HHS Public Access

Author manuscript

J Pers Soc Psychol. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

J Pers Soc Psychol. 2015 December ; 109(6): 1090–1104. doi:10.1037/a0039517.

TraitEnactments as Density Distributions: The Role of Actors, Situations, and Observers in Explaining Stability and Variability

William Fleeson and **Mary Kate Law**

Wake Forest University

William Fleeson: FleesonW@wfu.edu

Abstract

The purposes of this paper were to determine (i)whether the high consistency of individual differences in average aggregated behavior is due to actors' personalities or to consistency in the situations those actors encounter; and(ii)whether the surprisingly high within-person variability in trait enactment is a real phenomenon corroborated by observers, or merely in individuals' heads. Although traits are supposed to describe what individuals are like in everyday life, little evidence exists about the enactment of trait content in everyday life. Past experience-sampling studies have found both highly variable and highly consistent trait enactment, but were restricted to self-report data and to naturally occurring situations. The current study used experience-sampling in controlled lab environments with 97 targets and 183 observers to address these shortcomings. Targets attended hour-long lab sessions 20x each and observers rated targets' behavior. Parameters of distributions were highly consistent (r's ~ .80), revealing that actors were responsible for consistency, not situations. Nonetheless, observer ratings revealed that most variability in trait enactment was within-person, confirming that even when people put it on the line in ways that affect others, they still varied rapidly in the traits they enacted. In the face of two historically vexing objections to traits, this paper supports the density distributions model of traits and argues that trait conceptualizations must accommodate large within-person variability.

The purpose of this paper is twofold. The first purpose is to discover whether the demonstrated strong consistency in average behavior is mainly due to the powerful effects of situations on behavior (e.g., Fleeson, 2007; Ross & Nisbett, 2011), or is rather largely a contribution of the actor and his or her personality. That is, do people act the same way from week to week because of who they are or because they repeat the same situations? The second purpose is to test whether large within-person variability (inconsistency) in trait enactment is evident in the observable aspects of enactment reported by observers. That is, do people change their outward behavior from moment to moment as much as they change their inward behavior from moment to moment?

There are at least three reasons we believe these purposes may be important. First, because single behaviors are only moderately consistent, the high consistency of average behaviors has been essential for the validity of the trait approach (Epstein, 1979; Fleeson & Noftle, 2009). However, a primary objection to the apparently high consistency of average

behaviors is that relevant studies have not controlled for situational influence. If people differ in their everyday situations (Sherman, Nave, & Funder, 2010; Srivastava, Angelo, & Vallereux, 2008), and if situations predict trait enactment (Fleeson, 2007; Fournier, Moskowitz, & Zuroff, 2008; Furr & Funder, 2004; Leikas, Lönnqvist, & Verkasalo, 2012), then consistency may be largely due to situations. To solve this historically irrefutable objection, we propose a novel method to test the consistency of average behavior.

Second, there is a long-standing and recently intensified interest in within-person variability in personality-related constructs (Baird, Le, & Lucas, 2006; Clifton & Kuper, 2011; Côté, Moskowitz, & Zuroff, 2012; Distel et al., 2012; Erickson, Newman, & Pincus, 2009; Fleeson, 2001; Fleisher, Woehr, Edwards, & Cullen, 2011; Kiene, Tennen, & Armeli, 2008; Larsen, Augustine, & Prizmic, 2009; Mendoza-Denton & Ayduk, 2012; Minbashian, Wood, & Beckmann, 2010; Molloy, Ram, & Gest, 2011; Snyder, 1987). Large variability in personality-related concepts reinforces the need for social cognitive explanations of behavior (Mischel, 2004). However, self-reported adjectives may have overstated variability. Self-reported adjectives may not be grounded in concrete behaviors and they may include the potentially more pliable cognitive and affective aspects of enactment. It is unknown how much within-person variability exists in the aspects of enactment observable to independent others.

Third, although traits are intended as accounts of how people act in everyday life, the implications of trait ascriptions for how people act in everyday life are not known (Baumeister, Vohs, & Funder, 2007; Funder, 2001; Jackson et al., 2010). E.g., what does ascribing extraversion to a person imply about how extraverted the person acts in everyday life? The density distributions model (Fleeson, 2001) suggested that trait ascriptions refer to distributions of frequencies of acting at each level of the trait. Although the model has received some support (e.g., Fleeson & Gallagher, 2009), these two outstanding concerns about situational influence and self-reported data constrain its impact. Therefore, it is important to test the density distributions model using different data and study designs.

Actor Contribution to Consistency

The contribution of actors to their behavior remains doubted because of the alternative situation-based explanation for behavior. The doubt is important because consistency of behavior is important for inferring the existence of broad traits (G. W. Allport, 1937; Eysenck & Eysenck, 1985; Fleeson & Noftle, 2008). If a person acts similarly across situations and differently from others, then almost certainly something internal to the person must account for behavior. This internal factor is the actor contribution to behavior. For example, if some people act extraverted across situations whereas other people act introverted across situations, then something internal to those people is accounting for their introverted or extraverted behavior. Behavioral consistency has been challenged, because consistency of single behaviors is much lower than researchers expected (*r*'s of .1-.3, Mischel, 1968).

Epstein (1979) however, discovered that aggregated behavior averages are very consistent. Aggregated behavior averages are averages of a person's behavior across several occasions

of behavior. These behavior averages are much more consistent than are single behaviors (Epstein, 1979; Fleeson & Gallagher, 2009; Hartshorne & May, 1928; Rushton, Brainerd, & Pressley, 1983). Aggregated behavior averages calculated across just a few days are highly consistent with aggregated behavior averages calculated across a different few days, often with r's in the .90 range (Epstein, 1979; Fleeson, 2001; Fournier et al., 2008; Watson & Tellegen, 2002). For example, people who act introverted on average in one week are highly likely to act introverted in the next week. (G. W. Allport, 1937; Eysenck & Eysenck, 1985; Fleeson & Noftle, 2008).

This aggregated evidence shows that people act similarly to themselves across situations and differently from others. Thus, almost certainly something internal to the person must account for behavior (Fleeson, 2012). This internal factor is the actor contribution to behavior. Note that the fact of an actor contribution to behavior does not reveal the source or nature of the actor contribution; it reveals only that the actor contribution is something that differs across people and that is powerful in determining behavior. It could be a direct behavioral effect, such that people simply act differently across different situations. However, it could be due to a tendency to interpret different situations similarly, to evoke similar behaviors from others, or to transform situations (Buss, 1987; Snyder, 1983). For example, people who act agreeably from week to week may do so because that is the way they prefer to act, because they interpret situations as calling for agreeable behavior, because they evoke friendly behavior from others, or because they transform situations into agreeable ones. In all such cases, the actor contribution would lead to people acting differently from each other but similarly to themselves over time and across situations.

Alternative #1: Different People Have Different Situations

There is a significant outstanding objection to this interpretation, which undermines the conclusion that consistency of aggregated behavior averages represents an actor effect. Namely, consistency of aggregated behavior may be due to personality-independent differences in situations rather than to the actor's personality. Personality-independent differences in situations are differences between people in the objective or consensually interpreted situations they encounter. For example, people who act extraverted on average from week to week may do it because, for no reason of their own, they are in situations that objectively or consensually call for extraversion; if they were moved to situations calling for introversion, they might act introverted from week to week.

Situations influence behavior (Fleeson, 2007; Furr & Funder, 2004; Mendoza-Denton & Ayduk, 2012; Ross & Nisbett, 2011; Smith, Shoda, Cumming, & Smoll, 2009; Vansteelandt & Van Mechelen, 2004; Zayas, Shoda, & Ayduk, 2002), and recent studies have shown that people systematically differ in their daily situations (Sherman et al., 2010; Srivastava et al., 2008). Thus, individual differences in situations could lead to individual differences in average behavior (Zayas et al., 2002). Indeed, studies that demonstrated consistency of aggregated behavior, except one, allowed situations to vary across individuals[1]. (The one exception is the Hartshorne & May (1928) study, although it is debated whether this study demonstrated the consistency of aggregates.) What is needed is a study in which all individuals act in the same objective situations as each other. Individual differences in such

a study could not be explained by situations, because the situations did not differ across people in person-independent ways, so would be best explained as an actor effect.

Alternative #2: People Have the Same Situations as Themselves Over Time

Consistency of individual differences in aggregated behavior averages may reflect only that people stay in the same situations over time. People may act the same in one week as they do in another week because people are in the same objective situations one week as they are in the next. Furthermore, these situations may repeat themselves for largely personality-independent reasons, so themselves might not represent an actor effect. For example, a person might experience an easy commute to work, a repetitively bad boss, or a consistently helpful neighbor. As long as these repeating situations are different for different people, then individual differences would be consistent, even though they would not reflect an actor effect at all (Mischel, 2004; Mischel & Peake, 1982).

What is needed is a study in which each person experiences different situations over time. In such a study, consistent individual differences could not be explained by the situations, because they changed, so would best be explained as an actor effect. These two central criticisms of aggregation-based results are two main reasons for distrusting consistency of aggregated behavior and its implications for the behavioral validity of broad traits (Donnellan, Lucas, & Fleeson, 2009).

How to Measure Behavior in order to Assess Variability?

Self-report based research has revealed surprisingly large variability in behavioral trait enactments (Baird et al., 2006; Fleeson, 2007; Fournier et al., 2008; Minbashian et al., 2010). For example, most people, extraverts and introverts alike, enact wide ranges of extraverted and introverted behavior in the course of a few days. Much research has begun to explore this new domain, including behaviors at work, in mental health and health care, in interpersonal relationships, as a component of adolescence, as a key to adult development, and even among animals (Adams, Roberts, Milich, & Fillmore, 2011; Ahmed, van der Werf, Minnaert, & Kuyper, 2010; Anstey, 2004; Boker, Molenaar, & Nesselroade, 2009; Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Clifton & Kuper, 2011; Côté et al., 2012; Deboeck, Montpetit, Bergeman, & Boker, 2009; Kiene et al., 2008; Molloy et al., 2011; Stamps, Briffa, & Biro, 2012). Behavioral variability suggests that people may be highly flexible and responsive. Therefore, situational reactions may be an important aspect of personality (McAdams & Olson, 2010; Mischel, 2004).

