

Not So Fast: The (Not-Quite-Complete) Dissociation Between Accuracy and Confidence in Thin-Slice Impressions

Personality and Social Psychology Bulletin
36(2) 264–277

© 2010 by the Society for Personality
and Social Psychology, Inc
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/0146167209354519
<http://pspb.sagepub.com>



Daniel R. Ames,¹ Lara K. Kammrath,² Alexandra Suppes,¹
and Niall Bolger¹

Abstract

After decades of research highlighting the fallibility of first impressions, recent years have featured reports of valid impressions based on surprisingly limited information, such as photos and short videos. Yet beneath mean levels of accuracy lies tremendous variance—some snap judgments are well-founded, others wrongheaded. An essential question for perceivers, therefore, is whether and when to trust their initial intuitions about others. In three studies of first impressions based on photos and videos, the authors examined accuracy for Big Five trait judgments as well as corresponding reports of confidence. Overall, perceivers showed a limited ability to intuit which of their impressions were more accurate than others, although a curvilinear effect emerged: In the relatively few cases where perceivers reported an absolute lack of confidence, their accuracy was indeed comparatively low. Across the studies, judgment confidence was shaped by sources at the judgment level and the judge level that were unrelated to accuracy.

Keywords

thin-slice impressions, first impressions, confidence, calibration, metacognition

Received January 29, 2009; revision accepted July 20, 2009

For a good portion of the past century, social psychology's message to everyday perceivers about forming first impressions seemed, if not wholly depressing, at least relatively clear: Don't. As the century drew to a close, Ross and Nisbett (1991) surveyed the evidence and summed up perceivers' performance by wondering, "How could we be so wrong?" (p. 143). Perceivers appear to regularly use the wrong sorts of data, such as misguided stereotypes (e.g., Fiske, 1998), and they frequently fail to make the right sorts of inferential adjustments, especially situational discounting (e.g., Gilbert, 1998). Yet a recent surge of work on accuracy strikes a more optimistic tone. A mounting number of studies suggest that, in at least some cases and even with only impoverished thin slices of evidence about a target, perceivers show noteworthy validity in gauging target personality (e.g., Ambady, Bernieri, & Richeson, 2000; Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Zebrowitz & Collins, 1997). So when an everyday perceiver understandably asks, "Can I trust my first impressions of someone?" what does the current balance of person perception research have to say? Should people be encouraged to embrace their initial intuitions or resist them? Should scholars cautiously reply that the evidence is contradictory and the implications unclear?

Parts of the answers to these questions already exist, including findings about which domains are typically judged well and under what conditions (e.g., Carney, Colvin, & Hall, 2007; Gray, 2008; Letzring, Wells, & Funder, 2006). However, an essential question remains seemingly unresolved: Assuming perceivers are sometimes right and sometimes wrong in their snap impressions, can they tell the difference? This question of metacognition—whether confidence covaries with accuracy—has received considerable attention in other domains of judgment, but snap-impression research does not appear to have addressed whether perceivers can distinguish their valid first impressions from their faulty ones. In this article, we tackle the issue of metacognition in snap impressions directly, a matter essential to answering the perceiver's question, "Can I trust my first impressions?"

¹Columbia University, New York, NY

²Wilfrid Laurier University, Waterloo, ON, Canada

Corresponding Author:

Daniel Ames, Columbia Business School, Columbia University,
3022 Broadway Avenue, New York, NY 10027
Email: da358@columbia.edu

In brief, we expected to find some accuracy, a great deal of confidence, and limited calibration between accuracy and confidence in first impressions. Moreover, we expected to identify several factors that tend to increase confidence without raising accuracy, including both judgment-level factors (such as the target reminding the perceiver of a type of person) and judge-level factors (such as person perception self-efficacy). We also expected that advising perceivers explicitly to go with their gut would boost confidence but have no effect on accuracy. Evidence from three studies of first impressions was generally consistent with these expectations—although our results also revealed an intriguing curvilinear effect of calibration for those relatively few cases where perceivers claim no confidence whatsoever. These results are the first we know of to assess calibration for Big Five personality trait judgments in thin-slice impression paradigms, shedding new light on the link between confidence and accuracy and on the factors that appear to shape confidence while bearing no relation to accuracy. We believe our findings point toward new research directions and also extend the answers person perception scholars can offer to everyday perceivers wanting to know whether and when they should accept their first impressions.

Confidence and Calibration

If the question is “Can perceivers pick up anything valid about a person in a thin slice?” the answer that has emerged from the last few decades of research seems to be a resounding “yes” (see Gray, 2008, for a recent review). But if the question is “Do individual perceivers unfailingly do so?” the answer is surely “no.” Given that individual thin-slice impression accuracy likely varies considerably, do people have much awareness of when their thin-slice impressions of others are valid? If they do, there may be reason to encourage snap judgments and leave it to perceivers to separate their inferential wheat from the chaff. Indeed, if calibration were strong enough, even trivial average levels of accuracy might not pose a concern because individuals could identify those occasions when their inferences were on track.

Some past work points toward at least modest calibration in social judgments. Using a task in which participants judged the status, roles, and deceptive behavior of videotaped targets, Smith, Archer, and Costanzo (1991) found that participants’ ratings of confidence were associated with performance, and Patterson, Foster, and Bellmer (2001) documented within-judge, across-judgment calibration (although they did not find any between-judge calibration). Another line of research suggesting calibration comes from studies that find perceivers show at least some ability to identify and use valid cues in their judgments of others (e.g., Funder & Sneed, 1993; Gosling, Ko, Mannarelli, & Morris, 2002). To the extent that this is the case, perceivers might be able to recognize when they have used valid cues or, just as importantly, when no valid cues have been observed, and they could adjust their confidence accordingly.

Although it is possible that perceiver confidence is calibrated with accuracy in thin-slice judgments, we think calibration will nonetheless be limited for a number of reasons. Work on calibration in other domains has revealed modest or meager effects, including research on reading feelings and other mental states (e.g., Ames & Kammrath, 2004; Marangoni, Garcia, Ickes, & Teng, 1995; Realo et al., 2003), behavior prediction (e.g., Dunning, Griffin, Milojkovic, & Ross, 1990), and impressions in the context of longer-term relationships (Swann & Gill, 1997). One source of miscalibration could be the explicit use of invalid cues. Numerous studies show that perceivers overestimate the diagnosticity of selected cues—sometimes called *pseudodiagnostic cues* (e.g., Gill, Swann, & Silvera, 1998). Perceivers’ judgments of a target’s agreeableness, for instance, may be correlated with eye contact and smiling even though those cues are not related to the target’s actual agreeableness (Borkenau & Liebler, 1995; Funder & Sneed, 1993). Thus, people may rely on cues that they incorrectly believe are valid, leading them to be wrong but not in doubt.

Another limit to calibration has to do with restricted introspective access to inferential processes. Many scholars have argued that social judgments, including snap impressions, tend to form spontaneously, rapidly, and implicitly (Ambady et al., 2000; see also Uleman, Blader, & Todorov, 2005). Thus, perceivers may experience the output of their inferential system (“she’s warm” or “don’t trust him”) but not be able to excavate the sources and operations that led to the inference. As such, they may not be able to reliably assess how valid their inferences are.

Sources of Confidence

In response to the perceiver’s practical question about trusting her impressions, we have so far sketched an answer that says, in effect, “Sometimes you’re right, sometimes you’re not, and your confidence may not signal the difference.” We wish to go further, though, and identify factors that could prompt confidence in the absence of accuracy. Identifying conditions that may make people feel sure, but not be right, could help perceivers exercise caution when it may be most warranted. We expect that some sources of invalid confidence are judgment-specific (factors associated with a particular impression), whereas others are at the judge level (individual differences in perceivers). We outline two predictions for each level in the paragraphs that follow.

