



Published in final edited form as:

J Pers Soc Psychol. 2009 December ; 97(6): 1097–1114. doi:10.1037/a0016786.

The Implications of Big-Five Standing for the Distribution of Trait Manifestation in Behavior: Fifteen Experience-Sampling Studies and a Meta-Analysis

William Fleeson and

Department of Psychology, Wake Forest University, Winston-Salem, North Carolina, 27109

M. Patrick Gallagher

Department of Psychology and Neuroscience, Duke University, Durham, North Carolina, 27708

William Fleeson: FleesonW@wfu.edu; M. Patrick Gallagher: mpg9@duke.edu

Abstract

One of the fundamental questions in personality psychology is whether and how strongly trait standing relates to the traits that people actually manifest in their behavior, when faced with real pressures and real consequences of their actions. One reason this question is fundamental is the common belief that traits do not predict how individuals behave, which leads to the reasonable conclusion that traits are not important to study. However, this conclusion is surprising given that there is almost no data on the ability of traits to predict distributions of naturally occurring, representative behaviors of individuals (and that there are many studies showing that traits do indeed predict specific behaviors). This paper describes a meta-analysis of 15 experience-sampling studies, conducted over the course of eight years, amassing over 20,000 reports of trait manifestation in behavior. Participants reported traits on typical self-report questionnaires, then described their current behavior multiple times per day for several days, as the behavior was occurring. Results showed that traits, contrary to expectations, were strongly predictive of individual differences in trait manifestation in behavior, predicting average levels with correlations between .42 and .56 (approaching .60 for stringently restricted studies). Several other ways of summarizing trait manifestation in behavior were also predicted from traits. These studies provide evidence that traits are powerful predictors of actual manifestation of traits in behavior.

The purpose of this paper is to identify the implications that standing on each Big Five trait, as assessed by questionnaire, has for the distributions of trait manifestations in behavior. One of the basic questions in personality psychology is whether what people report they do in general and in reflection predicts what they do when confronted with real situations and when their actions have real consequences. That is, whether descriptions of people at the trait level are accurate as descriptions of the traits people actually manifest in their behavior, as they go about their lives and encounter situations and other influences on behavior. In fact, whether personality is required as an explanatory concept may depend on whether, and to what extent, personality is a determinant of social behavior. The received wisdom is that personality predictions vary from .00 to a maximum of .30 (or maybe .40) (Ross & Nisbett, 1991). As noted by Kenrick & Funder (1988), Swann & Seyle (2005), and others, many have taken this

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/journals/psp

level of prediction to mean that personality is not a strong predictor of behavior. However, there is in fact limited empirical knowledge about the actual size of the predictions, so it is not known whether, to what extent, or in what ways traits predict individuals' representative behaviors. For example, it is not known what relationship, if any, a description of an individual as moderately extraverted has to whether and how often the individual acts in an extraverted, moderately extraverted, or introverted manner. This paper describes 15 studies that investigate how strongly traits predict the distributions of actual trait manifestation in naturally-occurring, representative behaviors.

Personality psychology has yet to produce much knowledge about whether trait questionnaire scores predict the distribution of naturally occurring, representative behaviors, how strongly, and how deviations are accounted for in that prediction (Funder, 2001; Mehl, Gosling, & Pennebaker, 2006; Pytlik Zillig, Hemenover, & Dienstbier, 2002; Saucier & Goldberg, 1996; Wu & Clark, 2003) despite evidence that deviations are plentiful and perhaps plentiful enough to undermine the prediction entirely (Fleeson, 2001; Mischel, 1968). Part of the reason for this knowledge deficit may be two related and fundamental difficulties in obtaining a distribution of a person's trait manifestations in behavior: (1) it is very difficult to catalog the universe of behaviors that count as manifestations of each given trait, making it difficult to determine when a person is acting in the way described by the trait (Funder, 2001); and (2) it is difficult to record all of the behaviors of the person for an extended-enough period of time to obtain a comprehensive distribution of how the person acts. Thus, a second purpose of this paper is to describe the density distributions approach as a solution to these problems. The density distributions approach solves these problems by assessing behavior not in terms of specific, concrete actions but rather directly in terms of how much the behavior at the moment manifests the content of the trait. It assesses behavior of each individual several times a day for several days to form distributions for each individual for each trait, describing the frequency with which each individual acts at each level of the trait. These distributions can be related to trait standings from questionnaires in order to determine the implications of such trait standings for the distribution of behavior.

The relationship between trait scores and distributions of corresponding behavior can be broken down into at least five basic questions (summarized in Table 1).

1. Are there any implications of trait standing for trait manifestation in behavior at all? Although there are many studies showing that traits do predict specific behaviors, there is not evidence about whether traits predict a comprehensive account of trait manifestation in naturally-occurring behavior (Funder, 2001; Pytlig Zillig et al., 2002; Saucier & Goldberg, 1996; Wu & Clark, 2003).
2. How strong are the implications for distributions of trait manifestation in behavior? It is the magnitude of the prediction that is important to the relative usefulness of traits as explanatory concepts. The stronger the relationship, the more powerful and useful traits would be for understanding human psychology.
3. How often and how far do individuals deviate from their trait standing in their behavior? Individuals can and do adjust their behavior if circumstances require (individuals are less exuberant at funerals than at weddings), but it is not known whether deviations require such strong situations as weddings and funerals. or whether deviations are routine. The more deviations, the less predictive traits may be.
4. What summary function across multiple behaviors corresponds to trait standing? Because the same individual acts differently on different occasions to at least some extent, characterizing the entire distribution requires summarizing in some way across those behaviors. Possible summaries include at least the central tendency, the mode, or the most extreme. Determining which summary is implied by trait standing will

clarify what information about a person is conveyed by trait standing and may inform investigations into the mechanisms leading from trait standing to trait manifestation in behavior.

5. Are there further implications of trait standing for patterns of trait-manifestation in behavior? Beyond implicating a particular summary, trait standing may also implicate particular deviations from the summary. For example, individuals with a moderate standing on a trait may also be more variable in manifesting that trait.

Existing Knowledge About the Behavioral Manifestation of Traits

Knowledge about traits predominately concerns the correlation structure of individual differences (Funder, 2001; Saucier & Goldberg, 1996). This knowledge is impressive in its scope and its relevance and is powerful in its validity evidence (Funder, 2001; Wiggins, 1997). It is clear now that scores on trait inventories are highly reliable, even across the lifespan (McCrae & Costa, 2003), and are highly valid, predictive of positive and negative affect, life satisfaction, marital satisfaction and stability, career success, work-family conflict, and even length of life (Ozer & Benet-Martinez, 2006). Thus, neither the reality of traits nor their placement within psychological variables is in doubt.

Less evidence is available about whether, how, and how strongly traits become manifest in behavior (Furr, in press; Wu & Clark, 2003). Although there is a growing volume of evidence that traits have validity in predicting behaviors, this evidence consists mostly of unconnected facts and does not include an assessment of the relationship between traits and comprehensive distributions of behavior.

There are four current sources of knowledge about the trait-behavior relationship. One source consists of a growing number of studies in which one or a few specific behaviors are predicted by scores on a given trait (e.g., Borkenau & Mueller, 1992; Funder & Colvin, 1991; Mehl et al., 2006; Paunonen & Ashton, 2001). For example, Jensen-Campbell and colleagues (2002) showed that agreeableness predicted effortful control as measured by reaction times and performances on a card-sorting task. These studies show that there are at least some implications of trait standing for behavior. However, they do not predict the kind of comprehensive distributions of representative behaviors needed to answer the five basic questions about implications of trait standing for acting in the way described by the trait. Specifically, they predict only one or a few instances of behavior, often in the artificial setting of the lab, which may or may not be representative of how the individual acts in general and which does not allow calculations of patterns or trends in behavior.

The act frequency approach (Buss & Craik, 1983), which shares many goals with the current research, is a second source. It describes individuals by the frequency with which they enact prototypical acts of a given trait. After one group of subjects generates prototypical acts for a trait, other subjects retrospectively self-report how often they enact each of the prototypical acts. Thus, the act frequency approach shares with the present research the goal of determining how traits are manifest in behavior. However, the Act Frequency approach (a) attempts to identify categories of actions that manifest a trait, and (b) uses retrospective questionnaire measures of behavior frequencies, rather than reports that are made as the behavior is occurring, in actual situations.

Another important source of information about the consistency of behavior has arisen in the course of the person-situation debate (Epstein, 1979; Fleeson, 2004; Funder & Colvin, 1991; Shoda, Mischel, & Wright, 1994). However, most evidence relevant to this debate has described consistencies between behaviors rather than consistencies between behavior and trait standing as assessed by questionnaire. Mischel (1968) reviewed the available, limited evidence

to conclude that, typically, a given behavior predicts another behavior in a different situation at about a .3 level (Ross & Nisbett, 1991, later declared the upper limit to be .4). This conclusion was taken widely as evidence that the predictive power of traits was weak, because .3 was considered small, amounting to 9% of the variance in behavior being explained by personality. (There is reason to disagree with considering these effect sizes to be small, Funder & Ozer, 1983; Meyer et al., 2001.) Later, Epstein (1979) argued that aggregated behaviors are highly consistent, and he obtained much higher correlations predicting one aggregate from another aggregate. These higher correlations hold out hope that traits may be able to exceed .3 or .4 in their prediction of behavior, although many psychologists have concluded that .3 is the typical result and thus that traits are not useful (Ross & Nisbett, 1991).