Although provocative, conclusions about behavioral variability are tentative, because of the difficulty of identifying behaviors, quantifying behaviors' degree of trait enactment, assessing enactments across multiple moments, and calculating an index of variability across moments. Traditionally, researchers have used concrete behaviors (e.g., slapped someone on the back) to assess trait enactments (Jackson et al., 2010; Leikas et al., 2012; Wu & Clark,

[1]In the most compelling study to date regarding these issues, Moskowitz (1988) reported Cronbach's alpha estimated consistencies rather than observed correlations across aggregates because the number of situations (six) prevented full aggregation analyses. The promising alpha coefficients called for more definitive evidence with more situations in order to report empirically observed correlations between aggregates.

2003). However, there is ambiguity in identifying the trait content a given concrete behavior is enacting. Concrete behaviors also leave out other aspects of trait enactments such as affect, cognition, and the psychological meaning of the behavior (Leikas et al., 2012; Sherman, Nave, & Funder, 2009).

Building on the state concept (Steyer, Ferring, & Schmitt, 1992), Fleeson (2001) proposed an alternative assessment. Instead of assessing concrete behaviors, researchers could ask individuals to directly rate the degree to which they are enacting each trait content during a particular moment. Such ratings use the same dimensions, content, and scales as the personality traits, but describe a person in a particular moment rather than in general. For example, individuals rate how organized they are at the moment. Resulting trait enactments unambiguously relate to their corresponding trait contents. States are not only behaviors, in the narrow sense, but rather are the same broad adjective descriptions that are used to assess traits. They include the behavioral, affect, cognition, and motivational aspects of trait enactments.

The strength of capturing trait enactments broadly is also a weakness of the adjective solution. Adjectival assessments are more abstract than the concrete, observable aspects of trait enactment. “Organized” is more abstract than is “wrote a list of tasks”. Adjective assessments include the internal cognitive and affective aspects of trait enactment, which may be more pliable and so increase variability (Jackson et al., 2010; Noftle & Fleeson, 2010; Weisbuch, Slepian, Clarke, Ambady, & Veenstra-VanderWeele, 2010). “Friendly” includes feelings of warmth and beliefs about the other person, as well as outward behaviors. It is an open question whether trait enactment would be so variable were it limited to the independently observable aspects of enactment. Observable aspects of enactments might be less pliable, require extra effort, and have social consequences. These features mean that adjectival assessments of enactment might grossly inflate the estimates of within-person variability. Additionally, egoistic biases that may cause people to exaggerate their variability (Krueger, Ham, & Linford, 1996), inattention that may cause people to inaccurately report their enactment (Vazire, 2010); or repeated questioning may create a demand for exaggerated variability (Baird & Lucas, 2011).

Employing observers’ direct observations of trait enactment addresses some of these concerns. Observer reports are still subjective, potentially prone to inattention, and possibly affected by repeated questioning. However, observer reports are based on the observable aspects of trait enactment rather than on the private, internal aspects of trait enactment, and observer reports are independent of the individual conducting the action (Furr, 2009; Vazire, 2010).Instructing observers to use adjectives thus allows capitalizing on the advantages of adjectives, but also allows focus on the observable aspects of behavior rather than on the internal aspects of behavior and also provides independence in assessment. The use of adjectives preserves the psychological meaning of the actions, the direct translation of traits into actions, and the easy calculation of variability. The question is, when assessment retains the use of adjectives, but switches from self-report to observer report, does variability stay extremely high?[2]

Characterizing Individuals by Distributions of Trait Enactments

Rather than leave aggregation and variability as interesting but isolated and statistical abstractions, the density distributions approach combines them into a coherent account of what people are like in their everyday behavior (Fleeson, 2001). Traits can be taken, at least in part, to be descriptions of the trait content people actually enact in everyday life. For example, saying someone is agreeable can be taken in part as a claim that the person acts in agreeable ways in his or her daily life. However, not enough is known about the patterns of actual trait enactment in daily life. It is not known whether people enact a variety of traits in their daily life, how widely they range in the traits they enact, nor how consistent they are in their enactment of traits.

To find out the patterns of trait enactment in everyday life, Fleeson(2001) proposed assessing the frequencies with which each person enacted each level of each state. For example, this means assessing the frequency a person enacted very introverted behavior, the frequency the person enacted moderately introverted behavior, the frequency the person enacted moderately extraverted behavior, and so on.

Because it is believed that individuals pursue shifting goals and respond to varying situations (Epstein, 1983; Fleeson, 2007; McCabe & Fleeson, in press; Mendoza-Denton & Ayduk, 2012; Mischel, 2004; Perunovic, Heller, Ross, & Komar, 2011), individuals may enact different levels of states at different times. Such variability in states will lead to a distribution of states for each individual. The distribution tracks the number of times the individual enacts each given level of the state. Indeed, the high variability in such states underscores the need for characterizing people by entire distributions of states.

Based on the notion that actors have internal determinants of behavior, Fleeson (2001) also proposed that individuals will differ from each other in their distributions, specifically in their distributions' locations (centers), sizes, and shapes. Most notably, the aggregated means discovered by Epstein (1979) represent the center locations of distributions. Indeed, distribution central locations differed across individuals and were extremely consistent from week to week (*r*'s ~ .7 - .9; Fleeson, 2001). The concept of the distribution thus combines the variability in behavior with the stability of aggregated means, and in the process describes what people are like in their everyday life.

Nonetheless, two important questions about the density distributions approach are outstanding. First, because ESM collected data in uncontrolled settings, the source of stable individual differences in distributions may have been stable individual differences in situations rather than actors' internal personalities. Second, because ESM uses self-report, high variability may exist only in internal self-concepts, thoughts, or moods but not in observable aspects of trait enactment such as behaviors, expressions, or spoken words. Thus, this paper puts the density distributions approach to two stringent tests, simultaneously addressing two long unresolved issues in personality psychology.

[2]Conversely, self-reports of variability may underestimate the amount of variability. Even though self-reported adjectives have already shown very high amounts of variability, observer reports, based on more concrete behaviors, may show even higher amounts of variability.

The Present Research: Bringing ESM into the Lab

The goal of the current study was to use observer reports to obtain multiple occasions of trait enactment while maintaining uniform situations across participants. While it is impossible to follow participants throughout their daily lives and control their situations, we thought it may be possible to attain these desirable design characteristics by conducting the ESM study in the lab. Therefore, participants came to the lab across 10 to 20 weeks on twenty different occasions, lasting an hour each. For each session, participants were in a standardized situation. All participants experienced the same situations in the same order. As the situations were the same across participants, the differences between participants would not be due to situations. As the situations differed across occasions (with the exception of three repeats), consistency of average enactments would not be due to stability of situations, or to specific reactions. Additionally, because observers recorded subjects' behavior from behind one-way mirrors, variability and distributions would be due only to enactment in observable behaviors.

Method

Participants

This research is part of the IPSP study, in which participants came into the lab on up to 20 different occasions, once or twice per week for 10 to 20 weeks. "Targets" interacted in constantly changing groups of 2 to 4 (N = 97), whereas "observers" watched the targets from behind one-way mirrors (N = 97 primary observers and N = 86 secondary observers). Of the 321 participants who originally committed to the study, 22 dropped out prior to the study. Seven targets attended six or fewer sessions and 12 observers attended two or fewer sessions; thus, they were not included in any of the following analyses.

Participants were recruited through signs posted around campus, advertisements in the student newspaper and webpage, email listserves, snowballing, and class announcements in class. Interested individuals decided whether to participate at the end of an information session describing the study.

The study was conducted in two waves of enrollment and participation. The first wave occurred across the academic year, one session per week for 20 weeks (N = 62 targets, 118 observers). The second wave occurred in the spring semester of the same academic year, two sessions per week for 10 weeks (N = 35 targets, 65 observers). As the two waves followed the same procedure and contained samples with similar demographic backgrounds, the waves were combined. Wave 1 participants rated themselves as slightly more extraverted (*M's* = 4.53 vs. 4.29,*SD's* = .61, .47,$p < .10$) during the study than did Wave 2 participants; however, the samples did not significantly differ on any other Big Five states for self or observer ratings (all *p*'s > .15). Individuals were compensated $6 (1st semester) or $8 (2nd semester) per session, plus attendance-based bonuses.

Procedure

Participants first attended an information meeting describing the study. They were told that "the purpose of the study is to investigate how people act in a variety of activities and how

they rate their own behavior; we are conducting the study in the laboratory to make sure that participants all have the same situations and so that we can have observers conduct independent ratings of the behavior." At the end of the information meeting, participants provided consent and completed a short questionnaire. Only measures relevant to this paper will be described.

Session procedure—Each participant participated in up to 20 sessions. Groups of two to four targets participated together at a given time for 50 minutes (four was a good size for interaction and for observation). Sessions began with targets reading instructions and then participating in the instructed situation. Typically, targets sat at the sides of a small, four-sided table in a medium-sized room. Four wall-mounted cameras were positioned so each was focused on a target (video was not used in this paper). One-way mirrors lined two of the walls, and four observers were behind each mirror; this allowed observers watching the same target to be behind different mirrors. Observers wore headphones piping sound from the target room, and were instructed to focus on their assigned target at all times and not to talk or make other sounds.

After 20-25 minutes, targets rated their behavior during the previous 20 minutes, and continued the activity for 20-25 more minutes. At the end of the session, targets rated their behavior during the second 20 minutes. Each observer rated his or her assigned target after 20-25 minutes and at the end of each session. Concurrent observation was used to make the observer's task as similar as possible to the target's task.