First, because perceivers have limited access to their inferential processes, one possible source of confidence is the phenomenological experience of the inference itself—such as the extent to which an impression pops out. Perceivers may thus base their confidence on the strength of the output of the inferential system (“she seems very cold” or “he’s extremely emotional”), leading to a positive association between judgment extremity and judgment confidence. To take an analogy, consider the once-popular fortune-telling toy, the Magic 8-Ball, a liquid-filled sphere containing a

floating 20-faced die with an “answer” on each side (ranging from “My sources say no” to “Yes—definitely”), one of which would present itself to a window on the ball after shaking. This fortune-telling process is sheer chance, but the output could affect confidence in the quality of the judgment. So it may be with snap impressions. In some cases, an intuition of “sort of shy” might slowly float from the murky depths to the surface of a perceiver’s consciousness; in other cases, “totally introverted!” might instantly slam up against the window of consciousness. The content of this output, and the forcefulness with which it presents itself, may be taken as a signal of the quality of the inference. In short, we expect that judgment extremity (i.e., strongly agreeing or disagreeing that a target possesses a particular disposition) would predict judgment confidence, although extremity would not necessarily predict accuracy (e.g., the extreme inference could be based on a misguided stereotype).

Second, a source of confidence in thin-slice impressions may be consciously observed qualities of the target that activate perceivers’ explicit theories of persons, providing a conscious justification for a perceiver’s automatic intuitions. We call such cues *justifiers* because of their ability to justify confidence in an impression. As with judgment extremity, these justifiers could signal situations in which a perceiver feels sure without being right. For instance, a perceiver may experience an intuition that a target is highly conscientious; if the target also displays something specifically linked to the perceiver’s explicit theory of conscientiousness (e.g., the target wears glasses), the perceiver may be more confident in her judgment. Along the same lines, a perceiver may also feel more confident to the extent that the target matches an explicit person type (e.g., a nerd) or reminds the perceiver of someone he or she has met before (“she’s like my old roommate”). In each of these cases, the perceiver can use an explicit justification for an implicit feeling, which could serve to bolster confidence though not necessarily judgment accuracy.

We turn now to two sources of confidence at the level of the perceiver rather than the judgment itself. First, some perceivers may feel generally confident and comfortable in the domain of intuitive inferences and thinking with their gut. Others may prefer the careful use of logic and evidence (Cacioppo & Petty, 1982; Pacini & Epstein, 1999). We expect that those favoring an intuitive style would have greater confidence in their thin-slice impressions—they may be habitually comfortable with their snap judgments even if they are not right in any given instance. Indeed, we expect that preferences for processing style can be temporarily manipulated, with a resulting impact on confidence. Specifically, we believe that encouragement toward embracing an intuitive style would lead to higher levels of confidence in thin-slice impressions but to no improvement in accuracy.

A second potential source of confidence at the perceiver level is general person perception self-efficacy. Past research suggests that, accuracy aside, general beliefs about the self

predict estimates of one’s own performance (Ames & Kammrath, 2004; Ehrlinger & Dunning, 2003). If perceivers have limited access to the processes underlying a specific impression, they may fall back to such global beliefs (e.g., their assumed ability to read others). As such, we expect that person perception self-efficacy will predict confidence, though not necessarily accuracy, in a specific judgment.

Predictions and Plan of Study

What does person perception research have to say to perceivers faced with forming snap judgments of others? To extend the scholarly response to this question, we conducted three studies of first impressions, gauging accuracy and confidence. We expected some accuracy, considerable confidence, and limited covariance between the two. We also expected to reveal sources of confidence that were dissociated from levels of accuracy. Study 1 gauged impressions based on photographs; Study 2 focused on impressions based on short video clips. Study 3 also featured video clips; some participants were encouraged to adopt an intuitive thinking style, whereas others were told to be cautious about their intuitions or received no special instructions.

Study 1

In Study 1, participants judged targets depicted in photographs and reported their confidence in their impressions of each target. Based on previous research (e.g., Ambady et al., 2000), we hypothesized that perceivers would demonstrate above-chance levels of accuracy in their impressions but that accuracy would vary considerably from judgment to judgment and from perceiver to perceiver. The study design allowed us to examine the question of calibration both at the between-judge level (are judges who are generally more confident also more accurate?) and at the within-judge level (across judgments, is a judge likely to be more accurate when he or she is feeling more confident?). We expected that variation in confidence would show limited or no calibration with variation in accuracy at either level. In addition, we predicted that judgment extremity would be positively associated with judgment confidence, despite being unrelated to judgment accuracy.

Method

Participants. Twenty-five university undergraduates (14 female) participated in this study for cash reimbursement.

Procedure. Upon arrival, participants were given a photo packet and a rating packet. The photo packet contained photographs of 21 targets in random order. The rating packet contained 21 pages, one for each target. Participants were instructed to view a photograph, turn to the corresponding page in the rating packet, complete the items therein, and then move on to the next photograph. After viewing and making ratings for all 21 targets, participants were debriefed, paid, and released.

Photo judgments. For each photograph, judges rated their impressions of the target using the Ten-Item Personality Inventory, or TIPI (Gosling, Rentfrow, & Swann, 2003). The TIPI features two items for each Big Five dimension (extraversion, agreeableness, conscientiousness, emotional stability, and openness). Items (e.g., “She or he is extraverted, enthusiastic”) were rated on a 7-point scale ranging from –3 (*disagree strongly*) to 3 (*agree strongly*). After making all 10 trait ratings, judges also rated their overall confidence in their impression of the target on a 7-point scale ranging from –3 (*not confident at all*) to 3 (*extremely confident*).

Targets. Thirty Master’s of Business Administration (MBA) students were invited to be videotaped in a mock job interview. These targets were recruited to reflect diversity in personality. Twenty-one students agreed. Four still frames from the videos of each target were initially selected based on the target’s having a neutral expression (open eyes, closed mouth, minimum tension in facial muscles). Responses from pilot judges were used to identify the most neutral face of the four, which was used as the stimulus photograph in the photo packet.

Target accuracy criteria. In a prior MBA course, targets had recruited school peers and former work colleagues to provide ratings of their personality traits and management skills. Informants rated the targets’ Big Five personality traits on the TIPI. Targets had between 8 and 10 total informants, typically split evenly between work and school raters. Work colleagues generally knew the targets for 2 to 5 years; school colleagues were fellow members of randomly assigned small study groups who had worked together for 6 months.

For the personality accuracy criteria used in our present analyses, we used ratings provided by informants who gave the target a familiarity rating of 3 or 4 on a 4-point scale (indicating how well they knew the target; 1 = *not well at all*; 2 = *somewhat well*; 3 = *well*; and 4 = *extremely well*). Effect sizes and patterns of significance were comparable using criteria based on the target’s self ratings alone, ratings by targets’ school colleagues alone, ratings by targets’ work colleagues alone, and ratings by all informants regardless of familiarity.

Results

Accuracy. To compute aggregate judge accuracies, the mean judge rating for each trait of each target was computed across all 25 judges. These aggregate judgments were then correlated with the criterion values, separately for each trait domain and across all trait domains, as shown in Table 1. Aggregate accuracies were above zero for agreeableness, conscientiousness, and emotional stability, although only the latter two reached statistical significance ($p < .05$ based on $N = 21$ aggregate target judgments). Aggregate accuracies were near or below zero for extraversion and openness. Although some prior studies using photo stimuli have shown accuracy in judgments of extraversion, we suspect the mock interview context and our sample of still frames may have affected relevant cues.