However, all consistency studies predicted behavior from behavior. None of them predicted behavior from trait standing, leaving it unknown how trait standing predicts distributions of representative behaviors. (Epstein, 1979, included correlations between behavior and Guilford-Zimmerman Temperament Survey [Guilford & Zimmerman, 1949] scores but these correlations were low because the behaviors were not selected to represent those traits.)

A fourth source consists of three recent and important studies. Wu and Clark (2003) correlated frequencies of 55 specific behaviors (e.g., “hit something in anger”) recorded daily for two weeks with questionnaire measures of the specific traits of aggression, exhibitionism, and impulsivity. Each trait correlated about .4 to .5 with frequencies of corresponding behavioral items, showing that these three traits can predict behaviors reasonably strongly. Church, Katigbak, Reyes, Salanga, Miramontes, & Adams (2008) correlated Big 5 traits with the enactment of sets of ten specific trait-relevant behaviors on an average day, and found average correlations just under .4. Heller, Komar, and Lee (2007) calculated the correlation between averages of Big-5 states from an experience-sampling study and trait standing from a questionnaire, and found correlations around .4. These three studies are excellent examples of the kind of research we and others (Funder, 2001; Mischel, 2004) are arguing is important for the field.

Distributions of Accumulated Personality States

Answering all five questions about implications of trait standing requires that each individual's naturally-occurring trait manifestations be assessed in a comprehensive manner. Comprehensiveness entails at least three requirements: (1) assessing trait manifestation in behavior on numerous occasions of each individual in order to obtain a *wide range of each individual's trait manifestations in behavior*; (2) assessing behavior in naturally-occurring situations in order to obtain a *representative assessment of trait manifestations* (Bowers, 1973); and (3) having a *complete assessment apparatus* to ensure that all behavior that manifests a given trait is assessed and that no manifestations of traits are systematically missed. One way to accomplish this latter requirement would be a preliminary series of studies to catalog every type of behavior that manifests each trait. The enormity of this task may be one reason why personality psychology has yet to produce such knowledge about trait implications for distributions of naturally-occurring behavior, especially when combined with the fact that relevance of a given behavior for a given trait is likely context-dependent (e.g., slapping someone on the back may be an aggressive or a friendly act, depending on the context, Funder, 1991).

The present research proposes an alternative solution to this problem: rather than specify concrete activities to record, this approach directly assesses the degree to which the behavior expresses the content of the trait. This assessment is termed “personality state” assessment (Cattell, Cattell, & Rhymer, 1947; Fleeson, 2001; Fleeson, Malanos, & Achille, 2002; Fridhandler, 1986). A state is defined as having the same affective, behavioral, and cognitive

content as a corresponding trait (Pytlik Zillig et al. 2002), but as applying for a shorter duration. For example, an extraverted state has the same content as extraversion (talkativeness, energy, boldness, assertiveness, etc.), but applies as an accurate description for only a few minutes to a few hours rather than for the months or years that a trait description applies. This operationalization highlights the possibility that most trait terms may have originated as descriptions of behavior; only later may they have been generalized as traits (Eysenck & Eysenck, 1985). For example, to say that an individual is extraverted may be a generalization from his or her typical behavior.

Behavior will therefore be assessed with the same adjectives and numeric rating scales as are used to assess traits. Just as trait extraversion is assessed by self-reports of how talkative, bold, and assertive an individual is in general, e.g. from 1 to 7, state extraversion will be assessed by self-reports of how talkative, bold, and assertive an individual is at the moment, from 1 to 7. This definition of a state avoids the problem of trying to ascertain the concrete behaviors that fit into a trait. Rather, it transfers the content of the trait as a whole to the state. Thus, states are directly commensurate with traits and the five questions about implications of trait standing for trait manifestation in behavior can be answered directly.

At the outset, we acknowledge that this is only one way to assess behavior and that there are limitations to this method. However, there are also advantages. The personality state concept facilitates a comprehensive assessment of each individual's trait manifestation in behavior in naturally occurring and real situations. Multiple occasions are manageable with this method because the questionnaire can be brief enough to be completed multiple times per day for several days without being burdensome. Because all behaviors are rated for manifestation of all traits, few manifestations of traits are missed. The accumulated frequencies at each level of manifestation are then combined to form a distribution for each individual of each trait. Thus, the distribution provides a comprehensive account of the individuals' trait manifestation in behavior across several days.

Which State Level is Implicated by Trait Standing?

Because the same individual acts at different state levels on different occasions, one goal of this paper is to determine which of those levels is the one implicated by trait standing. At least three levels are plausibly implicated by trait standing: the central level, the most frequent level, and the most extreme level.

The central level of an individual's distribution of states is plausibly the one implicated by trait standing, because trait standing may represent a summary of the distribution, and the central tendency is a very good summary. Central tendency is a good summary of an individual's varied states because it balances out his or her higher scores with his or her lower scores, leaving the middle score. The median balances the frequencies of deviations whereas the mean balances the degree of deviation. The central tendency is consistent with a model of a trait as a causal force that determines behavior in conjunction with varying situational forces (e.g., Epstein, 1979), among other models.

The most frequent state level (i.e., the mode) may also be the level implicated by trait standing. Individuals may be describing their most frequent way of being when they complete trait questionnaires. The mode is also consistent with a model of traits in which a given level is the default level at which an individual acts unless there are pressures to deviate from that level.

An extreme level may also be the level implicated by trait standing, because trait standing may represent the amount of trait present when pushed to the limits, or the highest level the individual is capable of achieving (Paulhus & Martin, 1987; Wallace, 1966). The minimum may be the level implicated by trait standing because trait standing may represent the

individual's base or starting level of the trait, or the amount of a trait an individual possesses in the absence of any other forces.

Additionally, trait standing may have implications not only for a single level in the distribution but also for the width and shape of the distribution. Thus, we will test whether trait standings predicts individuals' standard deviations of states (width), skews of states (shape), and kurtoses of states (shape).

Overview of Studies and Analysis Plan

Fifteen studies will be reported in this manuscript. Each study has trait scores from standard questionnaires. Each study also has an experience-sampling component, in which participants report their current Big Five states several times a day for several days. The studies were designed for separate, individual purposes so there are variations across studies, including which traits and adjectives were assessed, the length of the experience-sampling portion, whether other questionnaires were included, and the number of participants (given the large number of data points per participant, each study has a small number of participants). The studies are by and large equivalent to each other, but unique features of individual studies that may have theoretical relevance will be pointed out. In addition, data from some studies have been published elsewhere for other purposes and data from some of the studies will be published elsewhere in the future for other purposes. Unlike many meta-analyses, all methodological details are known for the included studies, and these details are summarized for each of the 15 individual experience-sampling studies. Furthermore, all studies conducted in this lab on this topic between 1997 and 2004 were included, without prejudice, as long as at least 10 subjects participated and all necessary data (questionnaire and experience sampling) were available.

Studies 1 – 15

This is an unusual meta-analysis, in that almost none of the specific results in the current findings have been previously published. As a result, we describe the methods and results of the individual studies as they pertain to the current investigation in slightly more detail than is typical for a meta-analysis. However, because the first author was involved in all studies, they are very similar to each other methodologically. Therefore, we first describe the typical method employed in the 15 experience sampling studies. Table 2 lists the specific details of each study, such as the number of subjects. Then, a brief paragraph about each study will describe anything atypical about the study, the number of participants and reports excluded for data quality, and the reliability of the scales. Finally, the results for each individual study will be included in the Results section for the Mega-analysis.

Studies 1 – 15 General Methods

Participants—Participants were college-aged, male and female students at Wake Forest University. They participated in partial fulfillment of the requirements for an introductory psychology course. After excluding participants for the various reasons described below, the number of subjects in each study ranged from 12 to 63 (mean = 33.00, SD = 15.80).

Materials: standard trait questionnaires—Participants typically completed a standard Big Five questionnaire sometime before the experience sampling portion of the study. These questionnaires had 3-12 items for each factor, and participants were instructed to describe their behavior in general. The Big-Five are appropriately assessable with a large variety of adjectives (Goldberg, 1992); for these studies, adjectives were chosen that loaded on the correct factor in Goldberg (1992), were reliable in previous work, were distinct from each other, were easy to use to describe behavior, and had a minimal social desirability component. In most studies, at least one item per factor was reverse-keyed, so that a lower score indicated higher standing on

the factor (for example, one emotional stability item was “insecure”). Participants responded on scales of 1 to 7 or 1 to 6, with higher numbers meaning that the adjective was more descriptive. Wherever possible, all participants who completed the questionnaire were included in calculating reliabilities; sometimes only the participants who participated in the ESM portion of the study were included.

