# Knowing Me, Knowing You: The Accuracy and Unique Predictive Validity of Self-Ratings and Other-Ratings of Daily Behavior

Simine Vazire
Washington University in St. Louis

Matthias R. Mehl
University of Arizona

Many people assume that they know themselves better than anyone else knows them. Recent research on inaccuracies in self-perception, however, suggests that self-knowledge may be more limited than people typically assume. In this article, the authors examine the possibility that people may know a person as well as (or better than) that person knows himself or herself. In Study 1, the authors document the strength of laypeople's beliefs that the self is the best expert. In Study 2, the authors provide a direct test of self- and other-accuracy using an objective and representative behavioral criterion. To do this, the authors compared self- and other-ratings of daily behavior to real-life measures of act frequencies assessed unobtrusively over 4 days. Our results show that close others are as accurate as the self in predicting daily behavior. Furthermore, accuracy varies across behaviors for both the self and for others, and the two perspectives often independently predict behavior. These findings suggest that there is no single perspective from which a person is known best and that both the self and others possess unique insight into how a person typically behaves.

*Keywords:* accuracy, behavioral prediction, self-knowledge, informant reports, ecological momentary assessment

> Autobiography . . . is like asking a rabbit to tell us what he looks like hopping through the grasses of the field. How would he know?
> —Arthur Golden, *Memoirs of a Geisha*

If one wants to know what a person is like, the obvious solution is to ask him or her. Many people assume that this is the best approach; in other words, the self is often considered the single best expert when it comes to knowing how a person typically behaves. Pronin and her colleagues (Pronin, Kruger, Savitsky, & Ross, 2001) have shown that most people take it for granted that they know themselves better than others know them. Researchers frequently do the same. Seventy percent of studies published in the *Journal of Research in Personality* in 2003 involving personality assessments relied exclusively on self-reports (Vazire, 2006), presumably reflecting the implicit assumption that the self is the best source of information about what a person is like. This assumption was expressed more explicitly by an anonymous reviewer who commented on the use of informant ratings as a criterion measure for personality, saying, "the best criterion for a target's personality is his or her self-ratings . . . . Otherwise, the whole enterprise of personality assessment seriously needs to re-think itself" (personal communication, June 2003). This view is not surprising; the field of personality psychology has long relied on self-report inventories, reflecting the assumption that self-reports are the most valid method for assessing personality. In this article, we submit this assumption to a direct, empirical test by comparing the accuracy of self- and other-ratings of behavior.

As the opening quote by Arthur Golden implies, there are some important limitations to how well people can describe themselves. Recent research has brought to light many instances of glaring inaccuracies in self-perception (Bargh & Williams, 2006; Dunning, 2005; Epley & Dunning, 2006; Gosling, John, Craik, & Robins, 1998; John & Robins, 1994; Paulhus & John, 1998; Pronin & Kugler, 2007; Robins & John, 1997a; Sedikides & Strube, 1997; Wilson & Gilbert, 2003). Is it possible that the conventional wisdom is wrong and that others can see us more accurately than we see ourselves? If so, this would have far-reaching implications for our understanding of self-knowledge, interpersonal perception, and psychological assessment. Despite the inherent scientific and practical importance of this question, many researchers have given up hope of submitting it to a conclusive, empirical test and have essentially relegated the issue to the domain of philosophy. This is likely due to the fact that the long

Simine Vazire, Department of Psychology, Washington University in St. Louis; Matthias R. Mehl, Department of Psychology, University of Arizona.

Portions of this research were conducted as part of Simine Vazire's dissertation at the University of Texas at Austin. Portions of this research were supported by National Institute of Mental Health Grants MH64527-01A1 and MH52391 and by National Science Foundation Grant 0422924. We are grateful to Kimberly Chiew, Sam Gosling, Danny Heller, Nick Holtzman, Jason Rentfrow, Mike Strube, and Tal Yarkoni for their helpful comments on this article.

Correspondence concerning this article should be addressed to Simine Vazire, Washington University in St. Louis, Department of Psychology, Box 1125, 1 Brookings Drive, St. Louis, MO 63130. E-mail: svazire@artsci.wustl.edu

Journal of Personality and Social Psychology, 2008, Vol. 95, No. 5, 1202–1216
Copyright 2008 by the American Psychological Association 0022-3514/08/$12.00 DOI: 10.1037/a0013314

search for an objective accuracy criterion has proven extremely difficult (Kruglanski, 1989). This article tackles this issue by proposing a novel, naturalistic, behavioral accuracy criterion and provides the first comparison of the accuracy of self- and other-perceptions against such a criterion. The aim in this article is to (a) document the prevalence of the assumption that the self is the best expert on how a person typically behaves and (c) submit it to an empirical test by examining which perspective is more accurate and whether the two perspectives each provide unique insight.

How can theories of self- and other-perception inform this question? Theoretical models of self- and other-perception (e.g., Funder, 1999; Kenny, 1994, 2004) provide important conceptual frameworks for evaluating whether self-predictions or other-predictions of behavior are likely to be more accurate. The advantages and disadvantages of each perspective have been reviewed extensively elsewhere (e.g., Funder, 1999; Funder & Colvin, 1997; Hofstee, 1994; John & Robins, 1993; Kolar, Funder, & Colvin, 1996; Lucas & Baird, 2006; Paulhus & Vazire, 2007; Wilson, 2002; Wilson & Dunn, 2004). The consensus that emerges from these reviews is that both perspectives have their advantages and disadvantages. In addition, many features that seem like advantages may, in fact, hinder accuracy. For example, although the self has privileged access to mental states (e.g., thoughts, feelings, intentions), recent research suggests that this may in fact impair the accuracy of self-perceptions, for example by leading people to construct false reasons for their behavior (Bargh & Chartrand, 1999; Bargh & Williams, 2006; Nisbett & Wilson, 1977; Pronin & Kugler, 2007; Wilson & Dunn, 2004; Wilson & Gilbert, 2003).

In addition, there is a great deal of research on the many biases in self-perception (for a review, see Robins and John, 1997b). Self-perception is biased by motives such as the desire to create and maintain a positive self-concept (Taylor & Brown, 1988; Sedikides & Gregg, 2008), the desire to confirm one's existing self-views (Swann & Read, 1981), and the desire to improve oneself (Sedikides & Strube, 1997). In short, the existing body of research does not support granting an unequivocal advantage to self-perceptions. In the following section, we examine the existing empirical research on the accuracy of self- and other-perceptions.

## Empirical Research on Accuracy

The question of accuracy of behavior ratings or predictions is of critical importance for many fields. For example, self- and other-ratings are used to predict health behaviors, work performance, relationship outcomes, and academic performance (Ozer & Benet-Martínez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). However, systematic examinations of the relative predictive validity of self- and other-ratings are surprisingly rare in these literatures (cf. Agnew, Loving, & Drigotas, 2001; McAvay, Raue, & Brown, 2005; Rohrbaugh et al., 2004; Smith et al., 2007). As mentioned above, this is due at least in part to the methodological challenges involved with obtaining an appropriate criterion measure. Furthermore, the use of behavioral measures has declined overall in social and personality psychology research in recent years, perhaps suggesting that the field has developed in a way that "discourages researchers from observing behavior" (Baumeister, Vohs, & Funder, 2007, p. 402). However, several researchers have recently found novel and effective ways of dealing with the criterion problem and measuring behavior (e.g., Funder, Furr, & Colvin, 2000; Furr & Funder, 2007; Spain, Eaton, & Funder, 2000). In this section, we review the approaches other researchers have taken, with the aim of identifying the important advances and significant gaps remaining in research on the accuracy of self- and other-ratings of behavior.

Researchers testing the accuracy of self- and informant ratings have traditionally relied on one of three types of criterion measures: laboratory-based behavior, self-reports of real-world daily behavior, or objective measures of life outcomes. These approaches reflect different solutions to the apparent trade-off between two goals: (a) using a criterion measure that is independent of self- and informant ratings and (b) using an ecological criterion measure that captures a representative sample of targets' behavior.

The first goal is important because a criterion measure that is based on self- or informant ratings inevitably introduces shared method variance between the ratings whose accuracy is being evaluated and the accuracy criterion. This artificially biases the accuracy estimates toward the perspective whose ratings are being used as the criterion. For example, use of self-reports of behavior to test the accuracy of self- and informant predictions unduly favors self-predictions.

The second goal, external validity, is equally vital. Criterion measures that rely on objective but narrow behaviors (e.g., laboratory-based or single life outcome measures) are problematic because they may not reflect how a person typically behaves. Laboratory-based behavior may differ from how a person usually behaves for a variety of reasons (e.g., impression management, demand characteristics, situational affordances). Outcome measures, such as getting hired or fired or married or divorced are multiply determined and heavily influenced by factors outside of the person. Thus, a representative design—the random sampling of behavior from the full spectrum of daily life—is a methodological mandate for achieving external validity (Brunswick, 1956).

Until now, however, the two goals of obtaining external validity and methodological independence between predictors and criterion measures have been at odds. No available measure has adequately achieved both goals, and researchers have had to choose a criterion that compromised one or both of these goals. Nevertheless, the studies that have used these criterion measures have yielded important information about the relative strengths and limitations of self- and other-perceptions. Next, we review the results from studies, using each of the three types of criterion measures.

Studies comparing the accuracy of self- and informant ratings with a laboratory-based behavioral criterion mostly support the notion that informant ratings can be at least as accurate as self-ratings. For example, Kolar et al. (1996) found that the average correlation between informant reports and Riverside Behavioral Q-Sort codings was .44, whereas the average correlation between self-reports and the behavioral codings was .28. Consistent with these findings, John and Robins (1994) found that in a laboratory-based group managerial task, participants agreed more with the psychologists' ratings of performance when evaluating their peers (mean $r = .84$) than when evaluating themselves (mean $r = .58$). Finally, Levesque and Kenny (1993) found that even after only a brief exposure to targets, unacquainted informants were better at predicting targets' extraversion-related behavior (mean $r = .65$) than were the targets themselves (mean $r = .39$). Taken together, these studies show that when compared with a laboratory-based criterion, informant ratings are at least as accurate as self-ratings. However, as mentioned above, laboratory-based behavior may not be representative of how a person generally behaves in real life.