Table 1. Individual and Aggregate Accuracy at Zero Acquaintance (Study 1)

Trait	Individual Judge Accuracy		Aggregate Judge Accuracy
	Mean	Standard Deviation	
Extraversion	–0.06 _a	0.18	–0.08 _a
Agreeableness	0.03 _a	0.24	0.12 _{a,b}
Conscientiousness	0.26* _b	0.19	0.50* _b
Emotional stability	0.17* _b	0.22	0.46* _b
Openness	–0.08 _a	0.22	–0.23 _a
All	0.09*	0.09	.23*

Aggregate judge accuracies represent the correlation between aggregate perceiver judgments and criterion. Individual judge accuracies represent the average within-perceiver correlation between judgment and criterion. Values within a column that do not share a subscript differ at $p < .05$.

*significantly different from zero at $p < .05$.

The aggregate results confirm that our photo stimuli carried valid cues about agreeableness, conscientiousness, and emotional stability that were accessible to some extent to perceivers. But how accurate were perceivers in this paradigm, and how much did accuracy vary? For each individual perceiver, we correlated ratings with criterion values, separately for each trait domain and across all trait domains. Table 1 reveals that individual judge accuracies were generally lower than aggregate judge accuracies. In the case of agreeableness and conscientiousness, the mean individual judge accuracy ranged from .17 to .26, significantly above zero ($ts > 3.5$, $ps < .01$).

Confidence. The mean confidence rating across all participants and judgments was 0.10 and the modal confidence rating was 1 (on a scale from –3 to +3). The distribution of confidence ratings was negatively skewed: 30.1% of the confidence ratings were negative, whereas 45.7% of the confidence ratings were positive.

Calibration. No calibration was apparent at the judge level. The correlation between average judge confidence and Fisherized individual judge accuracy was nonsignificant, $r(23) = -.26$, ns .

We turned next to calibration at the judgment level. Within a perceiver, were judgments more accurate when they were more confidently held? To answer this question, we investigated whether the association between a judgment and its criterion was moderated by judgment confidence: A significant positive interaction between judgment and judgment confidence would indicate calibration at the within-judge level.

We conducted a multilevel regression analysis using the proc mixed procedure in SAS, taking advantage of the nested structure of the data (judgments within judges). We regressed criterion on judgment, confidence, and trait. We also included a Judgment \times Trait interaction term and a Judgment \times Confidence interaction term. We hypothesized a

significant Judgment \times Trait interaction, which would indicate differences in accuracy across trait domains. We hypothesized a nonsignificant Judgment \times Confidence interaction, which would indicate no moderation of accuracy by confidence.¹ The results of this multilevel analysis showed no evidence of calibration. The predicted Judgment \times Trait interaction was found, $F(4, 2589) = 8.33, p < .001$. The Judgment \times Confidence interaction was, as hypothesized, nonsignificant, $F(1, 2589) = .02, ns$.

Sources of confidence. We hypothesized that one source of judgment confidence would be the strength of the output of the intuitive processing system: judgment extremity (taken as an absolute value on the -3 to $+3$ trait rating scale). We conducted a multilevel regression analysis that regressed judgment confidence on judgment extremity. Judgment extremity was indeed a highly significant predictor of judgment confidence, $\beta = .13, F(1, 2599) = 87.53, p < .001$. We additionally hypothesized that judgment extremity would not be a reliable indicator of judgment accuracy. A multilevel analysis indicated that, as predicted, judgment extremity did not moderate the relationship between judgment and criterion, $F(1, 2587) = .80, ns$.

Discussion

On balance, perceivers in Study 1 showed some ability to gauge target traits on the basis of photographs. However, the variance in individual judge accuracies was considerable. As expected, perceivers appeared to show no ability to assess the validity of their impressions. Highly confident perceivers were no more accurate than were unconfident perceivers, and within a given perceiver, judgments that were more confidently held were not correspondingly more accurate. As hypothesized, confidence was associated with impression extremity, which did not covary with accuracy.

Study 2

Considerable research has shown that the amount of evidence and the visibility of relevant cues and behaviors affect first impression validity (e.g., Funder, 1999; Kenny, 2004). In Study 2, we sought to provide more and different evidence to perceivers, although still within a thin-slice paradigm. Participants watched 60-second videos of targets, reporting trait judgments as well as confidence at the level of each trait judgment. We again expected a small but significant level of individual judge accuracy, although we expected little or no calibration. Study 2 also investigated several predictors of judgment confidence. We predicted that properties of the inference (judgment extremity), the stimulus (justifiers), and the perceiver (person perception self-efficacy and preferred information processing style) would affect impression confidence but that these factors would be largely dissociated from impression accuracy.

Method

Participants. Fifty-seven university undergraduates (35 female) participated in this online study for extra credit in their introductory psychology course.

Procedure. Participants logged in to the study website via an online subject pool sign-up system. Half of the participants first completed a set of personality surveys and then completed the video judgment tasks. The remaining participants completed the video judgment tasks and then the set of personality surveys. Upon completion of both tasks, participants were directed to a debriefing webpage.

Video judgment tasks. For each of the 21 targets, participants first clicked on a link that played a 60-second video. After watching each video, participants made five trait ratings for the target. Each trait rating was anchored by two positive and two negative adjectives from the TIPI scale. Extraversion, for example, was rated with an item that ranged from -3 (*reserved, quiet*) to 3 (*extraverted, enthusiastic*). After each trait rating, participants indicated their confidence in that rating on a scale from -3 (very low) to 3 (very high). Participants then answered the question "Was there a specific behavior/observation that influenced your impression, or was it a more general perception?" on a scale from -3 (*general perception*) to 3 (*specific observation*). Participants also answered the questions "Does this person seem to fit a well defined type?" and "Does this person remind you of someone you've met before?" on scales from -3 (*definitely no*) to 3 (*definitely yes*). Targets were presented to participants in one of two random orders.

Target videos. For the stimulus video clips, we used the videotapes of the 21 targets engaged in mock job interviews described in Study 1. The 15-minute interview featured a range of questions including "Why should we hire you over other qualified applicants?" "Tell me a little bit about your personality—those characteristics that make you you," and "Describe a leadership experience, a project that you initiated or took charge of." Targets answered these questions while seated. A final interview question was asked while standing: "Tell me about your interests and hobbies outside of work." Two cameras were used to record the interviews: one camera captured a close-up that framed the target's head and shoulders, a second camera captured the target's full body.

A 60-second video clip for each target was created by editing together four 15-second clips, beginning with the interviewee's first spoken word in response to the questions noted above, including 8 seconds of the close-up and 7 seconds of the body shot. Audio tracks were run through a low-pass content filter, written by the researchers in the sound programming language Csound, to preserve paralinguistic features of the speech while making the semantic content inaudible. Like other researchers (e.g., Ambady, LaPlante, Nguyen, Rosenthal, & Levinson, 2002), we pursued this approach to preserve potentially informative nonverbal information (e.g., rate of speech) without

introducing confounds or confusion based on fragments of conversational content.