Experience sampling—Typically, participants reported on their behavior several times per day for one to two weeks, describing how they were acting during the previous 20 to 60 minutes. These reports were completed every two to four hours, and took about 1 to 2 minutes to complete. Most reports were scheduled for afternoon, evening, and night; for example, one typical schedule was noon, 3pm, 6pm, 9pm, and midnight. Each question appeared on a small Palm Pilot screen and participants responded by pressing a number with a plastic stylus. To encourage timely completion, participants regularly downloaded their data (usually every two days).

The first report normally occurred during an introductory session in which the procedure was explained. The unique nature and comprehensiveness of this methodology was stressed, and participants were urged to honestly complete as many reports as possible. At the end of these sessions participants were invited to withdraw for partial credit.

Participants were normally instructed to miss a report if it conflicted with an important task (e.g., driving, during an exam, while sleeping). Participants were also told they could complete a report earlier or later than the scheduled time (normally a few minutes early to two to three hours late). Reports were excluded if they were completed outside this time window (as recorded by the Palm Pilot), or were beyond the maximum number scheduled per day; thus, all reports were completed close in time to the described behavior. Reports were also excluded that included too many missing values (e.g., more than six) or too many of the same value (e.g., more than 20).

Personality states were assessed with the same format as traditional adjective-based Big-Five scales with the exception that, rather than describing themselves in general, participants described their behavior during a brief period (e.g., “During the last half-hour, how hard-working have you been?”). Unless otherwise stated, state reports used the same or a subset of adjectives as were used on the trait questionnaire. Because some studies assessed situational variables or affect, some studies assessed personality states with three items, and some studies only assessed personality states for some trait domains (to minimize the length of state reports). In most studies, participants responded with scales of 1 to 7 or 1 to 6. In order to conserve space, the specific items used in each study are not listed, but they are available upon request to the first author. A large variety of adjectives were included, selected from several sources (e.g., Goldberg, 1992; Saucier, 2002). Reliabilities are reported with each study.

Unique Methods for Each Individual Study

Study 1, ‘ESM Situations’—This study was largely similar to the standard study, with the exception that, because it assessed situations and affect in addition to behavior, only three traits (extraversion, agreeableness, and conscientiousness) were assessed in the ESM portion. Three participants were excluded for completing fewer than 20 reports, and 232 of the 1201 reports (19.3%) were excluded for one of the exclusionary reasons described above. Cronbach's alphas for the questionnaire measure were: E alpha = .53, A alpha = .76, C alpha = .58, ES alpha = .81, I alpha = .70. Cronbach's alphas for the ESM portion (calculated across all reports and participants) were: E alpha = .68, A alpha = .64, C alpha = .70.

Study 2, ‘ESM Standard 1’—Study 2 followed the standard methods. Two participants were excluded for not completing the standard Big-Five scale and three more were excluded

for completing fewer than 20 ESM reports, and 231 of the 2066 reports (11%) were excluded for one of the exclusionary reasons. Cronbach's alphas for the questionnaire were: E alpha = .78, A alpha = .78, C alpha = .79, ES alpha = .73, I alpha = .71. The ESM portion used a subset of the questionnaire adjectives, with the exception of replacing “imperturbable” and “uninquisitive” with “perturbable” and “inquisitive”. Cronbach's alphas for the ESM portion were: E alpha = .72, A alpha = .66, C alpha = .66, ES alpha = .62, I alpha = .69.

Study 3, ‘ESM Standard 2’—This study followed standard procedure. In total, 371 of the 1697 reports (21.9%) were excluded for one of the exclusionary reasons. “Trustful” was found not to be strongly related to the other four Agreeableness items (item-total correlation = -.18) and so it was excluded when calculating Agreeableness scores on the questionnaire. Cronbach's alphas for the five traits on this questionnaire were: E alpha = .79, A alpha = .69, C alpha = .86, ES alpha = .72, I alpha = .71. On ESM reports, participants were able to answer “na” if the item was not applicable. Cronbach's alphas for the ESM portion were: E alpha = .68, A alpha = .69, C alpha = .77, ES alpha = .68, I alpha = .74.

Study 4, ‘ESM Goals 1’—This study assessed goals and affect, leaving limited room for state assessment. 182 of the 1373 reports (13.3%) were excluded for exclusionary reasons. “Warm” was found not to be strongly related to the other two Agreeableness items (item-total correlation = -.16) and “intelligent” was found not to be strongly related to the other two Intellect items (item-total correlation = .04), and so these two items were excluded when calculating Agreeableness and Intellect on the trait questionnaire. Cronbach's alphas on the trait questionnaire were: E alpha = .62, A alpha = .29, C alpha = .73, ES alpha = .78, I alpha = .77. The ESM portion used the same 15 items that were used in the standard questionnaire, except “energetic,” which was replaced by “enthusiastic”. Cronbach's alphas for the ESM portion were: E alpha = .65, A alpha = .46, C alpha = .55, ES alpha = .68, I alpha = .70.

Study 5, ‘ESM Goals 2’—This study assessed goal pursuit and progress in addition to states, so only three of the Big-Five traits were assessed. 244 of the 1156 reports (21%) were excluded for one of the exclusionary reasons. “Unshakable” was found not to be strongly related to the other five Emotional Stability items (item-total correlation = -.02) and so was excluded when calculating Emotional Stability scores on the trait questionnaire. Cronbach's alphas for the traits were: E, Cronbach's alpha = .67, A alpha = .83, C alpha = .71, ES alpha = .65, I alpha = .62. The ESM reports used three of the six adjectives per trait that were used in the standard questionnaire. Participants were able to answer “na” if the item was not applicable. Cronbach's alphas for the ESM portion were: E alpha = .66, C alpha = .76, ES alpha = .73.

Study 6, ‘ESM Situations Long’—This study was much longer, but assessed situations, limiting assessment of states. 1163 of the 5688 reports were excluded for one of the exclusionary reasons. “Frivolous” was found not to be strongly related to the other five Conscientiousness items (item-total correlation = -.19), and so was excluded when calculating Conscientiousness scores on the trait questionnaire. Cronbach's alphas for the traits were: E alpha = .70, A alpha = .71, C alpha = .55, ES alpha = .73, I alpha = .69. Participants could indicate “irrelevant” for a particular adjective. Cronbach's alphas for the ESM portion were: E alpha = .61, C alpha = .51, ES alpha = .55.

Study 7, ‘ESM Adolescence’—This study assessed middle-school children (mean age = 12.9 years). Six participants were excluded and 445 of the 1080 reports (39%) were excluded for one of the exclusionary reasons. Cronbach's alphas for the trait questionnaire were: E alpha = .69, A alpha = .66, C alpha = .51, ES alpha = .67, I alpha = .61. “Rude” was not strongly related to the other four Agreeableness items (item-total correlation = .22) and was excluded

when calculating ESM Agreeableness scores. Cronbach's alphas for the ESM portion were: E alpha = .75, A alpha = .67, C alpha = .73, ES alpha = .53, I alpha = .67.

Study 8, 'ESM Opposites'—Five participants were excluded for not completing standard trait questionnaires and 294 of the 834 reports (35%) were excluded for one of the exclusionary reasons. “Frivolous” was not related to the other five Conscientiousness items, so it was excluded when calculating Conscientiousness scores on the trait questionnaire. Cronbach's alphas for the five traits were: E alpha = .68, A alpha = .75, C alpha = .74, ES alpha = .87, I alpha = .72. “Compliant” was found not to be strongly related to the other four state Extraversion items (item-total correlation = -.27) and was excluded when calculating ESM Extraversion scores. Cronbach's alphas for the ESM portion were: E alpha = .75, A alpha = .68, C alpha = .61, ES alpha = .85, I alpha = .68.

Study 9, 'Weekly'—In this study, participants completed ESM reports just once each week, for ten weeks. Cronbach's alphas for the five traits on the trait questionnaire were: Cronbach's alpha = .72, A alpha = .81, C alpha = .72, ES alpha = .69, I alpha = .76. Cronbach's alphas for the ESM portion were: E alpha = .68, A alpha = .80, C alpha = .68, ES alpha = .84, I alpha = .80.

Study 10, 'ESM Extraversion'—Because this study was primarily designed to measure aspects of the situation and goals, the only Big-Five trait measured in the ESM portion was Extraversion. Six participants were excluded for not completing enough valid ESM reports, and 232 of the 813 reports (28.5%) were excluded for one of the exclusionary reasons. “Imperturbable” was not strongly related to the other four Emotional Stability items (item-total correlation = .01) and was excluded when calculating questionnaire Emotional Stability scores. Cronbach's alphas for the five traits on this questionnaire were: E alpha = .78, A alpha = .69, C alpha = .80, ES alpha = .75, I alpha = .58. A “not applicable” option was also available. Cronbach's alpha for the ESM portion was .87 for Extraversion.