Studies with self-reports of real-world, daily behavior as a criterion paint a different picture of the relative accuracy of self- and informant ratings. In one study, Spain et al. (2000) compared self- and other-ratings of personality with momentarily assessed self-report ratings of emotion. In their study, self-ratings of extraversion and neuroticism were better than informant ratings at predicting self-reported positive and negative affect, respectively.

In a another study with real-world experience as a criterion, Shrauger and his colleagues (Shrauger, Ram, Greninger, & Mariano, 1996) compared the accuracy of predictions of life events (including common behaviors such as play cards, forget an appointment, and tell a joke that falls flat) by the self and close others against a self-report criterion measure obtained 6 months later and found that the self was more accurate than were close others. Although these studies suggest that self-ratings may have superior validity for predicting real-world daily behavior, the shared method variance between self-predictions and the self-report-based criterion measures makes it impossible to draw strong conclusions.

One final approach to tackling the criterion problem is to compare the usefulness of self- and other-ratings for predicting important, objective life outcomes. For example, Oltmanns, Turkheimer, and their colleagues (Fiedler, Oltmanns, & Turkheimer, 2004; Oltmanns, Gleason, Klonsky, & Turkheimer, 2005) have examined the value of self- and other-ratings of personality disorders for predicting success in military training. They have argued that other-ratings may be particularly useful for assessing personality disorders because lack of self-awareness is characteristic of many personality disorders. Their findings corroborate this view—other-ratings often provide incremental validity over self-ratings in predicting outcomes such as early discharge from the military.

Comparable findings regarding the incremental validity of ratings by close others over self-ratings have recently surfaced in the coping literature. For example, Rohrbaugh and his colleagues (2004) found that spouse ratings of patient efficacy predicted patient survival in the first 4 years following heart failure, above and beyond patient-rated self-efficacy. Similarly, Smith and his colleagues (2007) recently reported that spouse ratings, but not patient ratings, of patient hostility were associated with the severity of the patients' coronary artery calcification, an important risk factor for and component in coronary heart disease.

In summary, the three approaches that have been used to examine the accuracy of self- and informant ratings have provided conflicting answers. Studies with laboratory-based and life-outcome criterion measures suggest that well-acquainted others can be at least as accurate as the self, whereas studies with criterion measures based on real-world, daily behavior have favored the validity of self-reports. This pattern is confounded, however, by the fact that all studies with real-world, daily behavior as a criterion have relied on self-reports. Thus, the question of whether the self or close others can better predict a person's daily behavior has not yet been definitively answered. A new approach is needed to resolve this question.

## Measuring Behavior: An Ecological Act–Frequency Approach

From the evidence presented above, it is clear that what is needed is a criterion that is independent of self- and other-ratings (i.e., objective) and that measures behavior in a naturalistic setting (i.e., ecological). We propose that this can be achieved by combining the act frequency approach (AFA) to personality (e.g., Gosling et al., 1998) with naturalistic momentary assessment methods (e.g., Conner, Barrett, Tugade, & Tennen, 2007; Goodwin, Velicer, & Intille, 2008).

In the AFA, it is argued that daily behaviors form the core of what a person is like (Buss & Craik, 1983). This view is captured in Craik's (1993) question, "Is our task as personality psychologists to understand only the accentuated, highlighted, pronounced, revealed features of a person's life, or all of that life?" (p. 280) and his poignant answer, "lives are lived day by day, one day at a time, from day to day, day after day, day in day out. Lives as we experience and observe them are inherently quotidian" (p. 279).

It follows from this view that a good test of the accuracy of self- and other-ratings is to compare them with how a person actually behaves on a daily basis. For example, Gosling and his colleagues (1998) used the AFA to compare participants' self-reports of behavior in a laboratory task with observer codings of those behaviors. Their results show that people's perceptions of how they behaved were accurate for some behaviors (e.g., told a joke) but not others (e.g., interrupted others). This study illustrates the appeal of the AFA for our purposes: It provides a way of measuring what a person is like that is independent of self- and other-ratings. However, it did not address the question of the accuracy of informant reports.

Although we agree with AFA proponents that the frequency of displayed behavior captures the essence of what a person is like, laboratory-based measures of behavior may have limited external validity. Indeed, the AFA is based on the view that what a person is like is best captured by general trends in behavior rather than by single snapshots (Buss & Craik, 1985, p. 936). Thus, to meet our second goal, representativeness or external validity, we propose to take the AFA out of the laboratory and measure people's behavior in their everyday lives.

The obstacles to collecting such data are clear. Traditional experience-sampling methods, although externally valid, still rely on self-reports of behavior (Bolger, Davis, & Rafaeli, 2003; Conner et al., 2007). In addition, naturalistic behavioral measures are often quite obtrusive, which is problematic because this may lead to measurement-induced reactivity (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981). One solution is to make use of recent advances in mobile computing technologies to unobtrusively track people's daily behavior (Goodwin et al., 2008). One of these recent advances, the electronically activated recorder (EAR; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001), allows researchers to unobtrusively record ambient sounds directly from people's daily lives by asking participants to wear a small, pocket-sized digital audio recorder and a small lapel microphone. The EAR allows researchers to assess behavior in a way that is both objective and representative of real-world behavior (Mehl, 2007).

In a series of prior EAR studies, Mehl and colleagues have shown that a broad spectrum of acoustically detectable, daily behaviors (e.g., talking, listening to music, attending a class, laughing, arguing) can be assessed reliably and with low levels of reactivity from EAR-recorded ambient sounds (Mehl & Holleran, 2007), show very large between-persons variability and good temporal stability over a period of 4 weeks (Mehl & Pennebaker, 2003), and have good convergent validity with theoretically related trait measures such as the Big Five personality dimensions (Mehl, Gosling, & Pennebaker, 2006) and subclinical depression (Mehl, 2006). Thus, by using the EAR to obtain estimates of real-world

act frequencies, we can achieve a criterion measure that is both objective and externally valid.

## The Present Studies

The central question underlying our research is this: How accurate are self- and other-perceptions of what a person is like? We suspect that there is a widespread assumption that the self is the best expert for describing what a person is like. In Study 1, we document the existence of this assumption. In Study 2, we test its validity. Our research design in Study 2 is novel in two important ways. First, our accuracy criterion, based on self-report-free measures of real-world daily behavior, presents a unique constellation of methodological features that improve on previous approaches to measuring accuracy. Specifically, our EAR-based criterion measure is the first to combine objectivity and real-world representativeness. Second, we compare self-ratings with ratings from the best-acquainted informants possible: participants' closest friends and family members. This allows us to provide a strong test of the relative accuracy of self- and other-ratings of behavior. Use of anyone but the targets' closest friends and family members as our informants would have left open the possibility that there are still others out there who know the target better.

By incorporating these design features, we hoped to provide a rigorous and conclusive test of the relative accuracy and unique predictive validity of self- and other-ratings. To do this, we examined two research questions: Which perspective is more accurate overall? And, do the two perspectives provide unique insight? For each question, we examined both laypeople's intuitions and the actual empirical evidence. Below, we state our general predictions for each of these questions.

### *Question 1: Which Perspective Is More Accurate Overall?*

Our first aim was to determine whether one perspective, self or others, is generally more accurate than the other. With respect to laypeople's intuitions (Study 1), we predicted that the self would be rated as more accurate than others. We then tested the actual accuracy of the two perspectives (Study 2). We predicted that accuracy would vary across behaviors; that neither perspective, self or other, would be consistently more accurate for all behaviors; and that the overall accuracy of the two perspectives would be about equal. Specifically, we expected that the average accuracy correlation (across all behaviors) for the self would not be significantly different from the average accuracy correlation for informants.

### *Question 2: Do the Two Perspectives Provide Unique Insight?*

Given that we expected variation in accuracy across behaviors, we predicted that the self would be a better judge of some behaviors and that well-acquainted others would be a better judge of other behaviors. We used the Johari window (Luft & Ingham, 1955) as a framework to visually represent these expected differences (Figure 1). The Johari window partitions information about a person into four domains, according to whether self-knowledge for a particular trait is high or low and whether other-knowledge is high or low. We suspected that the lay predictions of self- and other-accuracy would fall mostly within the bottom right quadrant (known only to self). With respect to actual accuracy, we predicted that the behaviors would fall into all four quadrants of the model. In other words, we predicted that the two perspectives would provide both unique and overlapping information.

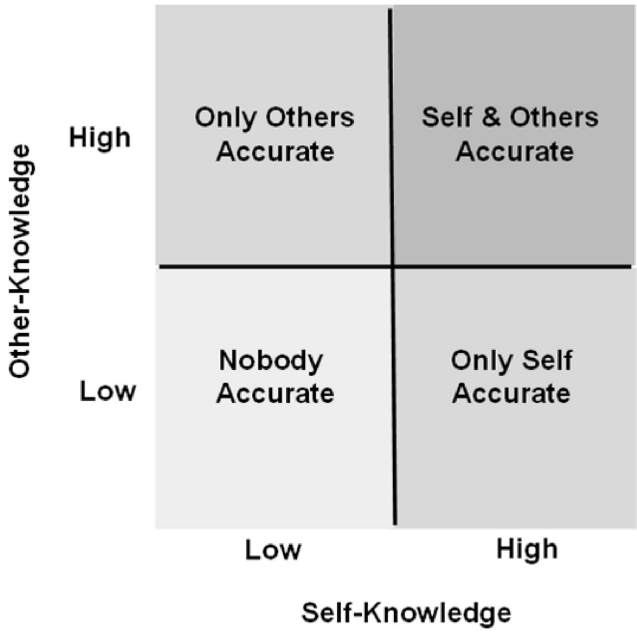


*Figure 1.* Adaptation of the Johari window (Luft & Ingham, 1955).

It is also possible that even when both perspectives are accurate, they are providing unique information. To test this, we conducted a multiple regression to see whether the two perspectives independently predicted behavior. If both self- and other-ratings turn out to be significant predictors of behavior when entered simultaneously into a regression, then they are providing unique, independent sources of knowledge about what a person is like.