Personality surveys. To assess information processing style, participants completed Pacini and Epstein's (1999) Rational-Experiential Inventory. The inventory contains 12 analytical items that originate from Cacioppo and Petty's (1982) Need for Cognition scale (e.g., "I have a logical mind") and 12 intuitive items that comprise the Faith in Intuition scale (e.g., "I often go by my instincts when deciding a course of action"). The 24 items were rated on 5-point scales from -2 (*definitely false*) to 2 (*definitely true*). To assess person perception self-efficacy, participants rated three items: "I am good at judging people's personality and character," "I am often able to read people's emotions correctly," and "I have difficulty assessing other people's strengths and weaknesses" (reverse scored). These items were rated on 4-point scales from 1 (*not like me at all*) to 4 (*very much describes me*) and were averaged to create the final measure. These personality scales were presented amidst several other unrelated personality scales (to prevent participants from overly focusing on the critical scales).

Results

Accuracy. Aggregate and individual judge accuracies were computed using the methods described in Study 1. As shown in Table 2, accuracies in this study were generally similar or slightly higher than those obtained in Study 1. Aggregate accuracies were significantly above zero for agreeableness, conscientiousness, and emotional stability ($ps < .05$ based on $N = 21$ aggregate target judgments) and slightly above zero for extraversion and openness. Individual judge accuracies were lower than aggregate judge accuracies. In the case of agreeableness, conscientiousness, and emotional stability, individual accuracies were significantly above chance ($ts > 4.5$, $ps < .001$).

Confidence. Judges were substantially more confident in their impressions in Study 2 than in Study 1. The mean confidence rating across all participants and judgments was 1.47 and the modal confidence rating was 2. The distribution of confidence ratings was even more negatively skewed than in Study 1: 5.8% of the confidence ratings were negative, whereas 82.3% of the confidence ratings were positive. A multilevel analysis revealed that within a judge confidence in judgments differed significantly across trait domains, $F(4, 5780) = 48.11$, $p < .001$: Participants were most confident when judging extraversion and least confident when judging openness.

Calibration. We found no evidence of calibration at the judge level; judges who were more confident on average were not correspondingly more accurate, as shown by a non-significant correlation between average judge confidence and Fisherized individual judge accuracy, $r(55) = .16$, ns .

To investigate calibration at the judgment level, we examined whether the relationship between judgment and criterion was moderated by judgment confidence. We conducted a multilevel analysis, regressing criterion on judgment,

Table 2. Thin-Slice Accuracy and Confidence (Study 2)

Trait	Individual Judge Confidence		Individual Judge Accuracy		Aggregate Judge Accuracy
	Mean	Standard Deviation	Mean	Standard Deviation	
Extraversion	1.75 ^a *	0.63	0.02 _a	0.21	0.03 _a
Agreeableness	1.42 ^a *	0.70	0.19 ^a *	0.24	0.51 ^a *
Conscientiousness	1.37 ^a *	0.73	0.17 ^a *	0.20	0.49 ^a *
Emotional stability	1.36 ^a *	0.65	0.15 ^a *	0.23	0.46 ^a *
Openness	1.21 ^a *	0.79	0.06 ^a *	0.22	0.15 _{a,b}
All	1.47 ^a *	0.55	0.14 ^a *	0.09	0.34 ^a *

Confidence measured on a scale from -3 (*extremely unconfident*) to 3 (*extremely confident*). Accuracy values are Pearson correlations, within-perceiver and in aggregate. Values within a column that do not share a subscript differ at $p < .05$.

^a $p < .10$, * $p < .05$.

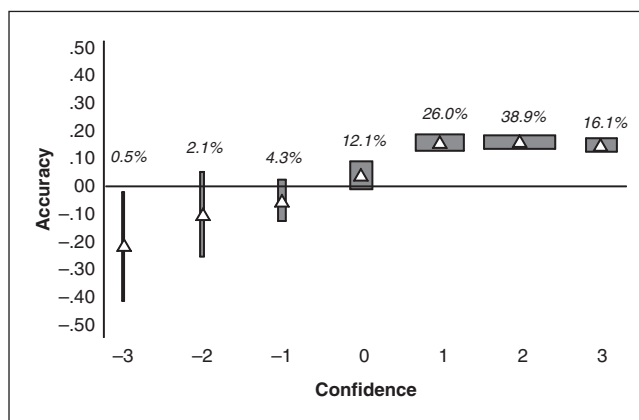


Figure 1. Accuracy and Confidence (Study 2)

Accuracy values (triangles) are standardized regression coefficients (β) using judgment to predict criterion within each level of confidence (X axis) across the entire data set. Block heights represent one standard error above or below the mean. Block widths and numerical values represent share of cases for a given level of confidence.

confidence, trait, a Judgment \times Trait interaction term, and a Judgment \times Confidence interaction term. The predicted Judgment \times Trait interaction was found, $F(4, 5606) = 8.33$, $p < .001$, indicating that accuracy significantly differed across trait domains, as it did in Study 1.

Unlike in Study 1, the Judgment \times Confidence interaction was marginally significant, $\beta = .05$, $F(1, 5606) = 3.39$, $p = .07$, suggesting potential calibration at the judgment level. To investigate this phenomenon more closely, we examined the relationship between judgment and criterion at each level of confidence across the entire data set. These relationships are shown in Figure 1. In contrast to a metacognition model showing a complete lack of calibration, the pattern of results suggested a curvilinear relationship between confidence and accuracy. When judges were fully unconfident in their

Table 3. Predictors of Confidence in Thin-Slice Impressions (Study 2)

Predictor	Standardized Coefficient (β)	F Statistic	p Value
Judgment level			
Judgment extremity	.33	1133.69	<.001
Saw something specific	.01	0.99	.32
Fits a type	.09	48.92	<.001
Reminds me	.06	17.37	<.001
Judge level			
Person perception efficacy	.14	4.44	.04
REI-faith in intuition	.09	1.85	.17
REI-need for cognition	-.15	7.31	<.01

Values reflect coefficients from a multilevel model of judgment confidence.

judgments, they indeed failed to show any judgment accuracy. When judges had some confidence in their judgments they showed modest judgment accuracy, but there was no difference in judgment accuracy between a judgment held with slight confidence (confidence = 1) and a judgment held with strong confidence (confidence = 3). This curvilinear relationship was tested statistically by running a second multilevel model that included the same predictors as the first model plus two additional terms to assess curvilinearity: a term for squared confidence and one for the Judgment \times Squared Confidence interaction. Confirming curvilinearity, this interaction term was significant, $F(1, 5604) = 5.65, p < .05$. A separate model featuring a term for cubed confidence did not reveal a significant cubic effect.

Sources of confidence. We conducted a multilevel analysis that regressed judgment confidence on four judgment-level predictors (judgment extremity, rating of target cue specificity, rating of target type-ness, and rating of target reminiscence) and three judge-level predictors (person perception self-efficacy, need for cognition, and faith in intuition). The results of this analysis are shown in Table 3.

As Table 3 reveals, perceivers were most confident in those impressions that were most extreme. They were also confident in their impressions to the extent that the target reminded them of someone or seemed to fit a type. In addition, perceivers with high person perception self-efficacy were especially confident, as were, to a lesser degree, those with an intuitive informational processing style.² Perceivers with an analytical informational processing style, on the other hand, were less confident in their impressions. Confirming dissociation of sources of accuracy and confidence, none of the predictors of confidence showed any association with judgment accuracy ($F_s < 3, ns$).