Study 11, 'ESM Triple Length'—Because of the length of this study (up to 59 days), only participants who scored high in Conscientiousness on the standard Big-Five questionnaire were selected to participate in the ESM portion. Also, to ensure that all participants used the behavior-descriptive adjectives the same way, participants memorized the definitions of the adjectives. 176 of the 2384 reports (7.4%) were excluded for one of the exclusionary reasons. “Unemotional” was not strongly related to the other four Emotional Stability items, and was excluded when calculating Emotional Stability scores on this questionnaire. Reliabilities for the traits (for all 269 participants who completed this scale) were: E, Cronbach's alpha = .76, A alpha = .75, C alpha = .72, ES alpha = .65, I alpha = .77. Imperturbable was not strongly related to the other four Emotional stability items (item-total correlation = .11), and was excluded when calculating ESM Emotional Stability scores. Cronbach's alphas for the ESM portion were: E alpha = .61, A alpha = .66, C alpha = .67, ES alpha = .63, I alpha = .36.

Study 12, 'ESM Adulthood'—Participants included non-traditional students (age range 18 to 51) who participated as part of their Theories of Personality course. Data from one participant who repeated the ESM portion is not included; 99 of the 2443 reports (4%) were excluded for one of the exclusionary reasons. Because this study was designed to assess affect and features of the situation, only three of the Big-Five traits were assessed during the ESM portion of the study. Traits were measured with the NEO rather than with adjectives in this study (Costa & McCrae, 1985). Cronbach's alphas for the five traits were: E alpha = .80, A alpha = .73, C alpha = .85, ES alpha = .87, I alpha = .71. Adjectives were used to assess states during the ESM portion of the study. “Unsympathetic” was found not to be strongly related to the other two Agreeableness items (item-total correlation = .25) and “disorganized” was found not to be

strongly related to the other two Conscientiousness items (item-total correlation = .00) and these two items were excluded in calculating their respective ESM trait scores. Cronbach's alphas for the ESM portion were: E alpha = .58, A alpha = .70, C alpha = .61.

Study 13, 'ESM Defined'—To ensure that all participants used the behavior-descriptive adjectives the same way, participants memorized the definitions of the adjectives used to describe their behavior. Five participants were excluded for not completing the standard questionnaire and three more for contributing too few valid ESM reports; 126 of the 1505 reports (8.4%) were excluded for one of the exclusionary reasons. “Unreflective” was found not to be strongly related to the other four Intellect items (item-total correlation = -.15), and so it was excluded when calculating Intellect scores on this questionnaire. Cronbach's alphas for the five traits on this questionnaire were: E, Cronbach's alpha = .89, A alpha = .79, C alpha = .79, ES alpha = .83, I alpha = .55. In this study, some participants completed the ESM portion using Palm Pilots, and others used paper diaries. Cronbach's alphas for the ESM portion were: E alpha = .75, A alpha = .79, C alpha = .77, ES alpha = .74, I alpha = .51.

Study 14, 'ESM Lab'—The ESM portion of the study was conducted in the laboratory. Participants participated in 10 50-minute sessions over the course of 10 weeks in groups of three or four. After each 20 minutes, participants rated their behavior during the preceding 20 minutes. Four participants were excluded for providing fewer than seven valid reports of behavior. Reliabilities for the five traits on this questionnaire were: E alpha = .76, A alpha = .59, C alpha = .39, ES alpha = .63, I alpha = .70. Cronbach's alphas for the ESM portion were: E alpha = .83, A alpha = .69, C alpha = .62, ES alpha = .60, I alpha = .74.

Study 15, 'ESM Standard 3'—210 of the 2204 reports were excluded for one of the exclusionary reasons. This study used bipolar adjectives. Reliabilities for the five traits on this questionnaire were: E, Cronbach's alpha = .77, A alpha = .69, C alpha = .89, ES alpha = .66, I alpha = .66. Some participants completed the ESM portion on palm pilots and others on paper. Cronbach's alphas for the ESM portion were: E alpha = .78, A alpha = .83, C alpha = .86, ES alpha = .76, I alpha = .68.

Study 16, Mega-analysis of All 15 Studies

Method

Selection of studies—With one exception, all studies included in this mega-analysis were conducted by the first author or his collaborators. All studies available at the inception of this paper were included without prejudice as long as they contained at least ten participants and the two essential types of data: (i) standard Big-Five questionnaire scores and (ii) Big Five ratings of behavior during experience sampling. Of 21 studies available, six studies were not included, leaving 15. Four were not included because standard questionnaire data were not available, one study was not included because participants were explicitly instructed to center their ratings on the number 4, eliminating any individual differences in behavior, and one study was excluded because there were only nine participants.

Combining data—Because we collected all of the data ourselves, we had the rare advantage of having access to all raw data. Thus, we were able to combine all data sets into one file for analysis for a “mega-analysis”, in addition to the more traditional meta-analysis (Steinberg et al., 1997). Participants in the 15 studies used several different questionnaires to describe their behavior. For ratings on the standard questionnaires, one study had a 1 to 5 scale, two studies had a 1 to 6 scale, and twelve studies had a 1 to 7 scale. For ratings during the ESM reports, four studies had a 1 to 6 scale and eleven studies had a 1 to 7 scale. To equate the scales across studies, we POMP transformed scores on both standard questionnaires and ESM reports

(Cohen, Cohen, West & Aiken, 2003). Numbers were re-scaled so 100 was the maximum possible, 0 was the minimum possible, and each intermediate score represented the portion of 100, or the “Percent Of Maximum Possible”. This resulted in scores on a scale of 0 to 100 on each of the Big Five factors for all standard questionnaires and ESM reports. Because POMP is a linear transformation, it has no effect on the results, other than making them more interpretable and comparable. The scores obtained on each Big Five domain for each standard questionnaire, or for each ESM report, were then combined to form the data sets for mega-analysis.

Combining the data resulted in 21,871 concatenated ESM reports provided by 495 participants from 15 different studies. Each participant provided an average of 44.18 (SD = 35.32) reports. With the exception of Extraversion, which was assessed in every study [N (participants) = 495, n (reports) = 21,798 (not including reports for which the Extraversion score was missing), average number of reports = 44.04], different studies assessed different subsets of Big-Five traits. Agreeableness was assessed in 12 studies (N = 399, n = 15,388 average number of reports = 38.57), Conscientiousness in 14 studies (N = 474, n = 20,922, average number of reports = 44.14), Emotional Stability in 12 studies (N = 386, n = 17,827, average number of reports = 46.18), and Intellect in 10 studies (N = 311, n = 12,544 average number of reports = 40.33).

Missing data—For state reports that had missing items, the state score was calculated with the completed items. For some state reports, state scores were missing for one or more Big Five domains (state scores would be missing if all items were not completed or marked “n/a”). As a check of the potential impact of missing data, we counted missing state scores for each trait. For Extraversion, 73 of 21,871 possible state scores were missing (0.33%; average number of missing scores per participant = 0.14). For Agreeableness, 497 of 15,885 possible state scores were missing (3.13%; average number of missing scores per participant = 1.24). For Conscientiousness, 368 of 21,290 possible state scores were missing (1.73%; average number of missing scores per participant = 0.77). For Emotional Stability, 150 of 17,977 possible state scores were missing (0.83%; average number of missing scores per participant = 0.38). For Intellect, 28 of 12,572 possible state scores were missing (0.22%; average number of missing scores per participant = 0.09). Given the small percentages of missing state scores, it is unlikely that these missing data, even if they are not missing at random, would have any appreciable effect on the relationships between parameters of density distributions and trait scores. Consequently, density distributions were simply calculated with available state scores (missing data were not replaced or imputed).

Centering—It is possible that study differences in adjectives, time of year, or other administrative factors led to differences in mean-level of endorsement of the traits (or in other parameters such as variability, etc.). Such study differences could artificially inflate correlations, if they affected both the questionnaire scores and the ESM scores of a given study in one direction. For example, if the items in a particular study led to higher means in the ESM portion and also higher means on the questionnaire than for other studies, then the correlation between the means would be artificially increased because of the between-study differences. However, we are interested only in between-individual differences. To correct for this possibility, questionnaire scores and parameters were centered within-study before computing correlations. For the analyses predicting single states, single states were centered within-study before computing correlations. It is important to note that this correction eliminates any legitimate differences between studies in the trait levels of individuals, and thus may lead to a slight underestimate of the true correlations.

Results

Table 3 shows two parameters of the density distributions for each study and each trait separately. The mean refers to the average person's average level of the state across the ESM portion of the study. Because the data were POMP transformed, 50 would mean that the average state is equally close to each pole of the trait; numbers higher than 50 mean that the average person's average state is closer to the high end of the trait to the low end of the trait. The standard deviation refers to the amount of within-person variability for the average person in the study. A high standard deviation means that the average person varies quite a bit in how much his or her behavior manifests a given trait.