We tested these two research questions in two studies. In the first study, we examined people's lay beliefs about the accuracy of self- and other-perceptions of behavior by collecting predictions from 61 people who were not involved in the subsequent phases of the study. In the second study, we conducted three phases of data collection. In the first phase, we collected self-ratings of behavior from 80 new participants and asked them to nominate three people who know them well. In the second phase, we obtained informant reports of behavior from the self-selected informants via the Internet. In the third phase, we collected 4 days of behavioral observations from the 80 target participants with the EAR.[1]

[1] A portion of the EAR data presented here were used in Mehl, Vazire, Ramirez-Esparza, Slatcher, and Pennebaker (2007) for an analysis of sex differences in daily word use and in Hasler, Mehl, Bootzin, and Vazire (in press) for an analysis of diurnal rhythms in affect-associated everyday behaviors. The analyses and findings reported here do not overlap with those presented in those articles.

Table 1
*Lay Perceptions of the Accuracy of Self- and Other-Ratings of Daily Behavior: Study 1*

| | Perceived self-accuracy | | Perceived other-accuracy | | | |
|---|---|---|---|---|---|---|
| ACT behavior | *M* | *SD* | *M* | *SD* | *M* difference | Cohen's *d* |
| Alone | 5.80 | 1.33 | 3.43 | 1.52 | 2.38 | 1.66 |
| Crying | 5.44 | 1.86 | 3.10 | 1.88 | 2.34 | 1.25 |
| Reading | 5.57 | 1.55 | 3.31 | 1.74 | 2.26 | 1.37 |
| Indoors | 5.49 | 1.27 | 3.59 | 1.69 | 1.90 | 1.27 |
| Commuting | 4.98 | 1.85 | 3.13 | 1.59 | 1.85 | 1.07 |
| Talking one-on-one | 5.38 | 1.47 | 3.52 | 1.66 | 1.85 | 1.19 |
| Sleeping | 5.93 | 1.33 | 4.08 | 1.69 | 1.85 | 1.22 |
| Watching TV | 5.54 | 1.54 | 3.70 | 1.72 | 1.84 | 1.13 |
| Playing sports | 5.77 | 1.48 | 3.97 | 1.84 | 1.80 | 1.08 |
| On the phone | 5.31 | 1.69 | 3.57 | 1.61 | 1.74 | 1.05 |
| On the computer | 5.59 | 1.46 | 3.92 | 1.67 | 1.67 | 1.06 |
| Eating | 5.21 | 1.39 | 3.56 | 1.53 | 1.66 | 1.13 |
| Talking to same sex | 5.46 | 1.52 | 3.80 | 1.58 | 1.66 | 1.07 |
| Talking to opposite sex | 5.57 | 1.37 | 3.95 | 1.65 | 1.62 | 1.07 |
| Attending lecture | 6.07 | 1.58 | 4.49 | 1.73 | 1.57 | 0.95 |
| Outdoors | 4.92 | 1.75 | 3.39 | 1.36 | 1.52 | 0.98 |
| Listening to music | 5.41 | 1.49 | 3.95 | 1.80 | 1.46 | 0.88 |
| Arguing | 5.03 | 1.68 | 3.64 | 1.63 | 1.39 | 0.84 |
| At a coffee shop/bar/restaurant | 5.23 | 1.64 | 3.85 | 1.53 | 1.38 | 0.87 |
| Talking in a group | 5.28 | 1.59 | 4.05 | 1.64 | 1.23 | 0.76 |
| Socializing | 5.31 | 1.47 | 4.13 | 1.65 | 1.18 | 0.76 |
| With other people | 5.33 | 1.31 | 4.16 | 1.57 | 1.16 | 0.81 |
| Singing | 4.79 | 1.71 | 3.70 | 1.88 | 1.08 | 0.61 |
| At work | 5.93 | 1.70 | 5.02 | 1.48 | 0.92 | 0.57 |
| Laughing | 5.15 | 1.51 | 5.11 | 1.32 | 0.03 | 0.03 |

*Note.* $N = 61$. All Cohen's *d* values, except for laughing, are significant at $p < .01$, which is not significant. Behaviors are in descending order of the magnitude of the difference between perceived self-accuracy and other-accuracy. Ratings were done on a 7-point scale.

## Study 1

### *Method*

*Participants.* Sixty-one residents of Austin, Texas, were recruited through convenience sampling to provide the lay predictions. Eight research assistants each collected data from 4 to 12 participants whom they knew personally. Participants included friends, classmates, and family members of the research team. Fifty-seven percent of the participants were female (43% were male), and the mean age was 26 years old ($SD = 9.9$ years). The ethnic breakdown was 59% White, 26% Asian, and 15% Hispanic.

*Measures.* Lay predictions were made with the ACT questionnaire (appendix). The ACT was designed specifically for these studies to obtain ratings of the behaviors that can subsequently be assessed with the EAR. Thus, the behaviors rated on the ACT are meant to represent a broad range of daily behaviors that are detectable from ambient sounds.

For Study 1, the scale was altered from the version presented in the appendix. Specifically, participants rated each behavior twice on separate, 7-point Likert-type scales. On the first rating, participants were asked to rate "how accurate people are at predicting how much ***they themselves*** perform this behavior." On the second rating, participants were asked to rate "how accurate people are at predicting how much ***other people they know well*** perform this behavior" (emphasis in original). The two ratings were made side-by-side to emphasize the comparison between self- and other-accuracy (i.e., raters chose one number for self-accuracy and one number for other-accuracy, next to each other, for each item before moving on to the next item). The reliability of the lay predictions was acceptable; intraclass correlations (ICC) [2,*k*] = .67 for ratings of self-accuracy and .83 for ratings of other-accuracy.

### *Results and Discussion*

*Question 1: Which perspective is perceived to be more accurate overall?* We predicted that laypeople would consider self-ratings to be consistently more accurate than other-ratings. This prediction was clearly supported. Table 1 presents the average rating of self-accuracy and other-accuracy for each behavior, as well as the difference between these two means. Each of the 25 behaviors was rated as being significantly more accurately perceived by the self than by others, except for laughing, which was rated relatively high in accuracy for both perspectives. The average effect size for this difference was equivalent to a Cohen's *d* of .99, indicating that the perceived difference in accuracy between self and others is very large.

*Question 2: Are the two perspectives perceived to provide unique insight?* We next examined whether the participants thought that the two perspectives provide insight into different behaviors. Despite the fact that they favored self- over other-accuracy for every behavior, the magnitude of that difference could vary across behaviors, which would allow behaviors to fall

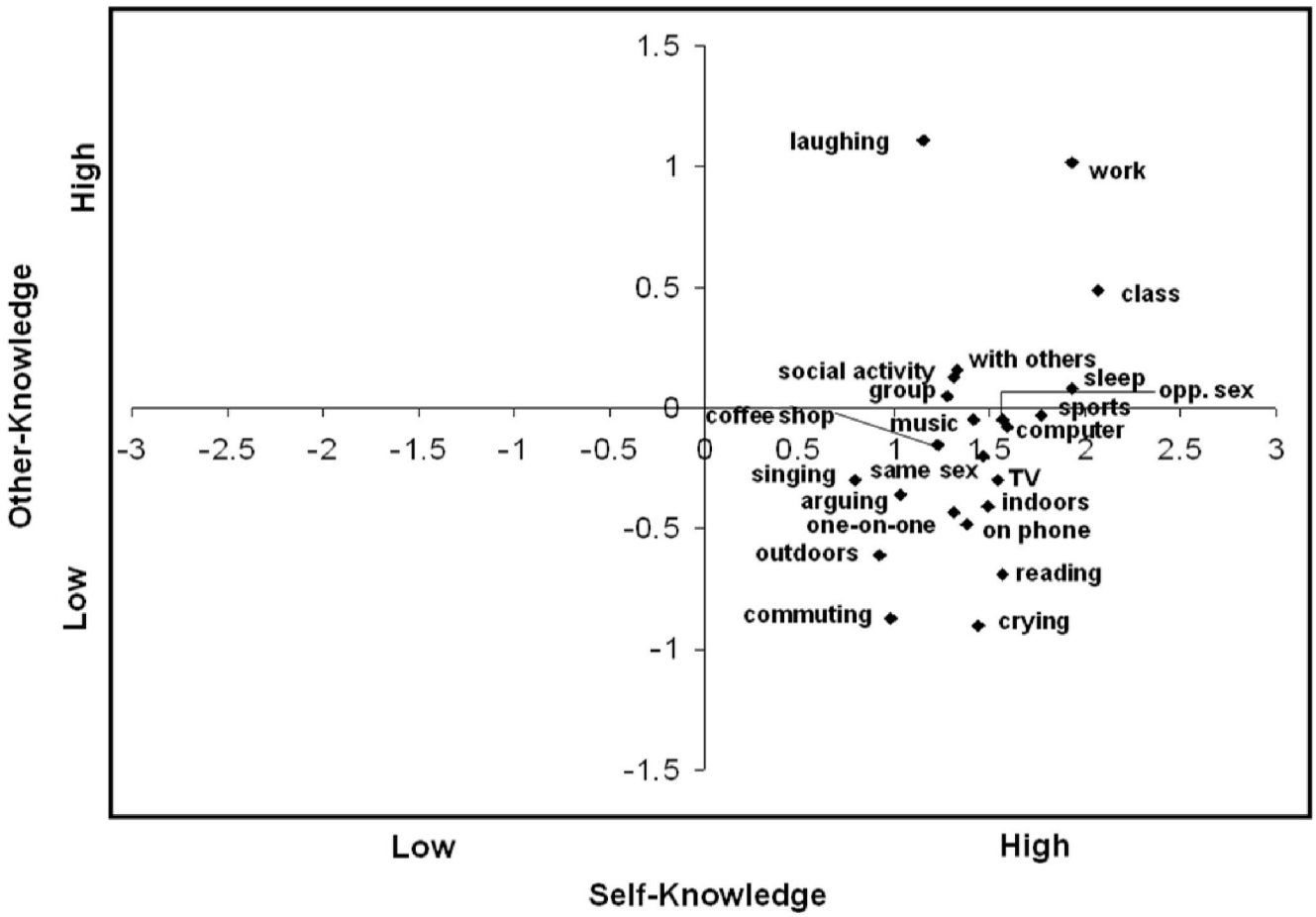


*Figure 2.* Lay predictions of self- and other-knowledge from Study 1. Self-knowledge is on the *x*-axis; other-knowledge is on the *y*-axis. All ratings were made on a 7-point Likert-type scale, and scores were centered on the midpoint (4). opp. = opposite.

in either the top right or bottom right quadrant of the Johari window. We, however, predicted that laypeople would consider most behaviors to be known to the self but not to others. Consistent with our predictions, lay perceptions about self- and other-accuracy for most behaviors fell into the bottom right quadrant (known only to self), with a few falling into the top right quadrant (known to self and others; Figure 2).