Discussion. Compared to perceivers who reacted to photographs in Study 1, Study 2's perceivers showed slightly greater accuracy when given 60 seconds of video on which to base their impressions. This converges with prior work showing that increased evidence is associated with increased

judgment validity (e.g., Funder, 1999). Whereas accuracy increased slightly with the shift from photos to videos, confidence increased substantially. The modal confidence rating was 2 on a scale from -3 to 3; only 5.8% of all confidence ratings were negative. Our results suggest that the most confident thin-slice judges are likely to be those who think they are very good at judging people in general, who have high faith in intuitive decision making, and who have low need for analytical decision making. Our results additionally showed that when a perceiver made multiple snap judgments, she or he was likely to be most confident when the impression was very extreme, when the target seemed to fit a type, or when the target reminded her or him of someone she or he knew.

Were these variations in confidence justified? Generally not, but with an interesting and potentially important qualification. Among that vast majority (more than 80%) of judgments endorsed with at least some confidence, confidence showed no association with accuracy. However, a curvilinear effect emerged when the comparatively few cases involving a complete lack of confidence were considered. In those cases, participants seemed to recognize the limited validity of their impressions. It is as if participants could tell when they got nothing—their utter lack of confidence was an indicator of a comparatively inaccurate judgment. But having some confidence and having extreme confidence were generally not associated with different levels of accuracy.

Study 3

Both Studies 1 and 2 showed a limited connection between confidence and accuracy in thin-slice judgments. Moreover, both highlighted factors that evoked confidence (including judgment extremity, person perception self-efficacy, and faith in intuition) but were unrelated to accuracy. In Study 3, we sought additional evidence for the dissociation between accuracy and confidence by seeking to manipulate confidence without affecting accuracy. The results from Study 2 concerning individual differences in faith in intuition suggest that receiving information advocating the validity of gut feelings may yield a temporary rise in confidence without a corresponding rise in accuracy. If this expectation is confirmed, it would clarify that the processes shaping confidence in snap impressions are at least partly distinct from the factors that yield accuracy in snap impressions. We tested this expectation in Study 3 by giving some participants information about the power of intuitions and gut feelings, cautioning other participants about the dangers of intuitions, and putting yet other participants in a control condition. Participants then recorded their impressions and confidence in their impressions of videotaped targets in a design similar to that used in Study 2.

Method

Participants. One hundred and ten university undergraduates (73 female) participated in this online study for extra

credit in their introductory psychology course. Of these, 70 participants (53 female) passed the manipulation check, described below, and were included in the final analyses. The final analyses included 24 participants in the control condition, 21 in the intuition condition, and 25 in the reason condition.

Procedure. Participants logged on to the study website via an online subject pool sign-up system. Participants completed a set of personality surveys and a set of video judgment tasks. Order of the personality surveys and video judgment tasks was counterbalanced across participants.

Video judgment tasks. The video judgment tasks began with the experimental manipulation. Participants in the reason and intuition conditions read information about the benefits of a particular form of decision making (Jordan, Whitfield, & Zeigler-Hill, 2007) before proceeding to the first target video. The manipulation instructions are noted in the appendix. Participants in the control condition were simply directed to the first video judgment task.

The video judgment tasks in Study 3 were highly similar to those in Study 2. Participants rated the personality of the target on the Big Five items described in Study 2 and indicated their confidence in these judgments. The rating instructions differed by condition. In the control condition, participants were instructed to “give your impression of this person’s personality and indicate your confidence in your judgments.” In the intuition condition, participants were instructed to “give your gut feelings about this person’s personality and indicate your confidence in your judgments.” In the reason condition, participants were instructed to “give your rational assessment about this person’s personality and indicate your confidence in your judgments.” Participants next answered the three questions about whether they had observed something specific about the target that influenced their impression, whether the target fit a well-defined type, and whether the target reminded them of someone.

Engagement check. Pilot testing of the design revealed that a small share of participants were not watching the videos in their entirety (perhaps because of technical issues or fatigue) and were instead basing their judgments on the still image of the target presented on the computer screen. To test for engagement, the 11th video was modified so that at 30 seconds a text screen appeared instructing participants to leave all the rating questions blank and to proceed immediately to the next video. Compliance with these instructions was used as a criterion for the inclusion of a participant’s data in the final analyses. Accuracy correlations in Study 3 were thus computed over 20 target judgments, rather than 21.

Personality surveys. Among other unrelated personality surveys, participants completed the three person perception self-efficacy items described in Study 2, which were averaged to create a composite person perception self-efficacy score. The intuitive analytical processing style measures

Table 4. Thin-Slice Accuracy and Confidence (Study 3)

Trait	Individual Judge Confidence		Individual Judge Accuracy		Aggregate Judge Accuracy
	Mean	Standard Deviation	Mean	Standard Deviation	
Extraversion	1.56 ^{*a}	0.76	0.07 ^{*a}	0.22	0.15 ^a
Agreeableness	1.31 ^{*b}	0.78	0.21 ^{*b}	0.21	0.53 ^{*a}
Conscientiousness	1.16 ^{*c}	0.78	0.13 ^{*b,c}	0.23	0.41 ^{*a}
Emotional stability	1.15 ^{*c}	0.84	0.14 ^{*b,c}	0.20	0.44 ^{*a}
Openness	0.99 ^{*d}	0.89	0.07 ^{*a,c}	0.21	0.24 ^a
All	1.24 [*]	0.76	0.15 [*]	0.12	0.38 [*]

Confidences and accuracies computed using participants in all three conditions. Values within a column that do not share a subscript differ at $p < .05$.

[†] $p < .10$, ^{*} $p < .05$.

were not included because we suspected they might interfere with the experimental manipulation.

Results

Accuracy. As shown in Table 4, accuracies observed in this study were similar to those seen in Study 2. Aggregate accuracies were significantly or marginally above zero for agreeableness, conscientiousness, and emotional stability ($ps < .05$ based on $N = 20$ aggregate target judgments) but not significantly above zero for extraversion and openness. Individual judge accuracies were lower than aggregate judge accuracies, with the average judge displaying accuracies modestly but significantly above zero on all trait domains ($ts > 2$, $ps < .05$).

To examine whether accuracy varied as a function of experimental condition or trait domain, a multilevel analysis was conducted, regressing criterion on judgment, trait, and condition. Additional terms included a Judgment \times Trait term (to assess whether accuracy differed across trait domains), a Judgment \times Condition term (to assess whether accuracy differed across experimental conditions), and the three-way interaction of Judgment \times Trait \times Condition. The only significant interaction was Judgment \times Trait, $F(4, 6834) = 2.58$, $p < .05$, indicating that accuracy varied by trait domain but not by experimental condition.

Confidence. Judges’ confidence in Study 3 was similar to that in Study 2. The mean confidence rating across all participants and judgments was 1.27 and the modal confidence rating was 2. The distribution of confidence ratings was again highly negatively skewed: 7.2% of the confidence ratings were negative, whereas 73.3% of the confidence ratings were positive. A multilevel analysis revealed that within a judge confidence in judgments differed significantly across trait domains, $F(4, 6637) = 57.84$, $p < .001$. As shown in Table 4, participants were again most confident when judging extraversion and least confident when judging openness.