The mega-analysis line of Table 3 shows the average across all available individuals (no centering was done here because there were no correlations that could be artificially inflated). This line shows what the behavior of a typical individual is like, characterized by the five dimensions of the Big 5. The average person's mean is close to the midpoint for extraversion and intellect, but much closer to the agreeable, conscientious, and emotionally stable end than to the disagreeable, unconscientious, and neurotic ends, respectively. The average person's SD is high for each of the traits, especially for extraversion and conscientiousness, meaning that the typical individual routinely and regularly expressed manifested levels of most traits in his or her behavior (Fleeson, 2001). For example, the SD of 17.37 for extraversion means that, assuming a normal distribution, the typical individual needed a range of about 70 to cover 95% of his or states. Because the scale is only 100 points wide, the typical person covered most of the possible ranges of extraversion states routinely and regularly.

Comparing the variation within people to the variation between people aids evaluation the extent to which individuals deviate from their trait standing. Table 3 also shows the between-person standard deviations of within-person means and standard deviations; this line indicates how much people differed from each other in their distributions. For four of the five traits, the within-person standard deviation was higher than the between-person standard deviation. Additionally, we conducted unconditional multilevel models in order to partition the total variance in states into that portion that represents differences between people in which traits they manifest and that portion that represents differences within each person in which traits they manifest on different occasions. As shown in Table 3, most of the variation in trait manifestation in behavior is within-person variation, that is, the same person manifesting different traits at different moments.

This high degree of within-person variability means that the majority of the time, participants reported acting at a different level than their typical levels. This lowers the likelihood that questionnaire scores will be able to predict such behaviors accurately, and verifies one source of concern about the ability of traits to predict trait manifestation in behavior.

Single states—Each participant's scores on standard Big-Five questionnaires were correlated with one of their randomly selected state reports. Ten trials were repeated with a different random selection of states in each trial. The average of these correlations are shown in Table 4. Several aspects of these results are important. First, the correlations are all significant. Thus, trait standing does significantly predict trait manifestation in behavior. Second, the magnitudes of the correlations are all below the traditional .40 barrier, consistent with the assumed but never tested claim that trait standing does not as strongly predict naturally-occurring trait manifestation in behavior at any given randomly selected moment. Third, however, the correlations vary from .18 for extraversion to .37 for intellect, with three of the traits having correlations over .30, suggesting that the typical correlation for predicting even single states is respectably high. This also means that participants do not simply report their trait standing on the ESM reports, but rather vary their responses from moment to moment.

Central tendency—Given the large variability in states, it is necessary to summarize them in some way. We first summarized states by the central tendency, using three different operationalizations of central tendency. The central tendency represents the location or center of an individual's distribution, and the three operationalizations differ in how they conceptualize the center. The mean represents the balancing point of the distribution, such that the sum of deviations above the mean is matched by the sum of deviations below the mean; the mean also represents the important aggregate of states (Epstein, 1979). The correlations between means and trait standing, as shown in Table 4, reveals that trait standing strongly predicts mean levels of trait manifestation in behavior. All correlations exceeded .40, and one reached as high as .56. This means that trait standing is a powerful predictor of averaged trait manifestation in behavior.

Correlations with the median of the distribution were almost as high, ranging from .41 to .55. The median is also a balancing point, but balances the number of deviations above and below rather than the magnitude of deviations. Finally, correlations with the mode were not as high, but still higher than correlations with single states, ranging from .28 to .48. These correlations suggest that trait standing refers to the central tendency of states.

Because trait standings were correlated across traits, we also wanted to determine the unique relationship of each Big-Five trait standing with the mean. Participants' means for each state were regressed onto the five questionnaire trait standings. These correlations were reduced slightly in magnitude, but remained high. Thus, intercorrelations among traits do not explain the strong predictive power of traits.

Alternate summaries: Maximum and minimum—The most extreme behavior is also a possible characteristic of the behavioral distribution that corresponds to trait standing. For example, the maximum state level may be implicated by competence views of traits and the minimum state level may be implicated by baseline views of traits (i.e., that trait standing reflects the amount of a trait present when no other factors are acting on behavior). Table 4 shows that both the maximum and the minimum state levels were predicted by trait standing. The magnitudes of predictions of the maximum state were similar to that for central tendency, ranging from .34 to .54, whereas the predictions of the minimum state were more similar to predictions of randomly selected states, ranging from .22 to .37.

Comparison of mean, median, mode, maximum, and minimum—In the above bivariate relationships, the mean appeared to be the strongest correlate of the trait standing. In order to determine whether the other summaries are related independently to trait standing, we conducted a series of multiple regressions. Each multiple regression predicted the trait standing from the mean and from one other summary, one trait at a time (note that this puts the questionnaire score into the DV position). For none of the five traits was the mean or mode significantly related to trait standing when holding the mean constant. For only one trait (intellect), did the minimum have a significant independent relationship to trait standing. However, for four of five traits (and with a p value of .06 for the fifth trait), the maximum was significantly related to trait standing even while controlling for the mean. In all 20 tests, the mean remained significantly related to trait standing, even while controlling for any other summary. These results suggest trait standing, as assessed by questionnaire, implicates primarily individuals' mean trait manifestation, but also provides independent information about individuals' maximum trait manifestation in behavior.

Other parameters of distributions—Trait standing may have implications not only for a single level in the distribution but also for the width and shape of the distribution. Table 4 shows that trait standing was only weakly related to the amount of variability in states, with correlations ranging from -.07 to .23. To be sure that a curvilinear relationship between trait

standing and variability was not underlying the weak linear relationship, we also correlated variability in states with the quadratic function of trait standing (while controlling for the linear function). These correlations were also very small, with only one being significant.

Table 4 shows that even skew, correlations ranging from $-.18$ to $-.28$, and kurtosis, correlations ranging from $-.02$ to $.27$, of state distributions had relationships to trait standing. However, these relationships were small.

Study length—It is possible that repeated completion of reports may have caused scale-use drift, fatigue, or increased accuracy. We divided each participant's reports into the first 24 completed (typically about 1 week's worth), the next 24 completed (typically a second week), and the remainder (note that studies varied in the number of reports participants completed). There were no significant differences between means of four of the five traits due to duration, suggesting that participants maintained remarkably stable standards across the studies. Conscientiousness was the only exception, increasing from $.61$ in the first week to $.66$ in the third week. This might be explained by the fact that the longest study preselected participants to be high in conscientiousness. Within-person standard deviations differed slightly and significantly from the first week to the second for all five states, and again to the third week for two more states. Standard deviation dropped about 1-2 points per week, out of standard deviations that ranged from about 12 to 18. This means that people were less variable in later periods of the studies, possibly suggesting a change in scale use. Finally, correlations between questionnaire scores and average state levels were strongest in the first week, slightly lower in the second week, and weakest for longer durations. Specifically, the three correlations for each week, respectively, were $.42/.32/.28$ (extraversion), $.54/.55/.50$ (agreeableness), $.45/.43/.47$ (conscientiousness), $.53/.46/.38$ (emotional stability), and $.55/.50/.42$ (intellect). These results suggest future research into the effects of study length on the quality of the data.

Effects of study factors—In a typical meta-analysis, the question of whether study factors are related to differences between study results would be addressed by formal tests. Such tests would compare groups of studies at different levels of the factor in question (for example, if some studies were conducted in the summer and some in the winter, a two-level variable would be created, and the variance in studies at one level (summer studies) would be compared to the variance in studies at the other level (winter studies) using a significance test). Although the total number of participants and of ESM reports in the present study is substantial, the number of studies (15) does not allow detection of study-level differences by this process. However, we selected certain study factors that seemed particularly relevant to the analyses, and then re-computed the central analyses for these select studies. The results for these studies, although not compared to the results for the other studies in a formal test, can provide some information about the potential effect of the identified study factors on the results. Because these analyses are only tentative, we limited them to the prediction of the average (mean) state level, which was the most predictable in the earlier analyses.

First, lower reliabilities of either the standard trait questionnaires or of the state measures are likely to reduce and put a ceiling on the correlation between the traits and behavior. Table 2 reports any cases in which reliabilities for a given scale were below $.65$. Although many studies had a couple of traits out of 10 that were low in reliability, three studies did not have any traits in either the self-report questionnaire or the ESM portion with reliabilities below $.65$ (ESM Standard 2, Weekly, and ESM Standard 3). We recalculated the correlation between questionnaire scores and the mean for those three studies combined, and the correlations were stronger. Specifically, the correlations were $.55$ (extraversion), $.59$ (agreeableness), $.53$ (conscientiousness), $.58$ (emotional stability), and $.68$ (intellect). These three studies suggest the average questionnaire to mean correlation approaches $.60$, even stronger than the already strong results for the mega-analysis as a whole. As an alternate procedure for considering

reliability, we recomputed the correlations between trait scores and mean state ratings, using only scales with reliabilities equal to or greater than .65. The resulting correlations were: for extraversion, .48 (N (studies) = 10, n (participants) = 311); for agreeableness, .55 (N = 9, n = 293); for conscientiousness, .50 (N = 7, n = 226); for emotional stability, .48 (N = 7, n = 220); for intellect, .52 (N = 7, n = 256). The average of these correlations is .51, very similar to the average obtained using all scales.