Overall, the data show that there is a strong assumption that people are their own best experts for all of the daily behaviors we examined. In Study 2, we proceed to empirically test the relative accuracy and unique predictive validity of self- and other-ratings of daily behavior.

## Study 2

### *Method*

*Phase I: Collection of self-reports.* Participants were 80 undergraduate students recruited mainly from introductory psychology courses and by flyers in the psychology department. The sample was 54% female (46% male). The ethnic makeup of the sample was 65% White, 21% Asian, 11% Latino, and 3% of another ethnicity. The majority of the participants (73%) were 18 years old, although their ages ranged from 18 years to 24 years ($M = 18.7$, $SD = 1.4$). Participants were compensated $50 for their participation. Seventy-nine participants completed the entire study. One participant dropped out of the study (i.e., did not come back for the follow-up) and also failed to provide valid EAR data. In addition, a few participants had sporadic missing data, so the effective sample size varies across analyses (as indicated in the results).

*Phase II: Collection of informant reports.* Each participant was asked to nominate three people who knew them well to provide ratings of their personality. Participants were asked to nominate their closest friend, one parent, and one romantic partner if possible. If participants could not provide any one of the three kinds of informants, they were told to nominate another person who knows them well. Participants were told that the informants' ratings would be kept completely confidential and that they themselves would never see their informants' ratings.

Informant ratings were collected via the Internet, following recommendations by Vazire (2006). Informants were contacted by e-mail and asked to complete an online questionnaire about how they see the target participants' personality. The online questionnaire included 110 items, mostly consisting of personality scales (e.g., the Big Five

Inventory) and the ACT. Altogether, the questionnaire took about 5–10 min to complete. Informants received a link and a unique identifying number in the email. Informants who did not complete the ratings were sent reminder e-mails after 2 weeks, 4 weeks, and 6 weeks. Participants were compensated at the end of the 3 weeks, regardless of whether the informants had completed their ratings. Informants were not compensated for their cooperation.

Two months after the end of the study, 76% of informants had completed the ratings, resulting in a total of 182 informant ratings, of which 87 were from friends, 55 from parents, and 21 from romantic partners. The remaining informant reports were from siblings (12), former-romantic partners (2), and one cousin, one grandmother, one great-grandmother, one former-friend, and one informant who did not indicate his or her relationship to the target. The informant ratings were aggregated to form a single, composite other-rating for each behavior. The average reliability of this composite was ICC [1, *k*] = .49 across the 20 behaviors, and the average pairwise informant agreement correlation was .32. The correlation between self-ratings and other-ratings are reported in Table 2 and averaged .38.

Five of the original 25 items were dropped from analyses in Study 2 for the following reasons: First, the first two ratings (time spent alone and time spent with others) were empirically redundant (average $r = -.72$) and were therefore subsequently aggregated into a single rating (time spent alone was reversed-scored). In addition, four behaviors were dropped due to the inadequacy of the criterion (i.e., the EAR codings; see below), thus leaving us with 20 behaviors.

Table 2
*Descriptive Statistics for Self- and Informant-Ratings of Daily Behavior: Study 2*

| | Self-rating | | Informant rating | | | |
|---|---|---|---|---|---|---|
| ACT behavior | *M* | *SD* | *M* | *SD* | Paired-sample *t* | Self-informant *r* |
| With other people | 4.14 | 1.39 | 4.60 | 1.20 | −2.82* | .37** |
| On the phone | 3.56 | 1.65 | 4.21 | 1.35 | −3.56** | .45** |
| Talking one-on-one | 4.92 | 1.22 | 4.90 | 1.14 | 0.11 | .10 |
| Talking in a group | 4.18 | 1.43 | 4.49 | 1.14 | -1.80 | .33** |
| Talking to same sex | 3.88 | 1.26 | 4.51 | 1.11 | −3.78** | .25* |
| Talking to opposite sex | 4.69 | 1.35 | 4.65 | 1.21 | 0.28 | .39** |
| Laughing | 5.41 | 1.29 | 5.24 | 1.12 | 1.08 | .33** |
| Singing | 4.60 | 1.60 | 3.91 | 1.44 | 3.20** | .23* |
| Crying | 3.19 | 1.67 | 3.06 | 1.25 | 0.69 | .49** |
| Arguing | 3.16 | 1.70 | 3.39 | 1.32 | −1.16 | .36** |
| Listening to music | 5.42 | 1.28 | 5.34 | 1.08 | 0.54 | .32** |
| Watching TV | 3.51 | 1.76 | 3.86 | 1.50 | −2.03* | .57** |
| On the computer | 5.15 | 1.53 | 5.13 | 1.32 | 0.17 | .48** |
| At work | 2.81 | 1.64 | 3.10 | 1.33 | −1.72 | .50** |
| Attending class | 4.88 | 1.21 | 5.17 | 1.18 | −2.29* | .56** |
| Socializing | 4.24 | 1.44 | 4.47 | 1.10 | −1.32 | .30** |
| Indoors | 4.68 | 1.10 | 4.52 | 0.94 | 1.13 | .27* |
| Outdoors | 3.92 | 1.11 | 3.60 | 0.94 | 2.51* | .39** |
| Commuting | 3.54 | 1.22 | 3.56 | 0.92 | −0.12 | .39** |
| At coffee shop/bar/ restaurant | 4.01 | 1.55 | 3.80 | 1.34 | 1.26 | .46** |

*Note.* For self-ratings, $N = 79$; for informant ratings, $N = 77$. Ratings were made on a 7-point Likert-type scale with the end points *much less than the average person* (1) *to much more than the average person* (7).
* $p < .05$, two-tailed, paired-sample *t* test, $df = 76$. ** $p < .01$, two-tailed, paired-sample *t* test, $df = 76$.

*Phase III: Assessment of daily behavior.* Behavior was measured with the electronically activated recorder (EAR; Mehl et al., 2001). The EAR is a digital audio recorder with an external microphone that records the ambient sounds of people's daily lives. The recorder can comfortably be carried in a pocket or purse, and the microphone, attached by a wire, can be worn on a shirt collar. The digital recorder was programmed to be on for 30 s every 12.5 min, producing 4.8 recordings per hour. Participants could not know when the recorder was on or off. For further details on the background of the method, see Mehl (2007).

Participants wore the EAR for four consecutive days during their waking hours. The monitoring was scheduled from Friday afternoon to Tuesday night to capture their behavior both during the week and over the weekend. Participants were thoroughly informed about the study's privacy and confidentiality policies, which included an opportunity to review and censor their recordings before releasing them to the experimenters. Participants were encouraged to wear the EAR as much as possible and to remove it only when its functioning was in jeopardy. When the participants returned to the lab to drop off the device, they completed a standard questionnaire assessing the perceived obtrusiveness of the device and their compliance with wearing it (Mehl & Holleran, 2007).

On average, over the 4-day monitoring period, the EAR provided 300 valid waking recordings ($SD = 104$) per participant, reflecting an overall high compliance. This compliance was corroborated by participants' self-reports; they reported having worn the EAR 72% of their time awake ($SD = 16\%$). Awareness of the EAR was rated on a 5-point scale (1 = *not at all*; 5 = *a great deal*) and yielded moderate ratings for the participants themselves ($M = 2.9$, $SD = 0.9$) and for the people around them ($M = 3.2$, $SD = 0.9$). Participants reported feeling only slightly constrained by the device ($M = 1.9$, $SD = 0.9$) and uncomfortable wearing it ($M = 2.1$, $SD = 1.0$). They further indicated that the recorder had only minor effects on their behavior ($M = 1.5$, $SD = 0.7$) and speech ($M = 1.3$, $SD = 0.6$) as well as the behavior of people around them ($M = 1.9$, $SD = 0.9$). Finally, research assistants coded how often the EAR was mentioned in conversations with others. On average, participants referred to the EAR in less than 3% of their recorded conversations ($M = 2.7$, $SD = 4.9$).

A team of 10 research assistants then coded acoustically detectable features of participants' behaviors with a revised version of the Social Environment Coding of Sound Inventory (SECSI; Mehl et al., 2006). The coding system was comprised of four major category clusters: the person's (a) 'current location (e.g., in apartment, outdoors, in transit), (b) activity (e.g., listening to music, on the computer, eating), (c) interaction (e.g., alone, on the phone, talking in person), and (d) mood (e.g., laughing, crying, sighing). Intercoder reliabilities for the SECSI categories were determined from a set of training EAR recordings (221 sound files) independently coded by all 10 research assistants. ICC[2,*k*] exceeded .70 for all categories except reading. Because of the low intercoder reliability ($r = .13$), reading was dropped from our analyses. In addition, three behaviors (playing sports, sleeping, and time spent eating) were dropped due to the fact that EAR did not adequately capture these behaviors. That is, the range and variance in behavior across participants was very small. In the case of playing sports and sleeping, this is likely due to the fact that participants were instructed to remove the EAR during these activities. For the remaining behaviors, the raw codings were converted into time-use

estimates by calculating the percentage of a person's valid (i.e., compliant and codable) waking EAR recordings in which a SECSI category applied (e.g., percentage of sound files in which the participant was talking, at home, or listening to music).

*Measures.* Participants and informants completed a battery of questionnaires, including the ACT. In this study, the raters were asked to indicate how much the target performs the given behavior compared with the average person. Responses were made on a 7-point Likert-type scale with the anchors *much less than the average person* (1) and *much more than the average person* (7). We chose to use a Likert-type scale rather than absolute behavior counts for several reasons. First, absolute behavior counts of common daily behaviors are hard to estimate (Sudman, Bradburn, & Schwarz, 1996). For example, it is very difficult to estimate what percentage of one's waking time one spends with others. Second, the EAR data are better suited to relative comparisons across participants than to absolute behavior counts.