Laboratory situations were chosen on the basis of five criteria: (i) they should engage individuals in the activity; (ii) they were reasonably representative of the situations participants encounter in daily life; (iii) they showed substantial between-subject variability in previous research; (iv) they were unstructured and unconstrained enough to allow a wide range of behaviors from the individuals in the situation (so as not to inflate or deflate estimates of variability or accuracy); and (v) they provided a variety of settings and tasks so as to provide opportunity for each level of each of the Big 5 states. The following situations occurred in the following order (if two short situations occurred during the same situation, they are labeled A and B): (1) Bingo, (2) discuss medical ethics dilemmas, (3) tell an embarrassing story, (4) plan a safe sex campaign, (5) twister, (6A), sculpt personality traits, (6B) analyze a painting, (7) role play a compensation committee, (8) play a board game, (9) interpret drawings, (10) study, (11) review trait words, and paint, (12A) solve the parking problem on campus, (12B) plan a party, (13) no instructions, (14) plan and present a speech and debate, (15) no instruction 2, (16) play charades, (17) collaborate on one participant's study project, (18) word review and paint 2 (19A) discuss sweatshops, (19B) complete questionnaires (20) play a board game 2. Observers did not rate targets after situation 19B. Because of scheduling issues or errors, on eight occasions participants completed a backup activity.

Scheduling and target-observer triads—Two observers were assigned to a particular target for the duration of the study. Although there may be some advantages to switching observers' targets, in this study each target was always assigned to the same observer. The main rationale for this choice was to avoid inflating within-target variance due to different

observers, which increased the power of the study and ensured that within-subject variation was due to changes in the same person, not due to different raters. (Occasionally, observers observed a different target, due to scheduling difficulties – that data is not reported here.) Halfway through the study, one of the two observers for each target was reassigned to a different task, not relevant to the present paper, based primarily on compatibility of schedules. The primary observer was the one who observed the target the most times, and analyses use the primary observers' data unless otherwise stated.

Observers were mostly assigned to targets at random, taking into account schedule availability. Some initial assignments changed and some groups rearranged due to scheduling or a participant leaving the study. Two of the observers started as targets but were switched to the observer role after 1 or 2 sessions for balancing triads; only their observer data is included.

We rotated targets through the groups, in order to have targets interact with different people each session. The average target interacted with 27-28 different people over of the course of the 20 sessions (because of occasional repeats, the average target had 45 interaction partners). We attempted to have the interaction partners be entirely new each time, but scheduling restrictions meant that the average target had seen 6.9 of the partners one previous time, 2.6 partners two previous times each, 1 partner three previous times, .4 partner four previous times, and .2 partner five previous times.

Attendance was very good, with observers attending 85% of the assigned sessions and targets attending 90% of the assigned sessions, producing 3492 self-reports and 4838 independent observer reports of those same behaviors. Reports were discarded if an experimenter noted a problematic occurrence, such as observers' headphones not working, a participant being late, observers overhearing targets discuss their self-reports, or a participant responding the same number to all adjectives. In the end, 183 target or observer reports were removed, resulting in 3455 target self-reports, and 2973 self-reports that had a corresponding observer report from at least one observer (86%). The extensive nature of the IPSP study makes it one of the most comprehensive studies ever to obtain multiple observed behavior ratings across a large sample of individuals.

Materials

Targets and observers used the same standard Big-Five adjective-based scales (Goldberg, 1992) to rate targets. Four bipolar adjective scales assessed each Big-Five factor: extraversion (unenergetic-energetic, bold-timid, talkative-silent, and unassertive-assertive), agreeableness (warm-cold, rude-polite, stingy-generous, distrustful-trustful), conscientiousness (careless-thorough, hardworking-lazy, serious-frivolous, inefficient-efficient), emotional stability (insecure-secure, nervous-at ease, relaxed-tense, unexcitable-excitable), intellect (unimaginative-imaginative, uninquisitive-curious, uncreative-creative, imperceptive-perceptive). Participants rated how well the adjective described them, or their target, during the previous 20 minutes, on 1-7 scales. Target (t) and observer (o) Cronbach's alphas were calculated across all reports and were as follows: extraversion, .80 (t), .84 (o) agreeableness, .73 (t), .77 (o), conscientiousness, .73 (t), .74 (o), emotional stability, .76 (t), .82 (o), intellect, .84 (t), .84 (o). Targets rated themselves as less conscientious than primary

observers rated them (*M*'s = 4.27, 4.56, *SD's* = .56, .54, *t* (96) = 4.42, *p*< .001), and trended towards rating themselves as more extraverted than primary observers did (*M*'s = 4.44, 4.32, SD's = .57, .70, t(96) = 1.67, *p*<.10), but did not differ in ratings of the other Big 5 traits.

Participants rated the situation during the previous 20 minutes on eight dimensions: others liked the participant, others included the participant, challenge, competitiveness, evaluativeness, control, liked the other participants, and interestingness. Ratings were on seven point scales. The first two ratings were completed during all 20 sessions, and the remaining situation ratings were completed during only the second 10 sessions. Participants also rated other items not used in this report.

Results

Consistency of individual differences in reaction to standardized situations

The first analyses address whether consistency of individual differences in average behaviors was primarily due to actors or situations. In previous studies, consistency of averages may have resulted from individuals encountering similar situations across their daily lives, and individuals differing from each other in the situations they chose. In the current study, participants experienced the same situations; so any individual differences that emerge would not be due to the situations. The situations also differed across time, so any consistency in individual differences would be due to the actor's contribution.

Individual differences in average trait enactments—The average of each trait's enactments across the 20 sessions was calculated separately for each participant. Figure 1 shows histograms for each trait. These results reveal that individuals differed in their average trait enactments, even though they all participated in the same situations. These differences therefore are unlikely to be due to the situations. However, these differences in averages could have been due to factors that differed randomly across participants. Capricious responding or subtle, uncontrolled and unsystematic differences in situations (e.g., daily weather or other participants in the situations) could have led to individual differences in behavior. Such random factors create a possible explanation for differences between people in their averages; the next analyses rule out this explanation.

Stability of distribution parameters across standardized situations—If individual differences in averages were due to random factors, they would not be consistent over time. If they were due to individuals, however, the averages would stay in the same relative positions over time. Importantly, because the situations differed over time in this study, consistency cannot be due to repeated reactions to the same situations.

Two assessments of each participant's averages were obtained from different portions of the data (two averages per trait per participant). For each participant, his or her reports were divided into two sets: the first 10 sessions and the second 10 sessions.

Averages for each set of sessions were calculated for each trait, and correlations across participants of these averages were computed. Correlations from the first half of the sessions to the second half of sessions were all between .7 and .9 (extraversion .79, agreeableness .

82, conscientiousness .71, emotional stability .87, intellect .84). These high correlations indicate high consistency of averages. Individuals had different averages from each other in each set of situations, and those differences were in the same relative positions in one set of situations as they were in the other set. Because situations were the same across participants and different from set to set, these correlations mean that actors were stable because of who they are.

Because three of the sessions had the same instructions as a previous session, we repeated these analyses without those sessions included. Results were highly similar. Note also that averages were slightly higher in the first set of sessions for extraversion (*M*'s = 4.51, 4.38, *SD's* = .63, .59, t (96) = 3.46, $p < .01$), agreeableness (*M*'s = 5.05, 4.98, *SD's* = .63, .59, t (96) = 2.03, $p < .05$), conscientiousness (*M*'s = 4.35, 4.19,*SD's* = .59, .63, t (96) = 3.43, $p < .01$), and intellect (*M*'s = 4.62, 4.52, *SD's* = .68, .73, t (96) = 2.43, $p < .05$).

Discriminations between trait contents—Discriminant validity of the averages was tested by repeating the analyses above, but controlling for the other traits. Discriminant validity is important for showing that the consistency correlations between the traits were due to the content of the traits rather than to general response styles such as acquiescence. The second half average of each trait was predicted from the first half averages of all five traits together. The correlations along the diagonal of Table 1 show that stability remained high, even controlling for levels of the other traits. Thus, stability was not the result of a general tendency to acquiesce or similar response bias (Watson & Tellegen, 2002). The correlations in the off-diagonals were low, indicating that participants discriminated between the traits in their self-reports of enactment. The effects of actors on their means were specific to specific traits.

Consistency across explicitly identified situation types—The argument of the previous findings was that because the situations differed from the first half of sessions to the second half, consistency was evident across different types of situations. However, the differences between situations were implicit in the previous analyses, because the analyses did not explicitly refer to situations. Because we did assess some aspects of situations, we can also test consistency across situations that differed on these explicit aspects.

For each situation rating, we divided occasions by whether they were high or low on the particular dimension with a median split. For example, we divided all occasions into those that were rated high on challenge versus rated low on challenge (with a median split). Within each of these halves, we calculated the participants' average levels of each state (provided that the participant reported at least five occasions in each half). These averages reveal the typical ways the participant acted in pairs of explicitly different situation types.

Table 2 shows the levels of consistency across each type of situation pair. For all traits, the correlations were very high across all opposing situation pairs. For example, the correlation for emotional stability across the situation dimension of challenge was .79. This means that individuals who were more emotionally stable than others in challenging situations were also more emotionally stable than others in non-challenging situations. This pattern of

results means that actors were just as consistent in their aggregates across explicitly differing situations, as they were across implicitly differing situations.

Ipsative stability in enactment profiles—Ipsative consistency is the degree to which trait enactments stay in the same relative position compared to other enactments by the same individual, rather than compared to other individuals (Block & Haan, 1971; Fleeson & Noftle, 2008; Furr, 2008). To test ipsative stability, a profile was constructed for each participant for each half of the data, using the participant's average levels of all five traits for the half of data. A correlation was then computed for each person (n=5 for each correlation), indicating the degree to which the two profiles were similar across the two halves of data. Similar profiles occur for a participant when traits that were higher than others in the first half of data are also higher than others in the second half of the same participant's data.