Second, in many studies, states were measured with slightly different items from or a subset of the items used to assess traits on the questionnaire. However, in six of the studies, the items were the same in the ESM portion as they were in the trait questionnaire, except for an occasional single item that did not cohere with its trait (ESM Opposites, Weekly, ESM Extraversion, ESM Defined, ESM Lab, ESM Standard 3). For these six studies, correlations between trait scores and means of corresponding states were again very high: .51 (extraversion), .59 (agreeableness), .56 (conscientiousness), .55 (emotional stability), and .60 (intellect).

Third, we considered the role of the described time frame for state reports, because the length of time described may affect the accuracy of the description. The correlations between trait scores and the mean for studies in which the described frame was the previous twenty or thirty minutes were the lowest: .24 (extraversion), .44 (agreeableness), .23 (conscientiousness), .39 (emotional stability), and .32 (intellect). Correlations between trait scores and mean reports based on the previous hour were similar to those for all studies combined: .41 (extraversion), .58 (agreeableness), .49 (conscientiousness), .51 (emotional stability), and .67 (intellect). This was also true for studies with reports describing the previous three hours: .52 (extraversion), .50 (agreeableness), .56 (conscientiousness), .57 (emotional stability), and .40 (intellect). These results suggest that 1-hour and 3-hour time frames may provide more accurate reports. It could be that describing longer time frames is more like describing trait standing. Longer lengths of time may include more behaviors that participants summarize over, and this summarization process may be similar to how participants answered trait questionnaires. However, the 1-hour reports were slightly better in these 15 studies than were the 3-hour reports (average = .53 vs. .51 for 3-hour reports). Unfortunately, any statistical test of study-level factors would have an N of 15, and so we are unable to test these differences statistically. Explaining these different results for different time frames may be a fruitful area for further research.

Figure 1 depicts the average of the correlations between traits and average states. Correlations were z -transformed, averaged, and the average was un-transformed. Figure 1 shows that traits were strongly predictive of average trait manifestation in behavior. Even when all 15 studies were included, the correlations were strong. However, when limited to studies with high reliability or high item overlap, correlations were even stronger, approaching .60.

Meta-analysis

As a second, more traditional, approach to combining all 15 studies, we also computed all results separately by study, and then combined the separate results to obtain the meta-analysis result. We limited these analyses to the predictions of single states (Table 5), means (Table 6), and maximums (Table 7) because of their theoretical relevance, because the results are highly similar to those for the mega-analysis, as shown in Table 4, and to conserve space.

For these three analyses, the correlations between standard questionnaire scores and each behavior summary were first transformed from Pearson's r to Fisher's z' , then weighted by $N-3$ (Cooper, 1998) before the mean was calculated across all studies. Means were then transformed back to Pearson's correlation coefficients. Results for each study separately are shown, as well as the meta-analytic average. As can be seen, the results are very similar to those for the mega-analysis. Specifically, trait standing predicts single behaviors at a respectable but not strong

level, means at strong levels, and maximums at levels nearly as strong. These results verify the mega-analysis results with a different and more traditional analytic technique.

Discussion

A basic question in personality psychology has been whether what people say they do in general and in reflection predicts what they do when confronted with real situations and when their actions have real consequences. The situationist challenge has been that traits do not predict what people do – when people are put into real situations, the immediate pressures overcome any general tendencies. Thus, there is a need to assess trait manifestation in the many behaviors that are actually carried out in real situations, and compare those to the assessments of what people say they do in general and in reflection. Results showed that traits did indeed predict actual trait manifestation in behavior, that they predicted the central tendency of wide (high variability) distributions, and that they predicted such central tendencies with a degree of strength. Questionnaire measures of traits have compelling validity as measures of the manner in which individuals actually behave in real situations; the predictions are strong enough to support traits as important descriptions of the traits people actually manifest in their behavior.

The relationship between trait scores and distributions of corresponding trait-manifestation in behavior was specified in terms of five basic questions as shown in Table 1; these studies have provided answers to each of the questions. (1) There are indeed implications of trait standing for trait manifestation. Traits were consistently correlated with states. (2) The implications were surprisingly strong, much stronger than the .3 to .4 commonly believed to be the ceiling for predictive power of trait standing. The correlations of trait standing with mean levels of behavior were also much stronger than the observed .18 to .37 correlations between questionnaire scores and single states. Whereas a .30 correlation means that 9% of the variance in how individuals act at the moment is accounted for by their trait standing, a .50 correlation means that 25% of the variance in individuals' average state levels is accounted for by their trait standing – at least a doubling and perhaps a tripling of the explained variance. (3) Individuals do frequently and widely deviate from their trait standing in their behavior, rather than mainly acting within a small range; most of the variance in trait manifestation is within-person, not between-person. Thus, this behavior must be summarized in some way for single-number trait standings to predict it. (4) The summary that trait standing seems to predict most strongly is the central tendency of states, as assessed by the mean. The mean was slightly more consistent with trait scores than was the median, and the median did not predict trait score when controlling for the mean. This implies that when individuals complete questionnaires assessing the Big Five, they are indicating the balancing point of their states on the given dimension – the total amount they deviate from that level in the upper direction on that trait is about the same as the total amount they deviate from that level in the lower direction. For example, an individual who scores a 5 on a 7-point extraversion scale has a balancing point at 5: he or she acts extraverted at a 6 or 7 level about as often as he or she acts extraverted at a 1, 2, 3, or 4 level.

(5) A variety of other functions of states were assessed, including the mode, minimum, maximum, and standard deviation. The maximum was strongly related to trait level, and it significantly predicted trait level over and above the mean's prediction. Thus, trait scores may also provide independent information about individuals' maximum state levels, providing some support to capability models of traits. The higher end of the scales were scored as the more desirable poles of the Big Five traits (Anderson, 1971), and these more desirable states may reflect the way participants want to or strive to be. Thus, trait scores may also represent individuals' abilities to achieve their desired states.

One of the strongest reasons the person-situation debate has endured is the conviction that traits never predict behavior higher than the .3 or .4 level. One response to this claim was to argue

that correlations as low as .1 to .4 may in fact represent strong effect sizes. For example, most important health-maintenance behaviors and most powerful situations have effects lower than or equal to .3 (Funder & Ozer, 1983; Meyer et al., 2001). The studies in this paper brought empirical evidence to bear about whether the traits people have predict the traits they actually manifest in their behavior, across a wide swath of typical occasions and situations. The resulting correlations comfortably surpassed .3 and even .4. This evidence combined with the strong predictions of life outcomes (Ozer & Benet-Martinez, 2006) casts strong doubt on the contention that traits do not predict behavior or that they have a .3 to .4 ceiling. In fact, far from being irrelevant, traits appear to be necessary for a full understanding of behavior, given the large amount of variance in trait manifestation in behavior they predict.

However, trait standing as assessed by questionnaire was not identical to trait standing as revealed in everyday behavior, in at least two ways. First, the typical individual's distribution of states is very wide, meaning that the typical individual routinely acts in a diverse range of ways. Single-number accounts of an individual's behavior do not refer to or explain this part of the individual's behavior, which the results revealed to be the lion's share of the variability. Rather, they can describe only the individual's average (Epstein, 1979; Fleeson, 2004; Mischel & Peake, 1983).

Second, the .42 to .56 correlations are strong but are not the .7 to .9 correlations that would be expected between two identical constructs. For example, Fleeson (2001) showed that state averages at one time predicted state averages at another time in the .7 to .9 range. There are at least three types of possible explanations for the .42 to .56 correlation not being as high as the .7 - .9 test-retest correlation for state averages. The first type includes non-shared method variance between questionnaires and experience-sampling. These were very different tasks for the participants: summarizing themselves as wholes versus only at the moment, retrospective versus concurrent reports, and one versus multiple reports (Lucas & Fujita, 2000). Additionally, questionnaire scores describe the person's behavior across a long period of time, at least a year, and across a diverse set of settings; in contrast, experience sampling covers only a few weeks of behavior, usually in one setting (university). Second, the questionnaire scores may be partially invalid as measures of how a person acts. For example, they may be based on cherished theories about the self, or be faulty due to inaccurate recall and summarization of actual behavior. Third, trait scores may refer to more than frequency of being in a given state (Funder, 2001; Revelle, 1995; Saucier & Goldberg, 1996). The current results showed that trait scores refer independently to the maximum. Trait scores may also refer to unique, special behaviors, or only to behaviors in relevant situations.