Table 5. Thin-Slice Accuracy and Confidence Across Experimental Conditions (Study 3)

Condition	Judge Confidence		Individual Judge Accuracy		Aggregate Judge Accuracy
	Mean	Standard Deviation	Mean	Standard Deviation	
Control	1.08 _a	0.87	0.16 _a	0.14	0.37* _a
Intuition	1.57 _b	0.71	0.14 _a	0.12	0.31* _a
Reason	1.11 _a	0.62	0.15 _a	0.10	0.37* _a

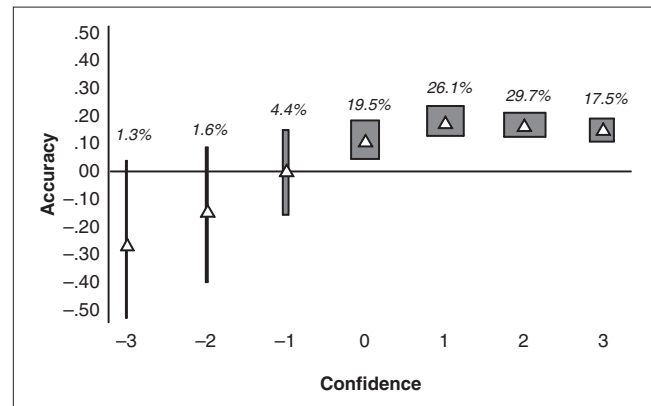
Confidences and accuracies computed across all trait domains. Values within a column that do not share a subscript differ at $p < .05$.

Table 5 reveals that whereas the experimental manipulations did not affect thin-slice accuracy, they did have a significant effect on thin-slice confidence, $F(2, 6637) = 3.01$, $p = .05$. As hypothesized, confidence was significantly higher in the intuition condition than in the control condition, $F(1, 6645) = 4.80$, $p < .05$; confidence in the reason condition did not differ significantly from that in the control condition, $F(1, 6645) = .01$, ns . The effect of condition on confidence did not significantly interact with trait domain, $F(8, 6637) = .48$, ns .

Calibration. We again found no evidence of calibration at the judge level. The correlation between judge mean confidence and Fisherized individual judge accuracy was not significant, $r(68) = .20$, ns . The association between judge mean confidence and individual judge accuracy was not moderated by experimental condition, $F(2, 64) = 1.08$, ns .

To investigate calibration at the judgment level, we conducted a multilevel analysis, regressing criterion on judgment, confidence, condition, a Judgment \times Condition interaction term, a Judgment \times Confidence interaction term, a Condition \times Confidence interaction term, and the three-way Judgment \times Confidence \times Condition term. The Judgment \times Confidence interaction term was significant, $F(1, 6571) = 6.93$, $p < .01$, suggesting calibration at the judgment level. The three-way interaction with condition, however, was not, $F(2, 6571) = 0.60$, ns , indicating that the experimental instructions did not affect calibration.

The evidence of calibration between confidence and accuracy at the judgment level in this study led us to again check for a possible curvilinear relationship. Figure 2 displays the relationship between judgment and criterion at each level of confidence across the entire data set. The data again suggested that accuracy jumped as perceivers went from no confidence to some confidence but failed to rise further once a positive level of confidence was attained. This curvilinear relationship was tested statistically by running a multilevel analysis that included the same predictors as in the first model plus additional terms to assess curvilinearity: a term for squared confidence and a term for the Judgment \times Squared Confidence interaction. Confirming curvilinearity,

**Figure 2.** Accuracy and Confidence (Study 3)

Accuracy values (triangles) are standardized regression coefficients (β) using judgment to predict criterion within each level of confidence (X axis) across the entire data set. Block heights represent one standard error above or below the mean. Block widths and numerical values represent share of cases for a given level of confidence.

the interaction term was significant, $F(1, 6569) = 13.53$, $p < .01$. These results suggest that in Study 3, like in Study 2, a relationship emerged between unconfidence and inaccuracy, even though among the impressions with at least some confidence (73.3% of cases) there was no evidence of calibration. A separate model featuring a term for cubed confidence did not reveal a significant cubic effect.

Sources of confidence. We conducted multilevel analyses within each condition, predicting judgment confidence with judgment extremity, specificity rating, type rating, remind rating, and person perception self-efficacy. The results of this analysis are shown in Table 6. Perceivers in all conditions showed highest confidence in their most extreme judgments. They were also more confident to the extent that the target seemed to fit a type or reminded them of someone. Perceivers in the reason and intuition conditions also were more confident to the extent that the target displayed something specific that caught their attention. These predictors of confidence shared no association with judgment accuracy, with one exception: In the case of specificity ratings, higher ratings of target observation specificity were associated with slightly lower accuracy, $\beta = -.03$, $F(1, 6546) = 5.69$, $p < .05$.

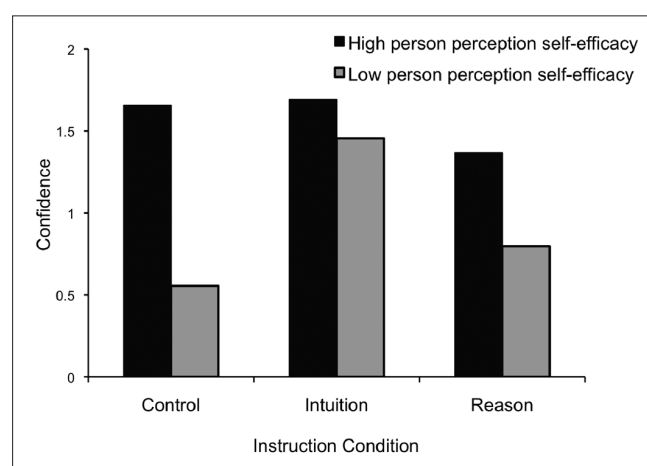
The effects of the first four potential sources of confidence did not differ significantly between conditions. The effect of the fifth source, person perception self-efficacy, did differ between conditions, $F(2, 6649) = 2.31$, $p < .10$. This interaction is shown in Figure 3. The effect of person perception self-efficacy on thin-slice impression confidence was significantly lower in the intuition condition ($\beta = .09$) than in the control condition, $\beta = .39$, $t(6649) = 2.08$, $p < .05$. In the reason condition, the effect of person perception self-efficacy ($\beta = .25$) did not significantly differ from that in the control condition, $t(6649) = 1.43$, ns . These findings suggest that the primary effect of the intuition instructions was to temporarily

Table 6. Predictors of Confidence in Thin-Slice Impressions by Condition (Study 3)

Predictor	Control Condition	Reason Condition	Intuition Condition
Judgment level			
Judgment extremity	.35***	.39***	.38***
Something specific	.00	.04*	.05*
Fits a type	.12**	.07**	.07**
Reminds me	.07**	.06**	.07**
Judge level			
Person perception self-efficacy	.32**	.18*	.10

Values reflect coefficients from a multilevel model of judgment confidence within each condition.

* $p < .05$, ** $p < .01$, *** $p < .001$.

**Figure 3.** Differential Experimental Effects for High and Low Self-Efficacy Perceivers (Study 3)

Values plotted are predicted values from a multilevel regression analysis. Person perception self-efficacy reflects participants' self-ratings of general person-reading ability. Confidence reflects participants' mean confidence rating across 100 thin-slice impressions.

boost the confidence of generally low-confidence perceivers to match the level of confidence typically displayed by generally high-confidence perceivers.

Participants in Study 3 who were encouraged to go with their gut showed greater confidence in their snap impressions of targets but they were no more accurate than were those in a control condition or those who were encouraged to use logic and evidence. In general, across and within conditions, judges who were more confident were not more accurate. However, as in Study 2, in the relatively small share of cases where perceivers indicated a complete lack of confidence, their judgments did indeed show a corresponding lack of accuracy. Thus, the within-judge, across-judgment calibration effect was a curvilinear one.

As expected, perceivers were more confident in impressions that were more extreme and those that were more

justifiable (an observation of a specific cue, the perception that the target fit a well-defined type, and the perception that the target was similar to a previous acquaintance). In the control and reason conditions, perceivers also expressed more confidence to the extent that they had high global self-efficacy in their ability to read others. Although impression extremity, impression justifiers, and global-self views were associated with higher confidence in snap judgments, these factors were not associated with greater accuracy.