The average correlation for ipsative stability across participants was .78 (93 of the 97 participants had positive profile correlations). A t-test revealed that the average correlation was significantly greater than 0, t (96) = 30.45, p< .001, indicating that ipsative stability existed.

Some of this high consistency may reflect normative differences between traits, rather than actual consistency (e.g., people are typically more agreeable than extraverted), such that the correlations above do not necessarily reflect consistency. To control for this, the overall average on each of the traits in each half of data was subtracted from each person's profile in the corresponding half. The resulting distinctive profiles indicate the degree to which each person's behavior deviated from the average pattern in the corresponding half of situations (Furr, 2008).

The average distinctive ipsative stability correlation was .67, t (96) = 18.76, p< .001. This means that even the distinctive ways actors acted were consistent from one set of situations to another. Consistency of behavior appears to be influenced by something the actor contributes.

Subjective Perception of Standardized Situations—This paper's goal is to address alternative explanations for the consistency of aggregated averages of behavior. Specifically, these alternative explanations are that individuals experience different situations for personality-independent reasons. Thus, we standardized the personality-independent aspects of situations.

We believe that one of the mechanisms that may produce the actor contribution is that individuals interpret different situations in different ways (Fleeson & Jayawickreme, in press). It is beyond the scope of this paper to systematically investigate the influences of situation perceptions on behavior, but we can examine whether people differed in their interpretations of situations.

Eight unconditional multilevel models were run on the situation ratings, one per situation rating. In each model, one of the situation ratings served as the dependent variable, and there were no independent variables. An intercept was calculated for each situation, and ratings were allowed to vary randomly both across situations and within situations. (Thus, person

was nested within situation in these analyses.) For example, we considered how much ratings of challenge differed across and within situations.

First, the variance of the intercept indicates the degree to which the situations had different ratings from each other on the given dimension. Indeed, the situations were rated significantly differently from each other on challenge, competitiveness, feeling liked, feeling included, evaluativeness, and interestingness, all *p*'s < .05, but not on control or liking of others. This means that the activities were consensually perceived as different from each other on multiple dimensions.

Second, the within-situation variance shows the degree to which individuals differed from each other in their perceptions of the same situation. Indeed, people differed from each other in how they interpreted the same situations on all eight dimensions of challenge, competitiveness, feeling liked, feeling included, evaluativeness, interestingness, control, and liking of others. all *p*'s < .001. Thus, interpretations of situations make a likely candidate for one source of the actor contribution.

Is Variability Present in the Independently Observable Aspects of Trait Enactment?

Different people act differently on different occasions. Some of this difference is due to the actions being conducted by different people (between-person variance), and some of this is due to the same person acting differently on different occasions (within-person variance). The second purpose of this paper is to address whether direct observations of trait enactment reveal a great deal of within-person variability, or whether the previously discovered large amounts of within-person variability are constrained to private, internal aspects of trait enactment. That is, when people reveal themselves publicly and interact with others, are they still highly variable?

The amount of variation in trait enactments answers this question: the total amount of variation is the sum of between-person variation plus within-person variation. Dividing each type of variation by the total determines the percentage of each. In the following analyses, observers' ratings of targets' enactments are used. This allows determination of target actors' variabilities, according to observers' perceptions.

The intraclass correlation, which is the amount of within-person variation in each trait, was calculated using an unconditional model in multilevel modeling, to take into account the nesting of occasions within targets. Each trait was entered as a dependent variable, and the intercept was allowed to vary randomly across participants. As an unconditional model, the constant was the only fixed and the only random effect. Figure 2 shows the amounts of variance between and within individuals for each of the five traits. As can be seen, most of the variation in trait enactment (between 57% and 78%) was within person; that is, due to the same person enacting different trait contents at different times. A smaller portion of trait enactment variability was due to people being different from each other.

F-tests on variances revealed that within-person variance was greater than between-person variance for extraversion, p (28,97) < .001, conscientiousness, p (28,97) < .001, and intellect, p (28,97) < .01, but was equivalent to between-person variance for agreeableness, p

(28, 97) =.08, and emotional stability, *p* (28,97) = .16. The large amount of within-person variability means that the typical individual regularly acts at different levels of the adjectives that are used to define the content of extraversion, agreeableness, conscientiousness, emotional stability, and intellect. This variability is in aspects of individuals' behavior that are visible to others.

The absolute amounts of both within- and between-person variability in trait enactment were comparable to but slightly lower than in previous studies. Within-person variability in the current study ranged from .48 to 1.05 versus .59 to 1.16 in Fleeson (2001). Between-person variability ranged from .29 to .43 in this study versus .44 to .79 in Fleeson (2001). Note that in all studies, within-person variability contains a higher portion of error variance than does between-person variability, which means that their reliable magnitudes are somewhat closer to each other than is represented here.

Large within-person variability means that participants treated these situations as very different from each other, enacting very different levels of traits from situation to situation. Additionally, within each situation, between-person variability was high, ranging from a standard deviation of .71 (conscientiousness while doing definition review and painting) to 1.39 (extraversion while collaborating on one person's project), meaning that people reacted differently from each other to each situation.[3]

Variability in self-ratings—Variability in self-ratings was very similar to variability in observer ratings. First, most of the variance was within-person. Second, the differences between traits were matched. Third, the absolute amounts of variance were similar. F-tests comparing target-report variance to observer-report variance revealed that target-rated within-person variance was not greater than observer-rated within-person variance for any trait, all *F's*< 1.28, all *p's*> .25. This suggests that targets were unlikely to be artificially exaggerating their variability.

Considering between-person variance, target-rated between-person differences were significantly higher than observer-rated between-person differences for extraversion, $p < .05$ and emotional stability, $p < .05$, with a trend for agreeableness, $p = .07$, but not for the other two traits. Thus, targets saw themselves as slightly more different from each other than observers reported.

[3]***Comparison to a "nominal, maximally variable person".*** Another comparison standard for variability would be the amount of variability produced by someone who had no consistency whatsoever but whose behavior was reasonable for the situations of the study. We can make an artificial person who is entirely inconsistent, but yet who acted in normal ways for this study, by randomly selecting one report from each session half (Ozer, personal communication). Because the reports were randomly selected, there is no connection between reports; and, thus, no consistency in behavior contributing to them. Because the reports were selected one per session half, they take into account constraints on trait enactment created by the situations.
The average variabilities across 10 randomly created nominal people were as follows: extraversion, 1.74; agreeableness, .88; conscientiousness, 1.02; emotional stability, 1.36; intellect, 1.40. Comparison to Figure 2 reveals that variability of real people was about 60% of the variability of maximally variable artificial people. On the one hand, this means that real people approach the variability of perfectly variable people, but on the other hand also means that consistency reduced single-enactment variability by about 40%. This is a large number when considering that single enactments were being evaluated and when compared to the 10% predictability of single behaviors sometimes held to be a limit (Ross & Nisbett, 2011).

Density Distributions as the Merging of Variability and Stability: Validity in Stringent Data

Density distributions are proposed to interpret the coexistence of high variability with high stability. Each person has a stable distribution of enactments. Variability exists because single enactments change easily and rapidly. Stability exists because different individuals have different locations, shapes, and sizes of distributions, and the distributions are stable over time. The next analyses test the remaining pieces of this proposal by testing whether the density distributions properties hold up using the most stringent data: observer reports in standardized situations.

For each individual and for each state dimension, the individual's mean, minimum, maximum, SD, skew, and kurtosis of enactments was calculated across the 40 reports. The "Ave. Person" line of Table 3 shows the distribution parameters for the average person (as the mean of the parameters across people), and can be combined to form a mental picture of the average person's distributions of enactments. For example, the average person was somewhat high on intellect on the average occasion, varied from very low to very high on intellect across the 20 sessions, but did not reach into the very lowest or very highest region of the trait. The slight negative skew indicates that the average person's intellect distribution had a tail towards lower intellect.

Table 3 also shows the amount of differences between people in their distributions (as the standard deviation of these parameters across people). To test whether observer reports formed consistent distributions for targets across two independent set of situations, two distributions of trait enactment were computed for each target, one for each half of the situations. Table 4 shows the correlations between the first 10 sessions and the second 10 sessions for all parameters. Stability correlations for the means were between .66 and .81, indicating that participants differed in their average levels of each trait, and that these differences stayed the same from the first set of sessions to the second set of sessions. Individual differences in maximum state, minimum state, and amount of variation in state enactment also remained stable from the first half of sessions to the second half. However, individual differences in skew were only moderately stable; and individual differences in kurtosis were stable only for conscientiousness. Thus, even though everyone had the same situations, people differed greatly in their enactments; even though enactments were based on observers' ratings, individuals both differed in the locations and sizes of their distributions and maintained the locations and sizes of their distributions for each trait across different sets of situations.

It is possible that stability of observer perceptions over time represent a fixedness of first impressions rather than consistency of the target's behavior over time (Leising, Gallrein, & Dufner, 2014). To test for this possibility, we correlated aggregated average behaviors across raters. We calculated the target's, the primary observer's, and the secondary observer's average rating of the target on each trait across all sessions. Because observers were not present at every session, their ratings were based on somewhat different samples of behavior from each other and from the target's ratings. As shown in Table 5, average behaviors were all positive, with some significant and some quite large. Significant correlations mean that actors differed from each other in the observable aspects of behavior and that different raters agreed on those differences regardless of rater bias. Thus, these

correlations ruled out fixedness of first impressions or other rater bias as a complete explanation for observed stability correlations (i.e., at least some of the stability was indeed due to actor differences).

At the same time, the correlations were not as high as those when the same observer was compared over time, and not all were significant, meaning that the correlations did not rule out this alternative as a partial explanation. It is important to keep in mind that targets and observers have different information (internal versus external) and different vantage points (Vazire, 2010), and that rater biases would also mask genuine agreement. Indeed, primary-secondary observer correlations were much stronger, despite being based on half the behavior. Even one observer's ratings on one set of behaviors correlated with a different observer's ratings on an entirely different set of behaviors. Some of the traits showed quite strong agreement, which means that individual differences in averages for at least those traits was largely or mostly a result of genuine individual differences in average behaviors. Lower agreement on the other traits may have been due to the dominance of internal aspects for those traits, leaving observers with less information. Future research should determine what portion of stability is due to such rater stability effects (Leising et al., 2014).

