations was stronger when people’s relational and epistemic motives could be satisfied more effectively, suggesting that a desire for shared reality with others drove the relationships we have observed.

We interpret the results of these studies as signaling the motivational aspects of social perception. Social perception functions as an important channel through which people’s motives to (a) feel connected to others and (b) confirm the accuracy of their perception can be fulfilled. These motives can be satisfied when perceivers reach agreement with others in social perception. It also seems likely that, in most cases, agreement about the self can more effectively satisfy these motives than agreement about a third-party target. The results from the present research not only support but also extend the predictions from self-verification theory (Swann, 2011) and shared-reality theory (Echterhoff et al., 2009). Specifically, by comparing the effect of agreement-about-self and agreement-about-third-party, the present research demonstrates that various self-verification effects previously discussed can be interpreted as a combination of “self” and “verification” and those effects might be quantitatively or qualitatively different from those from the combination of “third party” and “verification.” Therefore, it could be valuable to examine how findings from self-verification research change when the basis of verification is a third-party target, not the self. Moreover, given that familiarity moderated the effect of agreement-about-third-party in Study 4, it would be interesting to examine how clarity in self-concept might moderate the effect of agreement-about-self (Stinson, Wood, & Doxey, 2008). In terms of shared-reality theory, our findings imply that the effect of shared reality might vary depending on “what kind of reality is shared” between people. For future research, it could be interesting to examine how epistemic trust as reflected in good-judge evaluations created by different types of agreement influences the “saying-is-believing” effect, which has been the main phenomenon of shared-reality research.

The present findings have relevance in the domain of motivation more generally. In Higgins’ description of motivation (2012), people pursue three distinct ways of being effective: (a) achieving desired results (i.e., value), (b) establishing what is real (i.e., truth), and (c) managing what happens (i.e., control). As Etam and Higgins (2010) point out, previous research in social cognition has mainly examined the motivational implications of value effectiveness. Research on truth effectiveness, in particular, has been scarce. The present work suggests that individuals’ achieving congruence between their perception of a social target and others’ perception of that same target can enhance their truth effectiveness, which in turn, impacts social evaluations.

Looking beyond the present work, although we did not capture meta-perception (i.e., what a person thinks other people think about them), measuring such perception in the context of good-judge evaluations could be a promising future direction. The meta-perception and meta-accuracy literature has shown that perceivers tend to assume—often to an unwarranted degree—that others see them as they see themselves (e.g., Vazire & Carlson, 2010). Accordingly, if a person encounters a judge whose view of them departs from their own self-view, they may assume that this particular judge’s view departs...
also from the view of other judges. Future work might examine whether low agreement—about-self often leads to assumptions about low agreement-with-other-judges. Future work might also consider the dynamic process by which perceivers discover what people think of themselves based on the information about agreement-about-third-party and whether and how they adjust their views of the third-party targets in the direction of congruence with target self-views.

We believe the present work provides some initial answers to the question of “Who do people think is a good judge?” This topic deserves additional attention in part because it holds practical implications for how people relate and interact—and in part because, as our work suggests, it connects with some of our most basic relational and epistemic motives.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jesp.2016.07.008.

References