General Discussion

What should perceivers make of their first impressions of others—do these judgments tend to be valid instincts or misguided guesswork? And when should perceivers embrace their snap judgments? This work extends the field's answers to these questions. In three studies using thin-slice impression formation paradigms, we examined individual impression accuracy as well as subjective confidence. As we expected, and as others have shown, we found evidence of above-chance levels of accuracy for some trait inferences based solely on photographs or brief videos. However, there was substantial variance in validity, allowing us to gauge metacognition—that is, whether perceivers' reports of confidence could discriminate between more and less valid inferences. Across the studies, we found no evidence for between-judge calibration: The level of confidence expressed by one judge versus the next was unrelated to the judgmental validity displayed by one judge versus the next. In Studies 2 and 3, we found evidence of a curvilinear within-judge calibration effect. We pinpointed this to an apparent difference in validity among judges who showed absolutely no confidence in their impressions and those who showed at least some confidence. Perceivers thus show some ability to tell when they have completely missed the mark in their initial judgments; in effect, people seem to realize when they are simply guessing. While offering a degree of optimism about metacognition in first impressions, it is worth noting that this effect seems to be located in the relatively small number of cases in which people showed no or minimal confidence in their judgments. In the vast majority of cases (70% to 80% of the time), people showed some to strong confidence in their impressions, and variance within these levels of confidence was not associated with differences in validity. In short, first-impression accuracy tends to be modest, first-impression confidence tends to be relatively high, and for the vast majority of impressions, accuracy and confidence tend to be unrelated to one another.

If accuracy does not regularly predict confidence, what does? When might people feel sure without being right? Our results point to several judgment-level factors, including judgment extremity (e.g., "she's very warm" versus "she's somewhat warm"). The shading of the explicit output from the inferential system may be taken as a cue by perceivers for

how true their inference is. Returning to the Magic 8-Ball analogy from this article's introduction, one might imagine that thin-slice impressions operate something like the supposed fortune-telling device. Perceivers have limited or no insight into the operations that bring the message (e.g., "he's unreliable" or "she's calm") to the surface, but they may take the substance of the message and the forcefulness with which the message appears as a proxy for the message's validity (i.e., "If I got a quick and clear answer, it must be true").

We also identified a number of potential justifiers (such as "she fits a certain type of person," "she reminds me of someone," or "I saw something specific that influenced my impression") that may heighten feelings of confidence. The first two of these suggest a conscious recognition of a correspondence between the target and a seemingly relevant knowledge structure. A major stream of social psychological work has addressed the powerful and pervasive impact of shared stereotypes on impressions; other work highlights that idiosyncratic stereotypes may also play a major role in trait judgments (e.g., Kenny, 2004). Elsewhere, scholars have suggested that mental models of significant individual others are often transferred onto new acquaintances (e.g., Andersen & Berk, 1998). In all of these cases, the knowledge structure that is activated and serves as template for understanding a novel individual could yield no improvement to judgment validity—or could markedly undermine validity—while making the perceiver feel sure about his or her impression.

Along with factors at the judgment level, we identified judge-level effects on confidence. Studies 2 and 3 suggest that people who were more confident in their snap judgments tended to be those with high self-efficacy in their ability to read others. In Study 2, we found that people who had an intuitive information-processing style were more likely to have confidence in their snap impressions, whereas people with an analytical information processing style were less likely to do so, even though processing style was unrelated to accuracy. Moreover, in Study 3, we manipulated attitudes about an intuitive processing style; perceivers who received information advocating intuition showed greater confidence, although no more accuracy, than did those in a control condition or receiving information about the perils of intuition.

Limitations and Generalizability

These studies are the first we know of to examine both confidence and accuracy in perceivers' Big Five trait judgments of targets in thin-slice paradigms. They are also the first we know of to find curvilinear calibration effects for social judgments and to show evidence for the potential impact of justifiers on confidence (but not accuracy). Of course, this research has limitations. We believe the studies reported here have the strength of featuring targets who are managers-in-training performing a natural task (i.e., a job interview) with criterion impression ratings provided by a comparatively

large number of close acquaintance informants, but questions remain about how extensible the results are. It is possible that accuracy and calibration would somehow be more substantial had we used different targets, different materials, or different participants. A more specific concern revolves around levels of accuracy for extraversion: Prior work suggests that extraversion, as a comparatively visible trait, is often the best judged of the Big Five in thin-slice paradigms. Two factors might have affected our participants' abilities to accurately gauge extraversion in our video studies (Studies 2 and 3). First, we sampled the videos to focus on targets' speaking behavior, thereby eliminating potentially informative cues from silence or lack of interjection. Second, our targets were from a highly competitive graduate program where most students could be expected to show proficiency in interviewing skills. Even the staunch introverts among them may have been able to display adaptive conversational behavior in an interview context. It is worth noting that even if judgments of extraversion are excluded from our analyses, the results concerning calibration and judge- and judgment-level factors affecting confidence remain virtually identical.

To further test the alternative possibility that calibration would be substantial in other thin-slice contexts (i.e., our results are somehow not generalizable), we analyzed the results from two of the highest profile thin-slice studies in the past decade: the bedroom and office studies reported by Gosling et al. (2002).³ Whereas Gosling and colleagues' published report focuses on accuracy and cue usage, the bedroom and office study judges also rated their confidence in their impressions of each target for each Big Five dimension. In new analyses of both studies across all five trait judgment dimensions, we did not find significant calibration effects (predicting confidence with accuracy). Consistent with the present results, though, we did find that confidence was significantly related to judgment extremity in both studies. In other words, in two widely cited thin-slice studies that showed considerable evidence of judgmental accuracy, there was no evidence of calibration between judges' accuracy and their confidence, but there was evidence that confidence and judgment extremity were closely linked. Although not definitive, these additional tests increase our faith that the limited calibration revealed in our studies was not an unrepresentative by-product of our specific methods or materials but rather may be a more general phenomenon.

In sum, whereas further work is certainly warranted, we believe our results converge with findings about the limits of calibration in other areas of social judgment (e.g., Ames & Kamrath, 2004; Dunning et al., 1990; Marangoni et al., 1995; Swann & Gill, 1997). We also believe that future research could helpfully explore the extent to which the curvilinear calibration effects we found and the sources of confidence we identified apply beyond Big Five trait judgments in thin-slice impressions.

Answering the Perceiver's Question

So what should person perception scholars tell perceivers who understandably ask, "Should I trust my first impressions of someone?" In our introduction, we raised the possibility that psychologists should demur, responding that the evidence seems contradictory and the implications are unclear. Rather than shy away, we wish to articulate some provisional responses. In some cases, we draw on the results presented here; in other cases, we draw on others' research. In doing so, we take a number of speculative leaps but we have faith that by taking clear stances, we will elicit other researchers' evidence and arguments, thereby sharpening the discipline's collective answer. So, let us turn to our own eight-point reply to the everyday perceiver's question.

First, a blanket "yes" or "no" response is not supported by the research evidence. As these studies and others' research have shown, snap-impression accuracy is sometimes above chance, although when it emerges it tends to be modest and highly variable from one judge or judgment to the next. Thus, the question could be profitably rephrased as "When should I trust my first impressions of someone?"

Second, some domains are better judged than others, although validity depends on the availability and use of valid evidence (cf. Funder, 1999). For instance, although participants in our studies using photographs and videos did not fare well in gauging targets' openness, impressions of openness based on targets' offices, bedrooms, and personal websites seem to show considerable validity (e.g., Gosling, Gaddis, & Vazire, 2008).