Discussion

This paper addresses two serious but long unanswerable objections to the validity of Big 5 traits. First, are stable individual differences in aggregated means primarily due to situations, rather than being genuine actor contributions? The strong possibility that aggregated mean stability is at least partially, if not solely, due to situations, is grounded on two principles: 1) individuals' situations predict trait enactment (Fleeson, 2007; Fournier et al., 2008; Huang & Ryan, 2011), and 2) individuals experience different situations (Fleeson, 2007; Sherman et al., 2010; Srivastava et al., 2008). To control for situations' influence, and answer the longstanding question, the current study kept situations across participants similar and examined trait enactments in dissimilar situations. Results indicate that aggregated means remained highly stable, even across different situations. Therefore, we conclude that much of stability is due to the actor's contribution, because stability should not be due to differences in situations or to highly specific reactions to repeated situations.

Second, is trait enactment highly variable when observers report on the external, observable aspects? Previous studies have demonstrated high variability using self-reports of trait enactments (e.g., Baird et al., 2006; Charles & Pasupathi, 2003; Fleeson, 2001; Fleeson & Gallagher, 2009; Fournier et al., 2008), which may inflate variability because of perception biases, inclusion of internal cognition and emotion, lack of public consequences to internal states, and problems inherent to repeated assessment. Although individuals may contemplate various courses of action, or feel like taking extreme action, the pressures of real situations and the risks of negative public reaction may lead people to sticking close to their typical behavior in observable trait enactments. Results from this study's observer reports indicate that nonetheless within-person variability was very high. By using observer reports, only observable aspects of trait enactments were measured, which would consist primarily of behavior. Thus, high variability in observer ratings means that even the publically observable and concrete behavioral enactments of traits were highly variable. As social

cognitive approaches to personality have long argued, variability suggests discriminative flexibility in responding to life's situations (Mischel, 2004).

We suggest that personality psychology would be advanced by a model that accounts for both aspects of traits (i.e., stability due to actors' contribution and variability in actual trait enactment), that characterizes what trait levels describe about people's enactment of traits in life, and that proposes causal mechanisms for trait enactments. Whole trait theory attempts to account for these aspects for the Big 5 (Fleeson & Jayawickreme, in press). The density distributions part of the theory has been supported previously (Fleeson, 2001; Fleeson & Gallagher, 2009), and the addition of the current study provides strong support for the theory.

Implications for the Nature of Traits

Research on traits has clarified the structure of individual differences in the descriptiveness of trait terms (e.g., Chang, Connelly, & Geeza, 2012; Saucier, 2009), but has not ascertained the reference of trait terms or the descriptive implications of trait terms. The following questions required more research: What about a person's behavior is described by a trait term? How often are traits enacted in behaviors? How widely and frequently to people deviate from their traits? How consistent are people's behaviors?

The density distributions model provides an account of traits that is grounded in actual enactment of traits, discovers patterns of trait enactment, and reconciles trait and social-cognitive approaches to personality (Fleeson, 2001, 2007; Noftle & Fleeson, 2010). The density distributions model is indebted to the Act Frequency Approach (Buss & Craik, 1983), which made pioneering strides in describing patterns of trait enactment, but nonetheless relied on retrospective questionnaire assessments of behaviors. The density distributions approach employed experience-sampling studies in order to assess behaviors in the moment. However, weaknesses of experience-sampling studies raised important questions about the model; specifically to what extent were results caused by naturally occurring situations that individuals encountered and to what extent were results due to self-report data.

This paper applied novel methods. It found that trait enactments are indeed highly variable, including their observable aspects. It also found that internal determinants are largely responsible for the individual differences in trait enactments. This provides strong evidence in support of the density distributions account of the nature of traits. Additionally, those internal determinants must be of the sort that are capable of producing highly variable trait enactments. Fleeson and Jayawickreme (in press) proposed Whole Trait Theory to suggest that social cognitive mechanisms make good candidates for those internal determinants (Baumert & Schmitt, 2012; Mischel & Shoda, 1995; Read et al., 2010). Social-cognitive mechanisms, such as interpretations of situations and goal pursuits (Mischel, 1973), are both capable of producing consistent individual differences in behavior and also highly variable enactments of traits.

Implications of Stable Means

High consistency of aggregated means has been important because it demonstrates that traits have powerful effects on how people act over time. Traits not only exist, but are highly impactful. The demonstration of stable means, even in standardized objective situations, indicates that something the actor is bringing to the moment is responsible for individual differences in aggregated means. Because situations were very similar across participants in the current study, behavioral differences between participants could not be due to situations. The actor contribution is important because it shows that traits are properties of persons, and that their impact originates in the person. Thus, these data provide strong evidence for the validity of decontextualized, trait terms.

In fact, these data might have underestimated the actor contribution, because part of the actor contribution may very well be that different people select different situations (Ickes, Snyder, & Garcia, 1997), which in turn influence behavior. Since we did not allow people to select their situations, any such part of the actor contribution would not have been included in the present estimates of the actor contribution.

Importantly, actors in the current study responded in a person-specific manner to a new set of situations -- the second half of the data. Consistency across two different sets of situations means the actor contribution is broad, rather than only a collection of stable but highly specific reactions to highly specific features of situations. Traits affected how people acted in a wide variety of situations. These results were reinforced by the consistency across explicitly measured situations. These situation pairs were not just different from each other, but consisted of opposing ends of their particular spectrum (e.g., high levels of being liked by others versus low levels of being liked by others).

Although these results indicate that the actor’s contribution was responsible for stability of density distributions, the results do not specify the nature of that contribution. There are many possibilities for the contributions actors bring by which they end up acting similarly to themselves across a wide range of situations. The most straightforward and traditional rationale for why people act similarly across a range of situations is due to a direct behavioral effect, such that actors have habitual ways of acting (Costa & McCrae, 2006). For example, a person high on extraversion might just habitually be more adventurous. These trait enactments could also be the results of multiple traits acting on the same behavior in a given situation.

A social-cognitive possibility is that actors have habitual ways of interpreting or construing a wide range of situations, and these interpretations lead to repeating similar behaviors across a range of situations (Allport, 1937; Mischel & Shoda, 1995). For example, a person high on extraversion may often see situations as opportunities to have fun, and so be more adventurous. Finally, an interactionist possibility is that actors modify situations in similar ways over time, creating situations that in turn elicit similar behaviors (Buss, 1987; Zayas et al., 2002). For example, a person high on extraversion might encourage submissiveness in others, which in turn elicits extraversion in the extraverted individual.

Implications for Measurement of Trait Enactments

The idea behind state measurement is to measure the degree to which the trait quality is present in the person in the moment on a dimension, rather than count the occurrences of categorically defined concrete behaviors, emotions, or thoughts. This approach directly assesses the content in behavior, as opposed to a two-step procedure in which the behavior is first assessed by a category of behavior and then an inference is made about the relevance of that behavioral category to the trait. In fact, Allport (1966), Eysenck (1985), and Johnson (1997) all argued that the primary use of adjectives, such as friendly, assertive, thorough, etc., is describing behavior.

For example, “warm” is an accurate description of a person in the moment if he or she is doing something pleasant for others, thinking nice thoughts about them, and or is feeling pleasantly toward them. However, “warm” does not specify the specific actions, such as whether he or she smiled, slapped someone on the back, or shared dessert. Rather, it evaluates the meaning of the behavior (Leikas et al., 2012; Sherman et al., 2009).

Although there is compelling face validity to rating behavior by adjectives, there is legitimate concern about whether such abstract adjective ratings are based on anything concrete or otherwise observable to others. Observers’ ratings are based on such observable aspects of enactment. The demonstration of highly stable aggregates, highly variable enactments, and reliable density distributions in these observer ratings suggest that adjective ratings do incorporate observable aspects of enactment. Therefore, it is possible to retain the psychological meaning in behavior, without sacrificing concrete behavioral information, by using independent observers.

Limitations and Future Directions

Although the novel method of conducting an ESM study in the lab allowed observers to report on targets’ trait enactments and allowed situations to be standardized, it necessarily has the complementary limitation of taking place in the lab. The lab situation and the observations may have created some artificiality in behavior, and also made the situations similar to each other in one respect. However, we used a variety of situations that were reasonably representative of the situations participants were likely to encounter in daily life, and that were unstructured and unconstrained enough to allow for a wide range of behaviors from the individuals in the situation, such as playing a game, studying, and discussing political issues. Empirically, the data showed that each person acted very differently across the situations, verifying that situations were not too similar.

Although this study showed that situations were not responsible for stable aggregations, the main analyses did not directly measure the effect of situations. Rather, situation effects were addressed indirectly, by standardizing them across participants. Standardization did mean that situations were not responsible for individual differences because situations did not differ between participants. To mitigate the indirect attention to situations in the main analyses, additional analyses calculated aggregates in explicitly identified and opposing situation pairs. Consistency of aggregates was also high across such explicit situation pairs.

A third limitation is that the results did not rule out that some of the consistency across situations was due to rater biases rather than to actors' behavior. Although mean ratings were correlated across different raters, and sometimes quite strongly, the correlations were not all significant and not as high as the same-rater correlations. The significant correlations even with different raters and different situations mean that individual differences in average trait enactments were due to actor effects for at least some traits. However, for the remaining traits it is still unclear how much the lower correlations were due to different information (Vazire, 2010) or to rater effects (Leising et al., 2014). Thus, these results ruled out two long-standing objections to the conclusion that individual differences in aggregated averages are due to actors' personalities, and largely ruled out a third objection for some traits, but did not rule out the third objection for some of the traits.