*Procedure.* All self-report measures were obtained on the first day of the 3-week study. After consenting to participate, participants completed the informant nomination questionnaire, followed by a battery of self-report questionnaires, including the ACT. The self-report questionnaires were administered online, on one of the laboratory computers. Participants were seated in a private room with the door closed and were told to complete the questionnaires, taking as many breaks as they desired. Participants then started the EAR monitoring period immediately upon leaving this session, which was always on a Friday.

At the end of the EAR monitoring period (i.e., after Tuesday night), participants returned to the lab and completed a questionnaire regarding their experience with the EAR. Upon completion of the study, all participants were given the opportunity to listen to their EAR recordings and erase any parts they wished (fewer than 0.01% of the sound files were erased). Informants were contacted by e-mail shortly after participants completed the study and were asked to complete the on-line informant-report questionnaire.

Table 3
*Descriptive Statistics for Targets' EAR-Assessed Daily Behaviors: Study 2*

| ACT behavior | Intercoder reliability | Base rate | | | |
|---|---|---|---|---|---|
| | | Minimum | Maximum | *M* | *SD* |
| With other people | .97 | 3.6 | 67.3 | 33.5 | 14.9 |
| On the phone | .97 | 0.0 | 20.9 | 4.0 | 3.6 |
| Talking one-on-one | .94 | 1.3 | 21.8 | 9.1 | 4.9 |
| Talking in a group | .88 | 0.0 | 14.0 | 2.8 | 2.5 |
| Talking to same sex | .95 | 0.3 | 50.0 | 15.6 | 11.0 |
| Talking to opposite sex | .95 | 0.0 | 35.3 | 7.2 | 7.8 |
| Laughing | .89 | 0.6 | 28.4 | 8.1 | 5.9 |
| Singing | .74 | 0.0 | 20.8 | 2.8 | 3.3 |
| Crying | — | 0.0 | 2.1 | 0.1 | 0.3 |
| Arguing | — | 0.0 | 2.6 | 0.2 | 1.6 |
| Listening to music | .95 | 0.0 | 42.3 | 14.9 | 9.8 |
| TV on | .95 | 0.0 | 79.6 | 16.5 | 15.6 |
| On the computer | .87 | 0.0 | 42.2 | 7.0 | 8.8 |
| At work | — | 0.0 | 54.3 | 2.8 | 8.4 |
| Attending class | .99 | 0.0 | 12.5 | 4.1 | 3.4 |
| Socializing | .91 | 0.0 | 61.8 | 16.4 | 12.7 |
| Indoors | .88 | 13.8 | 90.8 | 60.9 | 15.8 |
| Outdoors | .90 | 0.0 | 12.6 | 3.6 | 2.4 |
| Commuting | .89 | 0.0 | 32.2 | 6.3 | 6.3 |
| At a coffee shop/bar/ restaurant | .91 | 0.0 | 18.9 | 2.7 | 3.3 |

*Note.* $N = 79$. Values presented are percentages of EAR files in which the respective behavior was present. Intercoder reliability coefficients indicate intercoder agreement and are intraclass correlations (ICC) based on a two-way random effects model, ICC[2,k]; intercoder agreement was computed from a training set of 221 EAR sound files independently coded by all 10 coders. Dashes indicate that reliability could not be determined due to lack of variance in the codings in the training set (i.e., the respective behavior was not present in any of the training sound files, and all coders correctly coded the behaviors as absent across all sound files).

## *Results and Discussion*

Accuracy was measured by correlating the self- and informant ratings of the targets' daily behavior with the EAR-assessed frequencies of their actual daily behavior. The descriptive statistics for the self- and other-ratings are presented in Table 2. The descriptive statistics for the EAR-derived behaviors are presented in Table 3.

*Question 1: Which perspective is more accurate overall?* In contrast to the ratings provided by the participants in Study 1, we predicted that the two perspectives would be equally accurate. Consistent with our prediction, the average accuracy across all 20 ACT behaviors was identical for the two perspectives ($r = .26$). Self-accuracy was statistically significant for 13 of the 20 behaviors (with an alpha level of .05, one tailed) and other-accuracy was statistically significant for 14 of the 20 behaviors (Table 4). This proportion is considerably higher than the number of significant correlations we would expect by chance alone (i.e., due to Type I error), namely, 1 out of 20.

In addition to computing the accuracy of the aggregated informant ratings, we also examined the accuracy of a single informant. This approach corrects for the fact that the reliability of the aggregated informants' ratings is boosted as a result of aggregation. Thus, we computed the accuracy of a single informant's rating for all of the participants who were rated by friends ($n = 65$). We chose to use friends because more participants had ratings from friends than from any other type of informant. We also felt that mixing types of informants in a single analysis would artificially reduce the accuracy of the single informant because of the important mean differences between types of informants (Vazire, Rentfrow, Mehl, & Gosling, 2005). If participants were rated by more than one friend, we used the rating from the friend who had known the participant the longest. The average accuracy of the single friend ratings across the 20 behaviors was .23—virtually the same as the average accuracy of the aggregate of three informant ratings ($r = .26$). Thus, the accuracy of other-ratings was not simply a result of the psychometric benefits of aggregation.

*Question 2: Do the two perspectives provide unique insight?* We predicted that the two perspectives would provide both overlapping and unique information and that the behaviors would fall into all four of the quadrants in the Johari window.

Table 4 presents the accuracy of ratings by the self, the aggregated informants, and the single friend. Consistent with our predictions, the accuracy correlations of self- and other-ratings ranged from 0 to .55 and covered all four quadrants of the model. This suggests that the two perspectives know about different behaviors. Unfortunately, due to our relatively small sample size, we had only limited power to detect reliable differences between self- and other-accuracy. To eval-

Table 4
*Accuracy of Self- and Informant-Ratings of Behavior: Correlations With EAR-Assessed Daily Behavior: Study 2*

| ACT behavior | Self | Aggregated informants | Single informant |
|---|---|---|---|
| With other people | .14 | .36** | .30** |
| On the phone | .37** | .40** | .32** |
| Talking one-on-one | −.06 | .25* | .22* |
| Talking in a group | .25* | .20* | .25* |
| Talking to same sex | .34** | .25* | .13 |
| Talking to opposite sex | .31** | .32** | .18† |
| Laughing | .23* | .25* | .13 |
| Singing | .34** | .29** | .34** |
| Crying | .18† | .16† | .19† |
| Arguing | .28** | −.05 | .09 |
| Listening to music | .40** | .34** | .26* |
| Watching TV | .55** | .39** | .36** |
| On the computer | .29** | .31** | .20† |
| At work | .25* | .35** | .22* |
| Attending class | .07 | .33** | .26* |
| Socializing | .18† | .30** | .27* |
| Indoors | .16† | .16† | .20† |
| Outdoors | .11 | .05 | .10 |
| Commuting | .27** | .16† | .14 |
| At a coffee shop/bar/restaurant | .27** | .15† | .24* |
| *M* | .26 | .26 | .23 |

*Note.* For self, $N = 79$; for aggregated informants, $N = 77$; for single informant, $N = 65$. Correlations are between ratings and EAR-coded behavior. Single informant ratings are based on friends' ratings.
* $p < .05$, one tailed. ** $p < .01$, one tailed. † $p < .10$, one tailed.

uate the statistical significance of the difference between these correlations, which share a variable, we used Hotelling's *t* test with the Williams modification (Kenny, Kashy, & Cook, 2006). Self-ratings were significantly more accurate than were aggregated other-ratings for arguing, $t(74) = 2.70$, $p = .01$ (*p*s two-tailed) and were slightly but not significantly more accurate than were other-ratings for watching TV, $t(74) = 1.79$, $p = .08$. Other-ratings were significantly more accurate than were self-ratings for talking one-on-one, $t(74) = 2.09$, $p = .04$, and attending a lecture, $t(74) = 2.60$, $p = .01$, and were marginally significantly more accurate than self-ratings for spending time with others, $t(74) = 1.83$, $p = .07$.

Note that there were also some commonalities across the two perspectives. For example, both were very accurate in predicting watching TV, talking on the phone, and listening to music, and both were inaccurate in predicting spending time indoors, outdoors, and crying. Overall, self-accuracy and other-accuracy were moderately, positively correlated with each other; that is, the vector correlation between the two columns of accuracy correlations (after these were transformed with Fisher's *r*-to-*z* formula) was positive and moderate in magnitude ($r = .31$). This is also reflected in Figure 3, which shows the accuracy of the two perspectives on all 20 ACT behaviors in the Johari window. The origin of the axes is set at $r = .20$ (the significance threshold for $p < .05$, one tailed) so that behaviors fall into the appropriate quadrants. That is, behaviors that were not rated accurately by the self or others (i.e., with accuracy correlations less than .20) should fall on the low end of self-knowledge and other-knowledge continua, respectively. Note that the behaviors fell into all four quadrants of the figure, though relatively few fell in the unknown (bottom left) quadrant.

To provide a more direct test of how much unique information each perspective provided, we ran multiple regressions with both self- and other-ratings. The results, presented in Table 5, reflect the unique contribution (i.e., beta weights) of each perspective above and beyond the overlap in the two perspectives. For example, both self- and other-ratings of listening to music contributed uniquely to predicting actual time spent listening to music. Both perspectives also contributed uniquely to accurate predictions of singing. Self-ratings provided unique accurate information for predictions of talking to people of the same-sex, arguing, watching TV, and commuting. Other-ratings provided unique, accurate information for predictions of spending time with others, talking one-on-one, talking on the phone, attending a lecture, working at a job, and participating in social activities. Another way to examine the unique predictive validity of each perspective is to enter one perspective (e.g., self) in the first step of a stepwise regression and examine the change in $R^2$ when the other perspective (e.g., informants) is added as a predictor in the second step. When self-ratings were entered in the first step of the regression, adding informant ratings in the second step significantly increased the $R^2$ for 7 of the 20 behaviors (on the phone, listening to music, singing, attending a lecture, at work, socializing, and talking one-on-one), and the average change in $R^2$ across the 20 behaviors was .04 ($p = .05$). When informant ratings were entered in the first step of the regression, adding self-ratings in the second step also significantly increased $R^2$ for 7 of the 20 behaviors (watching TV, listening to music, singing, talking to same-sex, arguing, at a coffee shop/bar/restaurant, and commuting), and the average change in $R^2$ across the 20 behaviors was also .04 ($p = .05$). These analyses provide further support for our prediction that the two perspectives are equally accurate and that both perspectives provide unique predictive validity.