An interesting finding was that within-person variability (but not mean levels) were slightly smaller with longer lengths of study. These results suggest that extensive self-monitoring and reporting of states may have affected the states. This may be a real effect such that monitoring one's own behavior changed that behavior to make it more desirable, fit preconceptions, or be more consistent (Korotitsch & Nelson-Gray, 1999). Although self-monitoring has long been known to produce desired behavioral changes, these changes have been in mean-levels (Korotitsch & Nelson-Gray, 1999), whereas the changes in the present studies were in variabilities. This effect could also be an artifactual effect, such that practice, fatigue, scale-use drift, increased self-concept clarity, or other effects change people's accuracy in self-reports over time. The effect was not large in the present studies, but future research is warranted on the effects of ESM on the monitored behavior.

Expected Behavior Distributions for Different Trait Standings

A basic descriptive goal of personality psychology is to determine the distributions of trait manifestation in behavior that are implied by given trait standings. For example, what is the

behavior of an extravert like, and how does it differ from the behavior of an introvert? Conversely, what distribution of extraversion manifestation in behavior qualifies one as an extravert or as an introvert?

The results of these studies allows computing expected distributions of states for any given level of trait questionnaires. For example, they allow computing the expected state distributions for those who score high on a questionnaire and comparing them to the expected state distributions of those who score low on the questionnaire, in order to determine how such questionnaire score differences translate into actual state differences in everyday life. Prediction equations were created from the standardized coefficients predicting means and standard deviations from z-scored questionnaire scores (in Table 4) and the grand means in Table 3. Plugging in +1 SD and -1 SD on questionnaire score produced the predicted state means and state standard deviations for those scoring high on the questionnaire (+1 SD) and those scoring low on the questionnaire (-1 SD).

Figure 2 provides normal distributions based on these resulting state means and state standard deviations for each trait. For example, the first panel shows the distribution of states for highly extraverted individuals (“extraverts”, +1 SD above the mean on the questionnaire) and for highly introverted individuals (“introverts”, -1 SD above the mean on the questionnaire). Thus, these are fairly highly extraverted and introverted individuals. It can be seen extraverts and introverts do enact different states, because the distributions do not completely overlap. However, there is also a great deal of overlap in how they act: Extraverts quite regularly act introverted and introverts quite regularly act extraverted. The difference between extraverts and introverts is not that they do different things, nor in the frequency of being in the tails of the distributions, but in the frequencies with which they enact mid-range extraverted and introverted behaviors. Extraverts are in moderately-extraverted states about 5-10% more often than are introverts; introverts are in moderately introverted states about 5-10% more often than are extraverts.

The other four traits have less overlap than does extraversion. For example, there is less overlap in intellect states between those high in questionnaire intellect and those low in questionnaire intellect. Again, the difference is mainly not in the extremes, but in the mid-range states. It's not that those high in intellect are always having deep thoughts, it's that they take a moderately creative approach somewhat more frequently than do those low in intellect. All traits demonstrate this same pattern in which the differences are not in the behaviors but in the frequency with which they enact those behaviors. For example, even those very low in trait agreeableness (1 SD below the mean) are quite friendly and polite most of the time; their standing as low on agreeableness is due to the fact that they are a little less warm a little more often, but not due to them being rude. Thus, these figures show what kind of trait manifestation in behavior can be expected of those low and those high in a trait, and also what distributions of trait manifestation qualify an individual as high or low on a trait.

Advantages and Disadvantages of the State as the Operationalization of Behavior

The state concept is only one way to assess the degree to which behavior actually manifests traits. This way focuses on transferring the content of the trait to the state intact, and on obtaining comprehensive distributions of trait manifestation. However, there are some disadvantages to this operationalization. The first two disadvantages are due to the use of self-reports (Gosling, Craik, John, & Robins, 1998). On the one hand, self-report likely induced shared method variance between traits and states. For example, there are likely individual differences in reference standards for items such as hard-working, and these reference standards likely apply consistently to states and to traits, artificially inflating their correlations. On the

other hand, participants may compare themselves across occasions, artificially inflating the amount of within-person variance in states. Additional self-report weaknesses likely also affected the results, and future research is called for to address these issues.

A third disadvantage is that the state concept prevents identification of the specific, concrete actions that manifest the trait (Church et al., 2008; Pytlik Zillig et al., 2002; Wu & Clark, 2003). For example, Wu and Clark were able to show that physical aggression more strongly represented the trait of aggression than did arguing. The correlations found in the current study may have been lower if participants only reported performance of specific behaviors, perhaps because participants would have differed in their construals of specific behaviors' relevance to traits, or perhaps because specific behaviors would not have captured the same affective, behavioral, and cognitive aspects of traits. In the Wu & Clark study, the correlations were about the same magnitude as those presented here; in the Church et al. study, they were slightly lower.

Fourth, it is not clear that all adjectives are equally applicable at state and trait levels, because they may inherently refer to patterns across situations (e.g., “consistent”). We deliberately tried to select adjectives that did apply to momentary behavior and to avoid adjectives that did not, but some adjectives indeed might have been more difficult to use. The isomorphism between traits and states is an important topic for future research (Fleeson et al., 2002).

There are reasons to believe these limitations may not have unduly influenced the results. First, experience-sampling and questionnaire completion are very different tasks for participants, and as Lucas & Fujita (2000) pointed out, their non-shared method variance may have in fact deflated the correlations. In fact, the means from questionnaires and from states were very different from each other for every trait (between 6 and 13 points difference), meaning that participants did respond differently to the ESM items than to the questionnaire items. Third, this method reproduced the expected correlations between traits and single states – around .30 – suggesting that the current method produced accurate results despite these potential problems. Finally, Wu & Clark (2003) found similar results using concrete behavioral categories, although for only three traits. Future research is certainly needed to address these questions with other types of reports. For example, we have ongoing research that has found reliable agreement between self-reports and observer reports of states, suggesting that they are valid indicators of behavior.

We see three complimentary advantages to using the state concept. First, it makes behavior commensurate with traits – a straightforward way to assess how extraverted someone is acting at a given moment is to rate how extraverted they are acting at the moment. States are commensurate with traits both in the content and in the continuous scales. States, like traits, incorporate the meaning of the action into their assessment. Thus, state assessment allows straightforward questions such as “do moderately extraverted individuals act moderately extraverted? How often?”

A second advantage to the state concept is that it avoids having to go through the several intermediate and potentially error-ridden stages of normatively identifying behaviors that exemplify the trait, identifying when a person is performing one of those behaviors, and inferring from the performance of the behavior the degree to which the person is displaying the trait. Most importantly, such a procedure leaves it uncertain to what extent the selection or omission of actions is responsible for the particular correlations observed in a given study or for a given trait. State assessment is comprehensive because it assesses all behaviors in terms of the degree of the presence of the trait content.

A third advantage is that states represent the broad sense of behavior, including affect and cognition, rather than only the narrow sense of physical actions. Pytlik Zillig et al. (2002) showed that affect and cognition are also important parts of the content of traits and that traits

differ in their relative mixture. States are flexible in this regard by transferring whatever mixture exists in a trait to the state.

We believe this research needs to be extended to other operationalizations of behavior. However, the particular operationalization of behavior is not the crux of the matter. The crux of the matter is the degree of correspondence between the abstract, trait-level description of a person and the in-the-moment description of the individual, when the individual is confronted with situational and other forces. The criticism of traits has been about that correspondence, with the situationist position being that in-the-moment descriptions do not correspond to trait descriptions, because in-the-moment forces overwhelm the trait forces. And for this crucial issue, the state operationalization is particularly well-suited, allowing direct assessment of the degree of correspondence.

The relatively strong correlations between the traits and the states suggest that this strategy was successful in discovering how traits are represented in individuals' distributions of behavior. The necessity of identifying the concrete behaviors representing each trait and state may suggest two complimentary research projects. One project identifies the content of states and traits (Church et al., 2008; Pytlik Zillig et al., 2002; Wu & Clark, 2003) and another addresses the correlates, causes, and consequences of states (Fleeson, 2007; McNiel & Fleeson, 2006; Wu & Clark, 2003). In fact, this study when combined with the Wu & Clark (2003) study makes a compelling case that personality has strong predictions of behavior. By using concrete behavior categories, Wu & Clark (2003) and Church et al. (2008) were able to show the predictive power of traits for select sets of behaviors. By using states, and by combining 15 studies, we were able to show the predictive power for a comprehensive assessment of all trait manifestations in behavior. The two methods converge on the conclusion that traits do strongly predict behaviors.

Limitations and Future Research

All the data included in this paper were gathered at least in part by the same research team. This has the potential disadvantage that unknown idiosyncrasies in the research team may account for some of the findings. However, this fact also provided a unique advantage, in that it allowed us to conduct a mega-analysis rather than a meta-analysis (Steinberg et al., 1997). Because we were in possession of all the data, we were able to combine it together (with appropriate transformations) and analyze all 15 studies simultaneously.