A final limitation is that scheduling was not entirely random. Random differences in situations cannot explain consistency over time, only systematic differences can. Participants may have differed systematically in the time of day or the types of people they interacted with in the sessions. Most of the scheduling was random, and great efforts were taken to make sure that participants rarely interacted in the same groups and participated at different times of day. However, some small differences between people in situations may have existed, and these small differences may have been responsible for some consistency. Furthermore, targets interacted with different sets of interaction partners, although this could not explain the consistency over time of aggregates. It does seem unlikely that the few minor situational effects that were not perfectly random had such large effects on behavior that they single-handedly accounted for the very high consistency correlations.

Conclusion

Although situations are a cause of trait enactment in behavior (Fleeson, 2007), and although individuals differ in typically encountered situations (Sherman et al., 2010; Srivastava et al., 2008), consistency in behavior over time, as revealed by the classic aggregation tradition (Epstein, 1979), is not due to situations, but rather appears to be due to individuals' personalities. Although self-reports of adjectival accounts of behavior suffer somewhat from subjectivity, bias, and lack of knowledge, large variability in behavior is a real, observable phenomenon based on observable aspects of trait enactment rather than existing only in the mind of the actor.

The evidence for these conclusions came from a large investigation of observed behavior. Individual differences in average levels of trait enactment were highly consistent (split half r's ~.80), across multiple laboratory situations that were similar across participants. Nonetheless, there was a tremendous amount of variability within each person's trait enactment, such that people differed from themselves more than they differed from others, and this finding was true even in observer-rated enactments. The density distributions model remained an accurate model of trait enactment even when challenged by standardized situations and observer ratings of trait enactment.

Acknowledgments

We would like to thank Sarah Ross and Elizabeth Gedmark for coordinating this research. William Fleeson, Department of Psychology, Wake Forest University, Winston-Salem, North Carolina, 27109. Research presented in this manuscript was supported by the National Institute of Mental Health of the National Institutes of Health under award number R01MH70571. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This project/publication was also made possible through the support of grants from the Templeton World Charity Foundation and the John Templeton Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Templeton World Charity Foundation or the John Templeton Foundation.

References

Adams ZW, Roberts WM, Milich R, Fillmore MT. Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? Psychological Assessment. 2011; 23(2): 427–436. http://doi.org/10.1037/a0022112. [PubMed: 21443365]

Ahmed W, van der Werf G, Minnaert A, Kuyper H. Students' daily emotions in the classroom: Intra-individual variability and appraisal correlates. British Journal of Educational Psychology. 2010; 80(4):583–597. http://doi.org/10.1348/000709910X498544. [PubMed: 20438661]

Allport, GW. Personality: a psychological interpretation. Oxford England: Holt; 1937.

Allport GW. Traits revisited. American Psychologist. 1966; 21(1):1–10. http://doi.org/10.1037/h0023295.

Anstey KJ. Within-Person Variability as a Dynamic Measure of Late-Life Development: New Methodologies and Future Directions. Gerontology. 2004; 50(4):255–258. http://doi.org/10.1159/000078355. [PubMed: 15258432]

Baird BM, Le K, Lucas RE. On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. Journal of Personality and Social Psychology. 2006; 90(3):512–527. http://doi.org/10.1037/0022-3514.90.3.512. [PubMed: 16594835]

Baird BM, Lucas RE. "… And how about now?": Effects of item redundancy on contextualized self-reports of personality. Journal of Personality. 2011; 79(5):1081–1112. http://doi.org/10.1111/j.1467-6494.2011.00716.x. [PubMed: 21241305]

Baumeister RF, Vohs KD, Funder DC. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? Perspectives on Psychological Science. 2007; 2(4):396–403. http://doi.org/10.1111/j.1745-6916.2007.00051.x. [PubMed: 26151975]

Baumert A, Schmitt M. Personality and information processing. European Journal of Personality. 2012; 26(2):87–89. http://doi.org/10.1002/per.1850.

Block, J.; Haan, N. Lives through time. Berkeley: Bancroft Books; 1971.

Boker SM, Molenaar PCM, Nesselroade JR. Issues in intraindividual variability: Individual differences in equilibria and dynamics over multiple time scales. Psychology and Aging. 2009; 24(4):858–862. http://doi.org/10.1037/a0017912. [PubMed: 20025401]

Borkenau P, Mauer N, Riemann R, Spinath FM, Angleitner A. Thin Slices of Behavior as Cues of Personality and Intelligence. Journal of Personality and Social Psychology. 2004; 86(4):599–614. http://doi.org/10.1037/0022-3514.86.4.599. [PubMed: 15053708]

Buss DM. Selection, evocation, and manipulation. Journal of Personality and Social Psychology. 1987; 53(6):1214–1221. http://doi.org/10.1037/0022-3514.53.6.1214. [PubMed: 3320336]

Buss DM, Craik KH. The act frequency approach to personality. Psychological Review. 1983; 90(2): 105–126. http://doi.org/10.1037/0033-295X.90.2.105.

Chang L, Connelly BS, Geeza AA. Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait–multimethod approach. Journal of Personality and Social Psychology. 2012; 102(2):408–426. http://doi.org/10.1037/a0025559. [PubMed: 21967007]

Charles ST, Pasupathi M. Age-related patterns of variability in self-descriptions: Implications for everyday affective experience. Psychology and Aging. 2003; 18(3):524–536. http://doi.org/10.1037/0882-7974.18.3.524. [PubMed: 14518813]

Clifton A, Kuper LE. Self-reported personality variability across the social network is associated with interpersonal dysfunction. Journal of Personality. 2011; 79(2):359–390. http://doi.org/10.1111/j.1467-6494.2010.00686.x. [PubMed: 21395592]

Costa, PTJ.; McCrae, RR. Trait and Factor Theories. In: Thomas, JC.; Segal, DL.; Hersen, M., editors. Comprehensive Handbook of Personality and Psychopathology, Vol 1: Personality and Everyday Functioning. Hoboken, NJ US: John Wiley & Sons Inc; 2006. p. 96-114.

Côté S, Moskowitz DS, Zuroff DC. Social relationships and intraindividual variability in interpersonal behavior: Correlates of interpersonal spin. Journal of Personality and Social Psychology. 2012; 102(3):646–659. http://doi.org/10.1037/a0025313. [PubMed: 21895381]

Deboeck PR, Montpetit MA, Bergeman CS, Boker SM. Using derivative estimates to describe intraindividual variability at multiple time scales. Psychological Methods. 2009; 14(4):367–386. http://doi.org/10.1037/a0016622. [PubMed: 19968398]

Distel MA, Roeling MP, Tielbeek JJ, van Toor D, Derom CA, Trull TJ, Boomsma DI. The covariation of trait anger and borderline personality: A bivariate twin-siblings study. Journal of Abnormal Psychology. 2012; 121(2):458–466. http://doi.org/10.1037/a0026393. [PubMed: 22149910]

Donnellan MB, Lucas RE, Fleeson W. Introduction to personality and assessment at age 40: Reflections on the legacy of the person-situation debate and the future of person-situation integration. Journal of Research in Personality. 2009; 43(2):117–119. http://doi.org/10.1016/j.jrp.2009.02.010.

Epstein S. The stability of behavior: I. On predicting most of the people much of the time. Journal of Personality and Social Psychology. 1979; 37(7):1097–1126. http://doi.org/10.1037/0022-3514.37.7.1097.

Epstein S. A research paradigm for the study of personality and emotions. Nebraska Symposium on Motivation Nebraska Symposium on Motivation. 1983:91–154. [PubMed: 6843719]

Erickson TM, Newman MG, Pincus AL. Predicting unpredictability: Do measures of interpersonal rigidity/flexibility and distress predict intraindividual variability in social perceptions and behavior? Journal of Personality and Social Psychology. 2009; 97(5):893–912. http://doi.org/10.1037/a0016515. [PubMed: 19857009]

Eysenck, HJ.; Eysenck, MW. Personality and individual differences: A natural science approach. Plenum Press; New York: 1985. Retrieved from http://www.getcited.org/pub/102380515

Fleeson W. Toward a structure- and process-integrated view of personality: Traits as density distributions of states. Journal of Personality and Social Psychology. 2001; 80(6):1011–1027. http://doi.org/10.1037/0022-3514.80.6.1011. [PubMed: 11414368]

Fleeson W. Situation-Based Contingencies Underlying Trait-Content Manifestation in Behavior. Journal of Personality. 2007; 75(4):825–862. http://doi.org/10.1111/j.1467-6494.2007.00458.x. [PubMed: 17576360]

Fleeson W. Perspectives on the Person. The Oxford Handbook of Personality and Social Psychology. 2012:33.

Fleeson W, Gallagher P. The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. Journal of Personality and Social Psychology. 2009; 97(6):1097–1114. http://doi.org/10.1037/a0016786. [PubMed: 19968421]

Fleeson, W.; Jayawickreme, E. Whole Trait Theory. Journal of Research in Personality. n.d. http://doi.org/10.1016/j.jrp.2014.10.009

Fleeson W, Noftle EE. Where does personality have its influence? A supermatrix of consistency concepts. Journal of Personality. 2008; 76(6):1355–1386. [PubMed: 19012652]

Fleeson W, Noftle EE. In favor of the synthetic resolution to the person-situation debate. Journal of Research in Personality. 2009; 43(2):150–154. http://doi.org/10.1016/j.jrp.2009.02.008.

Fleisher MS, Woehr DJ, Edwards BD, Cullen KL. Assessing within-person personality variability via frequency estimation: More evidence for a new measurement approach. Journal of Research in Personality. 2011; 45(6):535–548. http://doi.org/10.1016/j.jrp.2011.06.009.

Fournier MA, Moskowitz DS, Zuroff DC. Integrating dispositions, signatures, and the interpersonal domain. Journal of Personality and Social Psychology. 2008; 94(3):531–545. http://doi.org/10.1037/0022-3514.94.3.531. [PubMed: 18284296]

Funder DC. Personality. Annual Review of Psychology. 2001; 52:197–221. http://doi.org/10.1146/annurev.psych.52.1.197.