## General Discussion

The purpose of these studies was to show that laypeople assume that the self is the best expert of what a person is like and to test the validity of this assumption in the context of daily behavior—arguably one of the most important criteria for what a person is really like.

Study 1 tested the hypothesis that many people intuitively assume that the self is by far the best expert on what a person is like. The lay predictions collected in our study clearly supported this hypothesis. Predictions from laypeople show that they expected the self to be more accurate than others for just about every behavior we assessed. According to these ratings, there is little if anything that close others know about people's behavior that people do not know themselves.

The findings from Study 2, which examined the actual accuracy of self- and other-ratings, tell a different story. Consistent with our own predictions, both perspectives were accurate for more than half the daily behaviors examined. Furthermore, each perspective provided unique validity in predicting the behaviors. In sum, contrary to the lay predictions obtained in Study 1, we showed that the two perspectives have both unique and overlapping domains of knowledge and that others do have some insight about what a person is like that is not captured by self-perceptions. Next, we review our findings and their theoretical implications in more detail.

### *Overview of Results and Implications*

This study expanded on existing studies (e.g., Fiedler et al., 2004; John & Robins, 1994; Gosling et al., 1998; Kolar et al.,

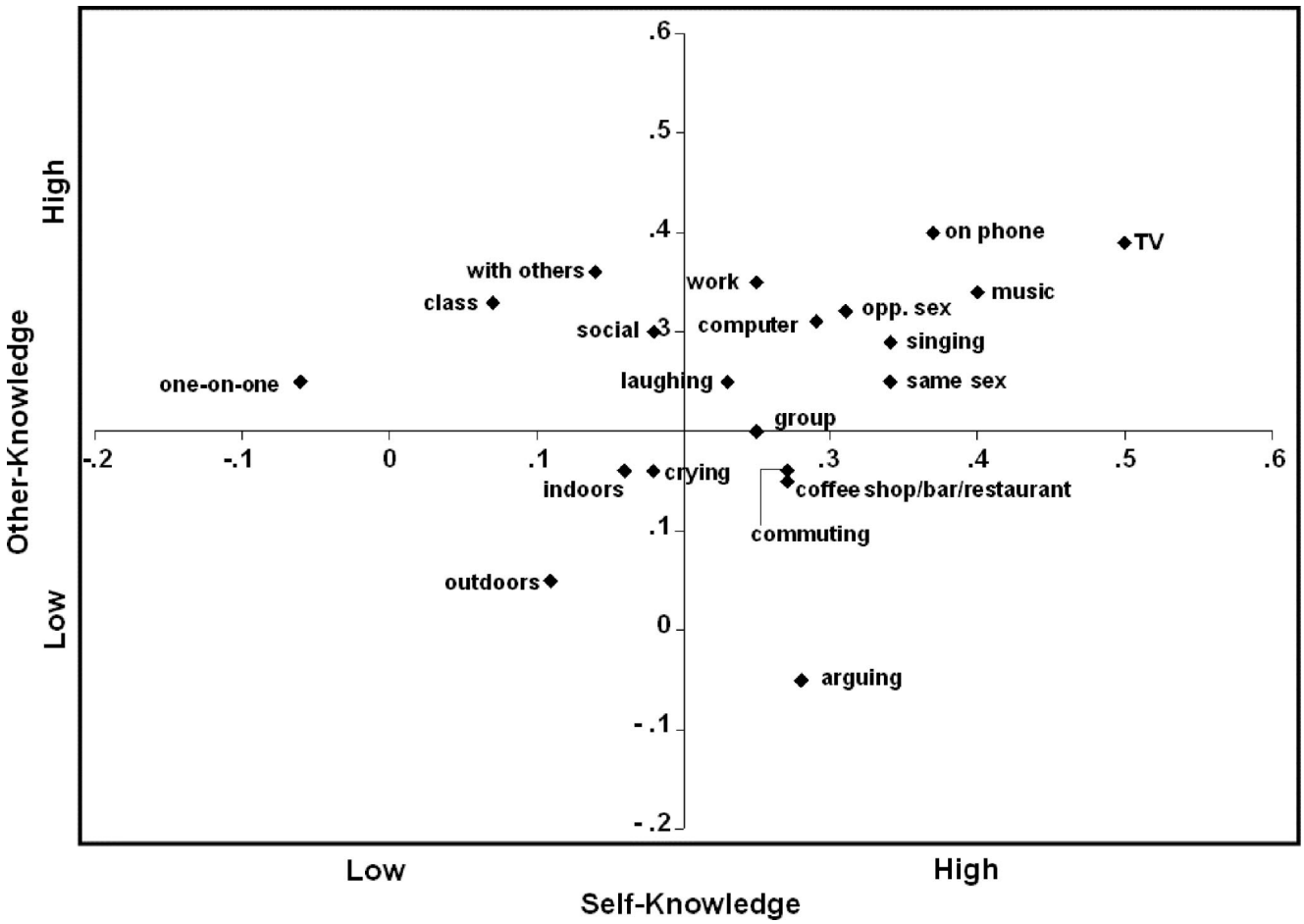


*Figure 3.* Accuracy of self- and other-ratings of behavior from Study 2. Self-accuracy ($r$) is plotted on the $x$-axis; other-accuracy ($r$) is on the $y$-axis. $r$s $\geq .20$ are significant at $p < .05$ (one tailed). The origin is set at $r = 20$, the threshold for significance, so that the behaviors that were rated with low or no accuracy fall into the unknown quadrants. opp. = opposite.

1996; Shrauger et al., 1996; Spain et al., 2000) in several ways. First, our criterion measure was based on unobtrusively observed real-world behavior rather than on momentarily assessed self-reports of behavior (e.g., via an experience sampling procedure). This allows for a more unbiased test of accuracy because the criterion measure has no shared method variance with either of the two predictors (self-ratings and other-ratings). Second, the daily behaviors we examined were sampled from the full spectrum of a person's daily life and aggregated over a period of 4 days, rather than being laboratory-based behaviors obtained in a single session. By following Brunswick's (1956) suggestions for a representative design, that is, by sampling evenly from people's natural settings, we were able to capture behavioral trends ecologically in the spirit of the act-frequency approach (Buss & Craik, 1983; 1985). Finally, another advantage of our method is that behavior was assessed in the environments that participants naturally select and create for themselves (Buss, 1987; Ickes, Snyder, & Garcia, 1997). Laboratory-based tests, in contrast, place people in situations that they may or may not have selected in their real lives.

The unique design of Study 2 allowed us to draw several conclusions. First, the fact that the behavior ratings were at all accurate provides more evidence that behavior is considerably consistent across situations (cf., Mischel, 1968). The fact that self- and other-ratings of how the targets typically behave predicted the targets' actual behavior over a 4-day period provides strong evidence that—despite high moment-to-moment variability—people do have noticeable and stable behavioral tendencies (Fleeson, 2001). Also, the magnitude of the multiple correlations (Table 5) are relatively large by today's standards (Hemphill, 2003; Meyer et al., 2001; Richard, Bond, & Stokes-Zoota, 2003). The accuracy levels obtained in Study 2 were particularly impressive given that the criterion measure was based on intermittent recordings over only 4 days—ultimately still a relatively small window of time (Epstein, 1979). Thus, any accuracy on the part of the self or the others is impressive, and large effect sizes cannot be expected.

Another important finding is that there were sizeable differences in the magnitude of the accuracy correlations across the 20 ACT behaviors. In some cases, this may reflect limitations of the EAR method for reliably detecting certain behaviors (e.g., low base-rate behaviors). However, for other behaviors it probably reflects genuine differences in the actual predictability of the behaviors. Due to the dearth of research on actual daily behavior, it would be

Table 5
*Standardized Beta Weights for Regression of EAR-Assessed Daily Behavior on Self- and Informant-Ratings of Daily Behavior: Study 2*

| | Self | | Aggregated informants β | | |
|---|---|---|---|---|---|
| ACT behavior | β | $\Delta R^2$ | β | $\Delta R^2$ | Multiple *R* |
| Watching TV | .50** | .17** | .11 | .01 | .57** |
| On the phone | .23* | .04* | .30* | .07* | .46** |
| Listening to music | .32** | .09** | .23* | .05* | .46** |
| With other people | .11 | .01 | .37** | .13** | .41** |
| Singing | .29* | .08* | .22* | .05* | .40** |
| Talking to same sex | .29* | .08 | .18 | .03 | .38** |
| Attending lecture | −.17 | .02 | .45** | .14** | .38** |
| Talking to opposite sex | .21† | .04† | .23† | .05 | .37** |
| On the computer | .18 | .02 | .22† | .04† | .35** |
| At work | .00 | .00 | .35** | .09** | .35** |
| Socializing | .13 | .02 | .27* | .06* | .33* |
| Arguing | .33** | .10** | −.17 | .03 | .31* |
| At a coffee shop/bar/ restaurant | .27* | .06 | .02 | .00 | .28* |
| Laughing | .14 | .02 | .20† | .04† | .28* |
| Talking in a group | .20† | .04† | .13 | .02 | .28† |
| Commuting | .25* | .05 | .07 | .00 | .28† |
| Talking one-on-one | −.08 | .01 | .26* | .07* | .27† |
| Crying | .16 | .02 | .08 | .01 | .21 |
| Indoors | .10 | .01 | .13 | .02 | .19 |
| Outdoors | .10 | .01 | .01 | .00 | .11 |
| *M* | .18 | | .18 | | .33 |

*Note.* $N = 77$. For each behavior, self- and aggregated informant-ratings were entered simultaneously into a regression predicting actual EAR-coded behavior, and standardized betas are reported here. $R^2$ change values are presented in parentheses.
* $p < .05$, two tailed. ** $p < .01$, two tailed. † $p < .10$, two tailed.

premature to propose a model that aims at systematically explaining these differences across behaviors. In the future, researchers should address this issue by examining the dimensions along which behavior varies and testing these dimensions as potential moderators of self- and other-accuracy.[2]

Most interesting, for our purposes, are the differences in accuracy between the two perspectives. The main goal of the study was to examine the relative accuracy and unique predictive validity of the two perspectives. That is, do self and others each have privileged knowledge about what a person is like, or can this be captured entirely by self-ratings? Our results provide strong evidence that the two perspectives each provide unique information. Perhaps the most important conclusion that can be drawn from this finding is that contrary to lay perceptions and many researchers' beliefs, the self is not always more accurate than others are, and others often provide valuable information that is not captured by self-ratings. The fact that there were any behaviors that others predicted better than the self and that other-ratings added incremental validity to self-ratings for many of the behaviors should cause researchers to carefully consider the limitations of self-reports and the usefulness of obtaining multiple perspectives on what a person is like.