Third, perceivers' natural or spontaneous impressions may often not be restricted to dimensions on which they can validly judge a target. Ames and Bianchi (2008) recently found that people judging targets based on photos, videos, and face-to-face interactions tended to focus their impressions on the targets' agreeableness even though they could not validly judge that dimension. Ames and Bianchi concluded that everyday impressions are often guided more by the demands of coordinating interpersonal relations rather than the supply of valid evidence. Thus, whereas perceivers can judge some things validly based on meager evidence, what they often do judge goes beyond what they can validly infer.

Fourth, pooling judgments often appears to improve validity, so perceivers might consider combining their observations with the judgments of others. Our studies highlighted what others have shown previously: Aggregated judgments averaged across perceivers generally show greater validity than individual judgments. Thus, pooling impressions, even with one or two other perceivers, could markedly improve validity.

Fifth, confidence is a fallible clue to accuracy. One perceiver feeling more or less sure than another perceiver does not reliably predict which perceiver is more accurate. That said, we offer a sixth point: If a perceiver has utterly no confidence whatsoever in his or her impression, that may be a sign that the judgment may truly be guesswork. The

difference between some and much confidence may not say a lot, but the difference between some confidence and none whatsoever may be a meaningful indicator.

Seventh, if a perceiver's confidence is somewhat high, she or he might reflect on its source. If it is because the perceiver feels generally gifted at forming first impressions or is drawn to intuitive thinking or because the target reminds the perceiver of someone or fits a type, beware that these factors may be misleading sources of confidence.

Eighth and last, we think the research evidence points to the value of challenging the question itself: Do you need to trust your first impression? In some cases, urgency and other constraints demand that perceivers quickly adopt some stance toward a target. Yet in many cases, there are opportunities to learn more, interacting over time in multiple contexts. The balance of research suggests this is well-advised, with the validity in impressions increasing as evidence accumulates (e.g., Funder & Colvin; 1988; Kenny, 2004; Paulhus & Bruce, 1992).

Not all scholars will agree with each of these points we offer, and some would no doubt wish to add points of their own. Nonetheless, we suspect many person perception researchers would agree that the balance of evidence suggests that the answer to the "Should I trust my first impressions" question is neither to universally discard snap judgments nor to categorically embrace them. A more complex and interesting answer is emerging. Whether the present findings about metacognition endure, are overturned, or are revised, we believe that the advice person perception scholarship offers to everyday perceivers will be incomplete unless it identifies the place of subjective confidence in first impressions.

Appendix

Instructions Prompting Intuitive [Analytical] Cognitive Processing, Study 3

Everyone approaches problems by integrating some degree of rational thought and intuition. It is clear that rational thinking generally improves decisions. Beyond this, however, paying attention to and trusting intuitions also seems to improve decisions. [In contrast, paying too much attention to and relying too much on intuitions often undermines decisions.] There is clear evidence that people who trust [can overcome] their intuitions in decision making are more successful in many areas of their lives. In business, people who can make quick decisions based on their instincts [use reason unbiased by intuition] make better decisions, and perform better in management and executive positions. They are also more popular socially in the workplace. Interestingly, people who rely on their gut feelings [keep their gut feelings out of their thinking] also tend to be more successful in romantic relationships, being less likely to divorce if they are married.

(continued)

Appendix (continued)

They are also more likely to be identified by others as community leaders. In many different ways, intuitive people [people who can keep their intuitions from influencing their decisions] seem to excel.

In the present study, we are interested in understanding why people who trust their intuitions [use reason unbiased by intuition] are more successful than others. In the tasks that you are about to perform, we would like to see how well you can read another person's personality using your gut feelings about that person [using logic and reason alone, overcoming your initial gut feelings].

Acknowledgments

We extend heartfelt thanks to Dana Carney, Sam Gosling, Tory Higgins, and Alex Smith.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The authors declared that they received no financial support for their research and/or authorship of this article.

Notes

1. The model (and all subsequent calibration models) included a random intercept and a random effect for judgment, which would allow for individual differences in accuracy. Because we were primarily interested in the fixed effects of judgment (i.e., accuracy) and Judgment \times Confidence (i.e., calibration), we did not interpret the random effects.
2. Person perception self-efficacy and intuitive information processing style were highly multicollinear. Each was a significant predictor of confidence when entered separately into the model.
3. We are grateful to Sam Gosling and colleagues for sharing their data with us.

References

- Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 201-271). San Diego, CA: Academic Press.
- Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., & Levinson, W. (2002). Surgeon's tone of voice: A clue to malpractice history. *Surgery*, 132, 5-9.
- Ames, D. R., & Bianchi, E. C. (2008). The agreeableness asymmetry in first impressions: Perceivers' impulse to (mis)judge agreeableness and how it is moderated by power. *Personality and Social Psychology Bulletin*, 34, 1719-1736.
- Ames, D. R., & Kammrath, L. K. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior*, 28, 187-209.
- Andersen, S. M., & Berk, M. S. (1998). The social-cognitive model of transference: Experiencing past relationships in the present. *Current Directions in Psychological Science*, 7, 1-7.
- Borkenau, P., & Liebler, A. (1995). Observed attributes as manifestations and cues of personality. *Journal of Personality*, 63, 1-25.
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, 86, 599-614.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116-131.
- 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.
- Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology*, 58, 568-581.
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, 84, 5-17.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 554-594). New York: McGraw-Hill.
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego: Academic Press.
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, 149-158.
- Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, 64, 479-490.
- Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, pp. 89-150). New York: McGraw-Hill.
- Gill, M. J., Swann, W. B., Jr., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology*, 75, 1101-1114.
- Gosling, S. D., Gaddis, S., & Vazire, S. (2008). First impressions from the environments that we create and inhabit. In J. Skowronski & N. Ambady (Eds.), *First impressions* (pp. 334-356). New York: Guilford.
- Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82, 379-398.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504-528.

- Gray, H. M. (2008). To what extent, and under what conditions, are first impressions valid? In N. Ambady & J. Skowronski (Eds.), *First impressions* (pp. 106-128). New York: Guilford.
- Jordan, C. H., Whitfield, M., & Zeigler-Hill, V. (2007). Intuition and the correspondence between implicit and explicit self-esteem. *Journal of Personality and Social Psychology*, 93, 1067-1079.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, 8, 265-280.
- Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Quantity and quality of available information affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, 91, 111-123.
- Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology*, 68, 854-869.
- Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76, 972-987.
- Patterson, M. L., Foster, J. L., & Bellmer, C. (2001). Another look at accuracy and confidence in social judgments. *Journal of Non-verbal Behavior*, 25, 207-219.
- Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintance-ship on the validity of personality impressions. *Journal of Personality and Social Psychology*, 63, 816-824.
- Realo, A., Allik, J., Nolvka, A., Valk, R., Ruus, T., Schmidt, M., et al. (2003). Mind reading ability: Beliefs and performance. *Journal of Research in Personality*, 37, 420-445.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.
- Smith, H. J., Archer, D., & Costanzo, M. (1991). "Just a hunch": Accuracy and awareness in person perception. *Journal of Non-verbal Behavior*, 15, 3-18.
- Swann, W. B., Jr., & Gill, M. J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology*, 73, 747-757.
- Uleman, J. S., Blader, S. L., & Todorov, A. (2005). Implicit impressions. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 362-392). New York: Oxford University Press.
- Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, 1, 204-223.