Furr RM. Framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. Journal of Personality. 2008; 76(5):1267–1316. http://doi.org/10.1111/j.1467-6494.2008.00521.x. [PubMed: 18705644]

Furr RM. The study of behaviour in personality psychology: Meaning, importance and measurement. European Journal of Personality. 2009; 23(5):437–453. http://doi.org/10.1002/per.726.

Furr RM, Funder DC. Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches. Journal of Research in Personality. 2004; 38(5):421–447. http://doi.org/10.1016/j.jrp.2003.10.001.

Hartshorne, H.; May, MA. Studies in the nature of character, [part] 1: Studies in deceit: book 1, General methods and results; book 2, Statistical methods and results. New York, NY US: MacMillan Co; 1928.

Huang JL, Ryan A. Beyond personality traits: A study of personality states and situational contingencies in customer service jobs. Personnel Psychology. 2011; 64(2):451–488. http://doi.org/10.1111/j.1744-6570.2011.01216.x.

Ickes, W.; Snyder, M.; Garcia, S. Personality influences on the choice of situations. In: Hogan, R.; Johnson, JA.; Briggs, SR., editors. Handbook of personality psychology. San Diego, CA, US: Academic Press; 1997. p. 165-195.

Jackson JJ, Wood D, Bogg T, Walton KE, Harms PD, Roberts BW. What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). Journal of Research in Personality. 2010; 44(4):501–511. http://doi.org/10.1016/j.jrp.2010.06.005. [PubMed: 21278818]

Johnson, JA. Units of analysis for the description and explanation of personality. In: Hogan, R.; Johnson, JA.; Briggs, SR., editors. Handbook of personality psychology. San Diego, CA US: Academic Press; 1997. p. 73-93.

Kiene SM, Tennen H, Armeli S. Today I'll use a condom, but who knows about tomorrow: A daily process study of variability in predictors of condom use. Health Psychology. 2008; 27(4):463–472. http://doi.org/10.1037/0278-6133.27.4.463. [PubMed: 18643004]

Krueger J, Ham JJ, Linford KM. Perceptions of behavioral consistency: Are people aware of the actor-observer effect? Psychological Science. 1996; 7(5):259–264. http://doi.org/10.1111/j.1467-9280.1996.tb00371.x.

Larsen RJ, Augustine AA, Prizmic Z. A process approach to emotion and personality: Using time as a facet of data. Cognition & Emotion. 2009; 23(7):1407–1426. http://doi.org/10.1080/02699930902851302.

Leikas S, Lönnqvist JE, Verkasalo M. Persons, situations, and behaviors: Consistency and variability of different behaviors in four interpersonal situations. Journal of Personality and Social Psychology. 2012; 103(6):1007–1022. http://doi.org/10.1037/a0030385. [PubMed: 23066881]

Leising D, Gallrein AMB, Dufner M. Judging the behavior of people we know: Objective assessment, confirmation of preexisting views, or both? Personality and Social Psychology Bulletin. 2014; 40(2):153–163. http://doi.org/10.1177/0146167213507287. [PubMed: 24107709]

McAdams DP, Olson BD. Personality development: Continuity and change over the life course. Annual Review of Psychology. 2010; 61:517–542. http://doi.org/10.1146/annurev.psych.093008.100507.

McCabe K, Fleeson W. What is extraversion for? Integrating trait and motivational perspectives and identifying the purpose of extraversion. Psychological Science. in press.

Mendoza-Denton R, Ayduk Ö. Personality and Social Interaction. The Oxford Handbook of Personality and Social Psychology. 2012:446.

Minbashian A, Wood RE, Beckmann N. Task-contingent conscientiousness as a unit of personality at work. Journal of Applied Psychology. 2010; 95(5):793–806. http://doi.org/10.1037/a0020016. [PubMed: 20718521]

Mischel, W. Personality and assessment. Hoboken, NJ US: John Wiley & Sons Inc; 1968.

Mischel W. Toward a cognitive social learning reconceptualization of personality. Psychological Review. 1973; 80(4):252–283. http://doi.org/10.1037/h0035002. [PubMed: 4721473]

Mischel W. Toward an integrative science of the person. Annual Review of Psychology. 2004; 55:1–22. http://doi.org/10.1146/annurev.psych.55.042902.130709.

Mischel W, Peake PK. Beyond déjà vu in the search for cross-situational consistency. Psychological Review. 1982; 89(6):730–755. http://doi.org/10.1037/0033-295X.89.6.730.

Mischel W, Shoda Y. A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. Psychological Review. 1995; 102(2):246–268. http://doi.org/10.1037/0033-295X.102.2.246. [PubMed: 7740090]

Molloy LE, Ram N, Gest SD. The storm and stress (or calm) of early adolescent self-concepts: Within- and between-subjects variability. Developmental Psychology. 2011; 47(6):1589–1607. http://doi.org/10.1037/a0025413. [PubMed: 21928883]

Moskowitz DS. Cross-situational generality in the laboratory: Dominance and friendliness. Journal of Personality and Social Psychology. 1988; 54(5):829–839. http://doi.org/10.1037/0022-3514.54.5.829.

Noftle EE, Fleeson W. Age differences in big five behavior averages and variabilities across the adult life span: Moving beyond retrospective, global summary accounts of personality. Psychology and Aging. 2010; 25(1):95–107. http://doi.org/10.1037/a0018199. [PubMed: 20230131]

Perunovic WQE, Heller D, Ross M, Komar S. The within-person dynamics of intrinsic and extrinsic motivation, affective states, and cultural identification: A diary study of bicultural individuals. Social Psychological and Personality Science. 2011; 2(6):635–641. http://doi.org/10.1177/1948550611405071.

Read SJ, Monroe BM, Brownstein AL, Yang Y, Chopra G, Miller LC. A neural network model of the structure and dynamics of human personality. Psychological Review. 2010; 117(1):61–92. http://doi.org/10.1037/a0018131. [PubMed: 20063964]

Ross, L.; Nisbett, RE. The Person and the Situation: Perspectives of Social Psychology. 2. Pinter & Martin Ltd; 2011.

Rushton JP, Brainerd CJ, Pressley M. Behavioral development and construct validity: The principle of aggregation. Psychological Bulletin. 1983; 94(1):18–38. http://doi.org/10.1037/0033-2909.94.1.18.

Saucier G. Recurrent personality dimensions in inclusive lexical studies: Indications for a big six structure. Journal of Personality. 2009; 77(5):1577–1614. http://doi.org/10.1111/j.1467-6494.2009.00593.x. [PubMed: 19678873]

Sherman RA, Nave CS, Funder DC. The apparent objectivity of behaviour is illusory. European Journal of Personality. 2009; 23(5):430–433.

Sherman RA, Nave CS, Funder DC. Situational similarity and personality predict behavioral consistency. Journal of Personality and Social Psychology. 2010; 99(2):330–343. http://doi.org/10.1037/a0019796. [PubMed: 20658847]

Smith RE, Shoda Y, Cumming SP, Smoll FL. Behavioral signatures at the ballpark: Intraindividual consistency of adults' situation-behavior patterns and their interpersonal consequences. Journal of Research in Personality. 2009; 43(2):187–195. http://doi.org/10.1016/j.jrp.2008.12.006.

Snyder M. The influence of individuals on situations: Implications for understanding the links between personality and social behavior. Journal of Personality. 1983; 51(3):497–516. http://doi.org/10.1111/j.1467-6494.1983.tb00342.x.

Snyder, M. Public appearances, private realities: The psychology of self-monitoring. New York, NY US: W H Freeman/Times Books/ Henry Holt & Co; 1987.

Srivastava S, Angelo KM, Vallereux SR. Extraversion and positive affect: A day reconstruction study of person-environment transactions. Journal of Research in Personality. 2008; 42(6):1613–1618. http://doi.org/10.1016/j.jrp.2008.05.002.

Stamps JA, Briffa M, Biro PA. Unpredictable animals: Individual differences in intraindividual variability (IIV). Animal Behaviour. 2012; 83(6):1325–1334. http://doi.org/10.1016/j.anbehav.2012.02.017.

Steyer R, Ferring D, Schmitt MJ. States and traits in psychological assessment. European Journal of Psychological Assessment. 1992; 8(2):79–98.

Vansteelandt K, Van Mechelen I. The personality triad in balance: Multidimensional individual differences in situation-behavior profiles. Journal of Research in Personality. 2004; 38(4):367–393. http://doi.org/10.1016/j.jrp.2003.08.001.

Vazire S. Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. Journal of Personality and Social Psychology. 2010; 98(2):281–300. http://doi.org/10.1037/a0017908. [PubMed: 20085401]

Watson D, Tellegen A. Aggregation, acquiescence, and the assessment of trait affectivity. Journal of Research in Personality. 2002; 36(6):589–597. http://doi.org/10.1016/S0092-6566(02)00509-3.

Weisbuch M, Slepian ML, Clarke A, Ambady N, Veenstra-VanderWeele J. Behavioral stability across time and situations: Nonverbal versus verbal consistency. Journal of Nonverbal Behavior. 2010; 34(1):43–56. http://doi.org/10.1007/s10919-009-0079-9. [PubMed: 20161668]

Wu KD, Clark LA. Relations between personality traits and self-reports of daily behavior. Journal of Research in Personality. 2003; 37(4):231–256. http://doi.org/10.1016/S0092-6566(02)00539-1.