The accuracy of the informants' ratings is all the more impressive when one considers that the informants were not necessarily people who live in the same town or interact with the target regularly. Many of the informants were family members and hometown friends (all participants were college students enrolled in summer school for the duration of the study). For them, the behavior ratings were inevitably based on impressions of the target in a context other than the one in which the behavior was measured (Kenny, 1994, 2004). For example, parents likely based their ratings on how the participants behaved at home.

The design of the study also provided the self with another potential advantage over informants: participants had the potential to control their own behavior. Certainly, local informants who interacted with the targets also had some opportunity to influence the target's behavior. Yet, ultimately, the participants themselves had the final word in how they behaved. Methodologically, this gives the self-ratings a slight leg-up in terms of accuracy. In theory, participants could have brought about accuracy by bringing their behavior over the next few days in line with the self-views they stated prior to the EAR monitoring (Swann & Read, 1981; Swann, Rentfrow, & Guinn, 2003). They could also have exerted a more subtle influence on their behavior by selecting environments or evoking behaviors in others that confirmed their self-views (Buss, 1987).

In short, the findings from this study show that the self and others both have unique insight into what a person is like. The two perspectives are about equally accurate, and each provides unique predictive validity about how a person behaves in daily life.

### *Limitations*

Despite its innovative design, the study has several potential limitations. The most important limitation concerns the restriction in the spectrum of daily behaviors measured. Technical, practical, and ethical constraints limited us to (a) assessing only acoustically detectable behaviors, (b) assessing behavior over a period of only four days, and (c) only assessing behaviors that participants were willing to have recorded and analyzed. These restrictions almost certainly compromised the content validity of our criterion measure. That is, there are a host of daily behaviors that were not covered by our 20-item ACT measure. Furthermore, the EAR method lends itself most immediately to identifying behaviors at a lower, molecular level (e.g., laughing, talking). As a result, our study could not capture many behaviors that are considered higher level acts (e.g., helping others, showing respect, acting competitively) within Vallacher and Wegner's (1987) action identification theory. Even though many of the current set of EAR behaviors are psychologically nontrivial (e.g., arguing, watching TV), it may often be more important to predict higher-level, molar actions rather than lower-level, molecular actions. Had we been able to capture a broader range of behaviors—both vertically and horizontally—we may have been able to provide a more detailed account of self- and other-knowledge, as well as analyses of moderators of self-

[2] We did attempt to examine three potential moderators of self- and other-accuracy: desirability, observability, and automaticity of the behavior. Specifically, eight research assistants who were not participants in either study rated the ACT behaviors on these dimensions (alphas were .93, .88, and .81, respectively), and their dimensional ratings were then correlated with the self- and other-accuracy correlations (transformed with Fisher's *r*-to-*z* formula) obtained in Study 2. The only significant finding was that the desirability of the behavior was positively associated with other-accuracy. However, we believe that these analyses are inconclusive due to the restricted number and range of behaviors examined in this study and should be the subject of a more comprehensive investigation in future research.

and other-accuracy. As technology improves and we become able to capture more behaviors in more varied contexts and across both lower and higher levels of identification, the landscape of self- and other-knowledge can be filled in with greater depth and precision.

Another limitation is that self- and informant ratings were made in comparison with the average person. This vague instruction leaves open the possibility that raters were imagining different average persons (i.e., idiosyncratic reference groups; cf. Heine, Lehman, Peng, & Greenholtz, 2002) or that errors in prediction are due to misestimating the average person's behavior rather than the target's behavior. Both of these phenomena would attenuate the accuracy of the ratings, thus making our results an underestimate of the actual accuracy of self- and informant ratings. It would be interesting to test these possibilities in future research, for example, by asking raters to give absolute frequency estimates of each behavior for both the target person and the average person.

It is interesting to note that the means of self- and other-ratings on the ACT questionnaire in Study 2 (Table 2) suggest that people may have had misconceptions about how the average person behaves. Specifically, several of the variables had means that were considerably far from the midpoint, either above or below, indicating that most people rated the target as above average or below average, respectively. In general, the means tend to show a positivity bias, or above-average effect. For example, self- and other-ratings were well above the midpoint of the scale for laughing and slightly below the midpoint for crying and arguing. Contrary to positive illusions theory, however, the bias was about as strong for other-ratings as for self-ratings. Nevertheless, these biases may indicate that people were not very good at estimating how the average person behaves, which in turn could have adversely affected accuracy levels.

A broader concern may be our claim that act frequencies of daily behaviors capture what a person is like. Most important, this definition of what a person is like misses the subjective, phenomenological, and coherent experience of being an individual (e.g., McAdams, 1995). We agree that these internal aspects of personality are also central to what a person is like, but they are not easily studied in this context. To set up a fair comparison of self- and other-accuracy (i.e., to avoid the possibility that shared method variance preemptively gives an undue advantage to one perspective), it would be necessary to assess these internal aspects objectively, that is, without having to rely on participants' explicit self-reports. For subjective personality phenomena that are deliberate in nature, this is by definition impossible. For subjective personality phenomena that are automatic in nature, however, it would potentially be possible to use modern implicit measures as objective accuracy criteria (e.g., Asendorpf, Banse, & Mücke, 2002). However, it is likely that even internal traits (e.g., positive affect) manifest themselves in observable behavior (e.g., laughing), so this aspect of a person is captured, albeit indirectly, in daily behavior.

Another problem with our criterion measure is that it could be argued that it has some method overlap with informant reports. That is, both informant ratings and behavior codings could be considered observer reports, and some of the correspondence between informants' ratings and behavior codings could be due to this shared method variance.

In response to this criticism, we argue that informant ratings are quite different from behavior codings. Although both are made from an outsider's perspective, the similarities end there. Informants were very well acquainted with the targets, and it is unlikely that when completing their ratings, they were relying on specific instances of the target's behavior. Indeed, what makes informants so valuable as judges of personality is that their impressions are presumably based on a great deal of experiences with the target, which they mentally interpret and aggregate (Vazire, Gosling, Dickey, & Schapiro, 2007). As Martin and Bateson (1993) put it, an informant making a rating "has played an active role in filtering, accumulating, weighting, and integrating information over a considerable period of time" (p. 81). Thus, we suspect that informant ratings are methodologically far more similar to self-ratings than either are to behavior codings.

A final limitation of this article is that it is unclear to what extent our findings generalize beyond U.S. college student samples. For example, the view that the self occupies a privileged position for predicting behavior (Study 1) may be more pronounced in individualistic cultures than in collectivistic cultures (Heine, 2007). Also, the Study 1 participants were all acquaintances of undergraduate students majoring in psychology and so may have held particularly strong beliefs about the capacity for self-insight. It would be interesting to test whether this bias exists equally among other populations and cultures that are habitually inclined to attend to close others. Another potential concern about generalizability is that self-knowledge about behavior may increase with age, as people have more opportunities to observe their own behavior and compare themselves with others. If this is the case, our finding that self-reports of behavior are no more accurate than other-reports (Study 2) may not hold among older adults.

### *Future Directions*

The findings from this article raise a host of questions about self- and other-knowledge. Perhaps the first question that comes to mind is why is the self not more accurate than others? One potential explanation is that the self does in fact possess greater knowledge about daily behavior than others do but that this knowledge was not fully tapped in our study. Indeed, we believe it is very likely that the levels of accuracy obtained in this study do not reflect the maximum accuracy possible. For example, it is possible that participants lacked the motivation or cognitive resources to provide the most accurate ratings possible at the time they completed their ratings. In addition, our study asked raters to describe how a person typically behaves rather than specifically asking them to predict future behavior or recall past behavior over a specific time period. It is possible that specifying the temporal perspective (retro- or prospective) could improve the accuracy of behavior ratings. However, there is no reason to believe these issues are more likely to affect self-ratings than other-ratings. Indeed, we suspect that both self and others ultimately possess greater knowledge than demonstrated in our study. For example, providing a reward for accurate ratings may increase the accuracy of self- and other-ratings. Future research should examine the conditions that promote accuracy to allow us to tap into the full extent of self- and other-knowledge (Funder, 1999).

As mentioned above, future research should also examine the characteristics of behaviors that may moderate self- and other-knowledge. For example, self-knowledge may be greater for deliberate than for automatic behaviors, whereas the opposite may be true for other-knowledge. We could not adequately test this question in our study due to the limited range of behaviors examined.

However, we hope that as measuring behavior becomes more and more practical, and as researchers are able to capture a broad range of behaviors that vary in intentionality and on other dimensions, future studies will shed light on this issue.

Another question for future research is which kinds of others are the most knowledgeable? Our studies lump together a variety of close others, including friends, romantic partners, and family members. However, it is likely that each of these perspectives has different knowledge about what a person is like. Is there one perspective that is most accurate overall? Or is each perspective an expert at predicting some behaviors but not others? We suspect that the latter is the case and that consistent with research on pragmatic or circumscribed accuracy (Gill & Swann, 2004; Swann, 1984), each perspective is most accurate for those behaviors that are most important in the context of their relationship with the target (cf. Kraemer et al., 2003). If this is the case, it is possible that our findings underestimate the actual amount of other-knowledge by aggregating different types of informants. Practically speaking, this would mean that accuracy could be greatly increased by use of the best expert for each kind of behavior.