Zayas V, Shoda Y, Ayduk ON. Personality in context: An interpersonal systems perspective. Journal of Personality. 2002; 70(6):851–900. http://doi.org/10.1111/1467-6494.05026. [PubMed: 12498358]

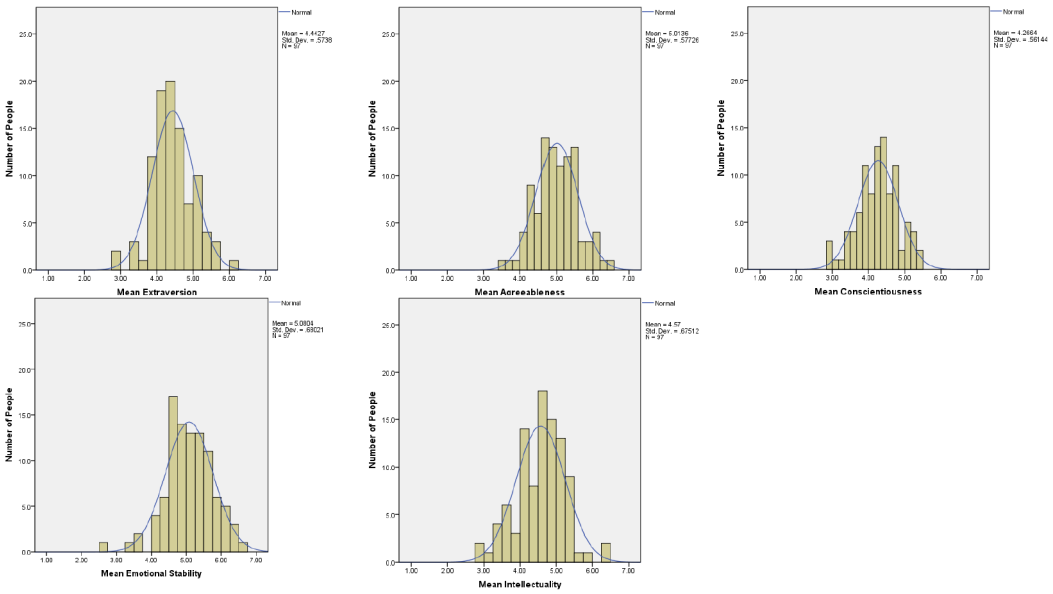

Figure 1.

Distributions of individual mean enactments of each of the five traits. Individuals had different means of each trait, despite the situations being standardized across situations. These differences could have been due to random factors or to something the actor contributed to the situation. Means were 4.4, 5.0 4.3, 5.1, and 4.6, and SD's were .57, .58, . 56, .68, and .68 for extraversion, agreeableness, conscientiousness, emotional stability, and intellect, respectively. N =97.

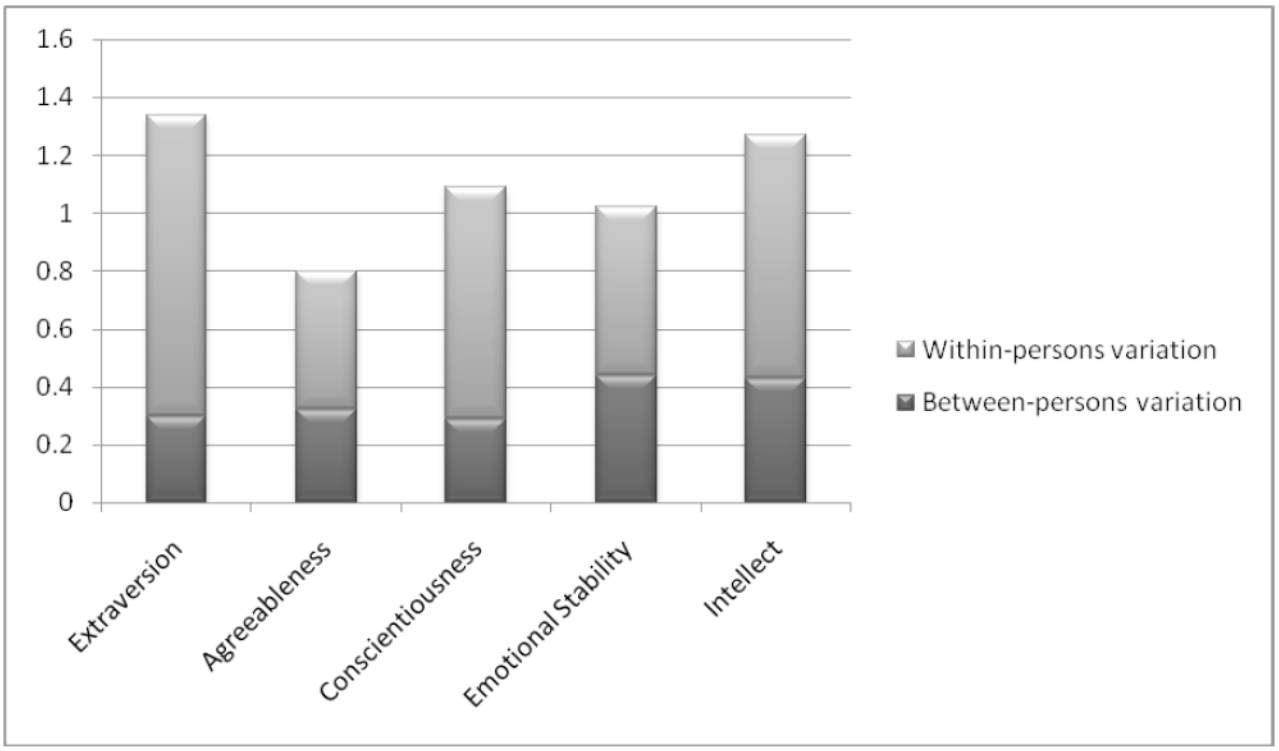

Figure 2.

Amount of within-person and between-person variance in the externally observable aspects of trait enactment. Most of the variance was within-person (extraversion, 78%; agreeableness, 60%; conscientiousness, 73%; emotional stability, 57%; intellect, 66%).

Table 1

Discriminant stability of density distribution locations (means).

Mean in 1st half of sessions	Mean in 2nd half of sessions				
	Extra.	**Agree.**	**Cons.**	**Emot. Sta.**	**Intell.**
Extraversion	.66***	-.11	-.02	-.09	-.04
Agreeableness	-.13	.69***	-.22*	.07	-.17
Conscientiousness	.11	.02	.64***	.06	.08
Emotional Stability	.14	.05	.02	.82***	.06
Intellect	-.02	.17*	.32**	-.02	.92***

N = 97. Table entries are unstandardized beta coefficients predicting 2nd half means from all 5 1st half means simultaneously. Higher predictions on the diagonal mean that participants discriminated among traits and that stability was specific to each trait.

Table 2

Consistency of aggregates across explicitly identified situations.

Situation Pair	Extraver.	Agreeab.	Conscie.	Emo. Sta.	Intellect
Low others liking you/High others liking you (n=63)	.50	.80	.67	.71	.72
Low inclusion/High inclusion (n=59)	.56	.78	.66	.88	.73
Low challenge/High challenge (n=61)	.62	.78	.69	.79	.70
No competitiveness/Some competitiveness (n=51)	.67	.84	.66	.86	.79
Not evaluated/Evaluated (n=48)	.49	.78	.68	.69	.73
Little control/Much control (n=50)	.50	.85	.67	.84	.70
Low liking of others/High liking of others (n=25)	.54	.82	.52	.67	.87
Uninteresting/Interesting (n=62)	.46	.84	.72	.78	.69

Numbers of participants varied because only participants with at least 5 reports in each situation level were included. For each situation, reports were divided into those below (or equal to) the median or those above the median. Aggregates were calculated in each portion of data, and correlations are between those aggregates. All correlations were statistically significant at the $p < .001$ level.

Table 3

Density distributions in observer ratings and standardized situations.

	Distribution Parameter					
Trait/Statistic	**Mean**	**Min.**	**Max.**	**Std. Dev.**	**Skew**	**Kurtosis**
Extraversion						
Ave. Person	4.32	2.03	6.04	1.10	-.47	-.06
Diffs btw. People	.70	.82	.69	.31	.60	1.18
Agreeableness						
Ave. Person	4.94	3.50	6.05	.64	-.29	.70
Diffs btw. People	.67	.94	.62	.20	.89	2.70
Conscientiousness						
Ave. Person	4.56	2.88	6.16	.82	-.01	.40
Diffs btw. People	.54	.92	.65	.30	.72	1.41
Emotional Stability						
Ave. Person	4.92	3.18	6.23	.78	-.42	.48
Diffs btw. People	.81	1.05	.65	.26	.70	1.96
Intellect						
Ave. Person	4.56	2.52	6.07	.90	-.48	.53
Diffs btw. People	.69	.95	.67	.29	.68	1.94

Table 4

Stability of density distribution parameters in observer ratings.

	Distribution Parameter					
Trait	**Mean**	**Min.**	**Max.**	**Std. Dev.**	**Skew**	**Kurtosis**
Extraversion	.67	.49	.55	.42	.14	.06
Agreeableness	.78	.56	.61	.37	.30	.09
Conscientiousness	.66	.49	.54	.49	.32	.23
Emotional Stability	.81	.63	.67	.53	.19	.19
Intellect	.70	.34	.59	.46	-.07	.11

N = 94 or 95. Correlations are between means in first half of data and second half of data, as reported by the observer. All correlations were significant at the $p < .001$ level for mean, minimum, maximum, and standard deviation. Skew stability correlations were significant for agreeableness and conscientiousness, $p < .01$, and kurtosis stability only for conscientiousness, $p < .05$.

Table 5

Correlations of Average Aggregated Behaviors Across Raters

Trait	Target-Prim. Obs.	Prim. Obs.-Sec. Obs.	Prim. Obs. Half 2 -Sec. Obs. Half 1
Extraversion	.32**	.46***	.35**
Agreeableness	.03	.11	.06
Conscientiousness	.31**	.19†	.10
Emotional Stability	.10	.46***	.41**
Intellect	.15	.22*	.15

N = 97 for target-primary observer correlations and 86 for primary observer-secondary observer correlations. Secondary observer ratings in first half of sessions correlated .77, .04, .98*, .50, and .82 with primary observer ratings in the second half of sessions, although these were based on only four targets' behavior.

*** $p < .001$

** $p < .01$

* $p < .05$

† $p < .10$