### *Conclusion*

The results from these studies show that contrary to the strong intuitions of most, people are not always their own best expert. We suspect that the findings challenge many people's assumptions about the privileged position of the self when it comes to knowing what a person is like. Rather than merely dealing another blow to self-knowledge, however, we see these findings as conveying an important positive message: A person's close friends and family know that person surprisingly well and can tell things about that person that he or she did not know before.

It is our hope that these findings have an important impact on how people think of the self and of others. As researchers, we should take seriously the possibility that the combination of multiple well-acquainted informants can (and often will) give us more valid assessments than can self-ratings (Hofstee, 1994) and that predictive validity can be maximized by exploiting both sources of information. However, perhaps the more powerful conclusion we can draw from this study is that all of us should take seriously the possibility that others who know us well may (and almost always will) see aspects of our personality and behavior that we simply do not see in ourselves. Listening to what our informants have to say about us—provided of course that they are willing to give honest feedback—could have important benefits for improving self-knowledge (Wilson, 2002). For example, it may be useful to draw on close others' impressions of us when trying to understand ourselves, when predicting our own future behavior, or even when making major life decisions (such as what career path to choose). Although these ideas are still speculative, the findings reported here provide grounds for exploring these possibilities further.

## References


Agnew, C. R., Loving, T. J., & Drigotas, S. M. (2001). Substituting the forest for the trees: Social networks and the prediction of romantic relationship state and fate. *Journal of Personality and Social Psychology, 81,* 1042–1047.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83,* 380–393.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462–479.

Bargh, J. A., & Williams, E. L. (2006). The automaticity of social life. *Current Directions in Psychological Science, 15,* 1–4.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2,* 396–403.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54,* 579–616.

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

Buss, D. M. (1987). Selection, evocation, manipulation. *Journal of Personality and Social Psychology, 53,* 1214–1221.

Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review, 90,* 105–126.

Buss, D. M., & Craik, K. H. (1985). Why not measure that trait? Alternative criteria for identifying important dispositions. *Journal of Personality and Social Psychology, 48,* 934–946.

Conner, T., Barrett, L. F., Tugade, M. M., & Tennen, H. (2007). Idiographic personality: The theory and practice of experience sampling. In R. W. Robins, R. C. Fraley, & R. Kreuger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 79–96). New York: Guilford Press.

Craik, K. H. (1993). Accentuated, revealed, and quotidian personalities. *Psychological Inquiry, 4,* 278–289.

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

Epley, N., & Dunning, D. (2006). The mixed blessings of self-knowledge in behavioral prediction: Enhanced discrimination but exacerbated bias. *Personality and Social Psychology Bulletin, 32,* 641–655.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37,* 1097–1126.

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

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80,* 1011–1027.

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

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

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside behavioral q-sort: A tool for the description of social behavior. *Journal of Personality, 68,* 451–489.

Furr, R. M., & Funder, D. C. (2007). Behavioral observation. In R. Robins, C. Fraley, & R. Krueger (Eds). *Handbook of research methods in personality psychology* (pp. 273–291). New York: Guilford Press.

Gill, M. J., & Swann, W. B., Jr. (2004). On what it means to someone: A matter of pragmatics. *Journal of Personality and Social Psychology, 86,* 405–418.

Goodwin, M. S., Velicer, W. F., & Intille, S. S. (2008). Telemetric monitoring in the behavior sciences. *Behavior Research Methods, 40,* 328–341.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with online codings by observers. *Journal of Personality and Social Psychology, 74,* 1337–1349.

Hasler, B. P., Mehl, M. R., Bootzin, R. R., & Vazire, S. (in press).

Prelimary evidence of diurnal rhythms in everyday behaviors associated with positive affect. *Journal of Research in Personality.*

Heine, S. J. (2007). Culture and motivation. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 714–733). New York: Guilford.

Heine, S., Lehman, D., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales: The reference-group effect. *Journal of Personality and Social Psychology, 82,* 903–918.

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58,* 78–79.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality, 8,* 149–162.

Ickes, W., Snyder, M., Garcia, S. (1997). Personality influences on the choice of situations. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 165–195). New York: Academic Press.

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

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66,* 206–219.

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

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review, 8,* 265–280.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis.* New York: Guilford Press.

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

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160,* 1566–1577.

Kruglanski, A. W. (1989). The psychology of being right: The problem of accuracy in social perception and cognition. *Psychological Bulletin, 106,* 395–409.

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

Lucas, R. E., & Baird, B. M. (2006). Global self-assessment. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 29–42). Washington, DC: American Psychological Association.

Luft, J., & Ingham, H. (1955). The Johari window, a graphic model of interpersonal awareness. *Proceedings of the Western Training Laboratory in Group Development.* Los Angeles: University of California, Los Angeles.

Martin, P., & Bateson, P. (1993). *Measuring behavior: An introductory guide.* Cambridge, England: Cambridge University Press.

McAdams, D. (1995). What do we know when we know a person? *Journal of Personality, 63,* 365–395.

McAvay, G., Raue, P., & Brown, E. (2005). Symptoms of depression in older home-care patients: Patient and informant reports. *Psychology and Aging, 20,* 507–518.

Mehl, M. R. (2006). The lay assessment of subclinical depression in daily life. *Psychological Assessment, 18,* 340–345.

Mehl, M. R. (2007). Eavesdropping on health: A naturalistic observation approach for social-health research. *Social and Personality Psychology Compass, 1,* 359–380.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90,* 862–877.

Mehl, M. R., & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the electronically activated recorder (EAR). *European Journal of Psychological Assessment, 23,* 248–257.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84,* 857–870.

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The electronically activated recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers, 33,* 517–523.

Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science, 317,* 82.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001, July 6). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56,* 128–165.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

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

Ozer, D., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57,* 401–421.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic bias in self-perceptions: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66,* 2024–2060.

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

Pronin, E., Kruger, J., Savitsky, K., & Ross, L. (2001). You don't know me, but I know you: The illusion of asymmetric insight. *Journal of Personality and Social Psychology, 81,* 639–656.

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of bias blind spot. *Journal of Experimental Social Psychology, 43,* 565–578.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7,* 331–363.

Roberts, B. W., Kuncel, N., Shiner, R. N., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socio-economic status, and cognitive ability for predicting important life outcomes. *Perspectives in Psychological Science, 2,* 313–345.

Robins, R. W., & John, O. P. (1997a). Effects of visual perspective and narcissism on self-perception: Is seeing believing? *Psychological Science, 8,* 37–42.

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

Rohrbaugh, M. J., Shoham, V., Coyne, J. C., Cranford, J. A., Sonnega, J. S., & Nicklas, J. M. (2004). Beyond the 'self' in self-efficacy: Spouse confidence predicts patient survival following heart failure. *Journal of Family Psychology, 18,* 184–193.

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science, 3,* 102–116.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 209–269). San Diego, CA: Academic Press.

Shrauger, J. S., Ram, D., Greninger, S. A., & Mariano, E. (1996). Accuracy of self-predictions versus judgments by knowledgeable others. *Personality and Social Psychology Bulletin, 22,* 1229–1243.

Smith, T. W., Uchino, B. N., Berg, C. A., Florsheim, P., Pearce, G., Hawkins, M., & Hopkins, P. N. (2007). Hostile personality traits and coronary artery calcification in middle-aged and older married couples: Different effects for self-reports versus spouse ratings. *Psychosomatic Medicine, 69,* 441–448.

Spain, J. S., Eaton, L. G., & Funder, D. C. (2000). Perspectives on personality: The relative accuracy of self versus others for the prediction of emotion and behavior. *Journal of Personality, 68,* 837–867.

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco, CA: Jossey-Bass.

Swann, W. B., Jr. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review, 91,* 457–477.

Swann, W. B., Jr., & Read, S. J. (1981). Self-verification processes: How we sustain our self-conceptions. *Journal of Experimental Social Psychology, 17,* 351–372.

Swann, W. B., Jr., Rentfrow, P. J., & Guinn, J. (2003). Self-verification: The search for coherence. In M. Leary and J. Tagney, *Handbook of self and identity*. New York: Guilford Press.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193–210.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94,* 3–15.

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

Vazire, S., Gosling, S. D., Dickey, A. S., & Schapiro, S. J. (2007). Measuring personality in nonhuman animals. In R. W. Robins, R. C. Fraley, and R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 190–206). New York: Guilford Press.

Vazire, S., Rentfrow, P. J., Mehl, M. R., & Gosling, S. D. (2005, January). *The multiple faces of reputation.* Paper presented at the meeting of the Society for Personality and Social Psychology, New Orleans, LA.

Webb, E. T., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences.* Boston: Houghton Mifflin.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Belknap Press/Harvard University Press.

Wilson, T. D., & Dunn, E. W. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annual Review of Psychology, 55,* 493–518.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345–411). San Diego: Academic Press.

## Appendix

## ACT

[This is the self-report version—the actual questionnaire was completed online.]

Compared to other people, how much do you do the following activities? Write a number next to each activity.

Much ***less*** than the average person 1 2 3 4 5 6 7 Much ***more*** than the average person

1. ____ Spend time by yourself
2. ____ Spend time with others
3. ____ Talk on the phone
4. ____ Talk with someone one-on-one
5. ____ Talk with people in groups (with more than just one other person)
6. ____ Talk with people of your own sex
7. ____ Talk with people of the other sex
8. ____ Laugh
9. ____ Sing or whistle
10. ____ Cry
11. ____ Argue or fight
12. ____ Listen to the radio or music
13. ____ Watch TV
14. ____ Spend time on the computer
15. ____ Read
16. ____ Work (at a job)
17. ____ Spend time eating (not the amount eaten, but the time)
18. ____ Attend class
19. ____ Spend time doing entertaining things (e.g., going to the movies, to a sporting event, playing arcade games)
20. ____ Sleep
21. ____ Spend time in a house or apartment (any, not just your own)
22. ____ Spend time outside
23. ____ Spend time in a car or bus
24. ____ Go to coffee shops, bars, or restaurants
25. ____ Exercise or play sports

Received December 21, 2007
Revision received June 2, 2008
Accepted June 13, 2008 ■