A second limitation concerned the reliabilities of the questionnaire measures of trait scores. Moderate reliabilities may have reduced the correlations between trait scores and behavior. They may have been due to the small sample size in the questionnaires (about 35 participants per study) or to the short questionnaires used to assess the Big Five (typically 20 to 30 items). Another contributor to unreliability may have been error variance introduced by some of the adjectives used in state reports. Adjectives such as “dependable” that load highly on Big Five factors in trait ratings may have been difficult for participants to apply to states (although adjectives for state reports were chosen to be applicable to momentary behaviors). Our analysis of studies with high reliability provides some evidence that high reliability increases the correlation between trait scores and behavior.

Similarly, means were correlated with each other across traits in both the questionnaires and the experience sampling. Correlations across means was not overly serious in that discriminative analyses (holding all other traits constant while predicting states) maintained high correlations between trait scores and state means (see Table 4). The correlations may have arisen from a genuine individual difference in overall activity level or may be due to scale use or positivity biases. Future research should investigate only one or two traits at a time so that

the experience sampling reports can also assess many items per trait, employ antonyms to increase reliability and reduce overall correlations, and employ more orthogonal markers (Saucier, 2002). However, most studies in the present research in fact employed negatively keyed items.

Finally, all studies but one (“ESM Adolescence”) took place in an American university setting, and all studies but two (“ESM Adolescence” and “ESM Adulthood”) had college-aged participants. This restriction in populations may have affected the estimates of the amount of typical deviation from one's trait standing, and by extension, estimates of the correlations between traits and trait manifestation in behavior. In fact, Nofle and Fleeson (under review) found some evidence for age differences in amount of variability, although the direction depended on the trait, and all age groups had high degrees of variability.

Conclusions

Whether traits have the ability to predict the way an individual acts in a comprehensive and representative set of behaviors is a basic question in personality psychology. These data provide the first case in which comprehensive, naturally-occurring, and representative distributions of trait manifestation in behavior were predicted by trait standing. The results were affirmative. Traits predicted average levels of everyday trait manifestation in behavior, and did so at a level comfortably beyond the supposed .3 or .4 limit of personality (Funder & Ozer, 1983; Ross & Nisbett, 1991). Although we wanted to include all relevant studies without prejudice to provide a complete result, when we limited the analyses to those studies that met stringent requirements, the results were even stronger, approaching correlations of .60 between questionnaire measures and mean state levels.

In addition to showing the implications of trait levels for frequencies of acting in the manner described by the trait, the results also provide a description of state differences between individuals with different traits -- what it means to be high or low in extraversion, agreeableness, conscientiousness, emotional stability, and intellect in everyday life. Figure 2 provides a depiction that shows that there is both quite a bit of overlap and also reliable and clear differences in the frequency with which individuals with different trait levels act in the way described by the trait. The difference is that people at higher levels of the trait consistently put in a few more moments at moderate levels of the trait.

Acknowledgments

We would like to thank Erik Nofle for comments on an earlier draft. Preparation of this article was supported by National Institute of Mental Health Grant R01 MH70571.

References

- Blumenthal TD. Extraversion, attention, and startle response reactivity. *Personality and Individual Differences* 2001;31:495–503.
- Borkenau P, Mueller B. Inferring act frequencies and traits from behavior observations. *Journal of Personality* 1992;3:553–573.
- Bowers KS. Situationism in psychology: An analysis and a critique. *Psychological Review* 1973;80:307–336. [PubMed: 4585116]
- Buss DM, Craik KH. The act frequency approach to personality. *Psychological Review* 1983;90:105–126.
- Cattell RB, Cattell AKS, Rhymer RM. P-technique demonstrated in determining psycho-physiological source traits in a normal individual. *Psychometrika* 1947;12:267–288. [PubMed: 18921433]

- Church AT, Katigbak MS, Reyes JAS, Salanga MGC, Miramontes LA, Adams NB. Prediction and cross-situational consistency of daily behavior across cultures: Testing trait and cultural psychology perspectives. *Journal of Research in Personality* 2008;42:1199–1215.
- Cohen, J.; Cohen, P.; West, SG.; Aiken, L. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd. Lawrence Erlbaum Associates; Mahwah, NJ: 2003.
- Cooper, HM. *Synthesizing research: a guide for literature reviews*. 3rd. Sage Publications; Thousand Oaks, CA: 1998.
- Costa, PT., Jr; McCrae, RR. *The NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources; 1985.
- Epstein S. The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology* 1979;37:1097–1126.
- Eysenck, HJ.; Eysenck, MW. *Personality and individual differences: A natural science approach*. New York: Plenum; 1985.
- Fleeson W. Towards a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology* 2001;80:1011–1027. [PubMed: 11414368]
- Fleeson W. Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions* 2004;13:83–87.
- Fleeson W. Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality* 2007;75:825–861. [PubMed: 17576360]
- Fleeson W, Malanos A, Achille N. An intra-individual process approach to the relationship between extraversion and positive affect: Is acting extraverted as “good” as being extraverted? *Journal of Personality and Social Psychology* 2002;83:1409–1422. [PubMed: 12500821]
- Fridhandler BM. Conceptual note on state, trait, and the state-trait distinction. *Journal of Personality and Social Psychology* 1986;50:169–174.
- Funder DC. Personality. *Annual Review of Psychology* 2001;52:197–221.
- Funder DC. Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality* 2006;40:21–34.
- Funder DC, Colvin CR. Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology* 1991;60:773–794. [PubMed: 2072255]
- Funder DC, Ozer DJ. Behavior as a function of the situation. *Journal of Personality and Social Psychology* 1983;44:107–112.
- Furr RM. Personality psychology as a truly behavioral science. *European Journal of Personality*. in press.
- Goldberg LR. The development of markers for the Big-Five factor structure. *Psychological Assessment* 1992;4:26–42.
- Gosling SD, John OP, Craik KH, Robins RW. Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology* 1998;74:1337–1349. [PubMed: 9599447]
- Guilford, JP.; Zimmerman, WS. *The manual of the Guilford-Zimmerman Temperament Survey*. Beverly Hills, CA: Sheridan Supply Co.; 1949.
- Heller D, Komar J, Lee WB. The dynamics of personality states, goals, and well-being. *Personality and Social Psychology Bulletin* 2007;33:898–910. [PubMed: 17483393]
- Jensen-Campbell L, Rosseli M, Workman KA. Agreeableness, conscientiousness and effortful control processes. *Journal of Research in Personality* 2002;36:476–489.
- Kenrick DT, Funder DC. Profiting from controversy: Lessons from the person-situation debate. *American Psychologist* 1988;43:23–24. [PubMed: 3279875]
- Korotitsch WJ, Nelson-Gray RO. An overview of self-monitoring research in assessment and treatment. *Psychological Assessment* 1999;11:415–425.
- Lucas RE, Fujita F. Factors influencing the relation between extraversion and pleasant affect. *Journal of Personality and Social Psychology* 2000;79:1039–1056. [PubMed: 11138753]
- McCrae, RR.; Costa, PT, Jr. *Personality in Adulthood: A five-factor theory perspective*. 2nd. New York, NY: Guilford Press; 2003.

- McNiel JM, Fleeson W. The causal effects of extraversion on positive affect and neuroticism on negative affect: Manipulating state extraversion and state neuroticism in an experimental approach. *Journal of Research in Personality* 2006;40:529–550.
- Mehl MR, Gosling SD, Pennebaker JW. Personality in its natural habitat: manifestations and implicit fold theories of personality in daily life. *Journal of Personality and Social Psychology* 2006;90:862–877. [PubMed: 16737378]
- Meyer GJ, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, Eisman EJ, Kubiszyn TW, Reed GM. Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist* 2001;56:128–165. [PubMed: 11279806]
- Mischel, W. *Personality and assessment*. New York: Wiley; 1968.
- Mischel W. Toward an integrative science of the person. *Annual Review of Psychology* 2004;55:1–22.
- Mischel W, Peake PK. Analyzing the construction of consistency in personality. *Nebraska symposium on motivation* 1982: Personality – current theory and research 1983;30:233–262.
- Muchinsky PM. The correction for attenuation. *Educational and Psychological Management* 1996;56:63–75.
- Noftle EE, Fleeson W. Age differences in Big Five behavior averages and variabilities across the adult lifespan: Moving beyond retrospective, global summary accounts of personality. under review.
- Ozer DJ, Benet-Martinez V. Personality and the prediction of consequential outcomes. *Annual Review of Psychology* 2006;57:401–421.
- Paulhus DL, Martin CL. The structure of personality capabilities. *Journal of Personality and Social Psychology* 1987;52:354–365.
- Paunonen SV, Ashton MC. Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology* 2001;81:524–539. [PubMed: 11554651]
- Pytlik Zillig LM, Hemenover SH, Dienstbier RA. What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin* 2002;28:847–858.
- Revelle W. Personality processes. *Annual Review of Psychology* 1995;46:295–328.
- Ross, L.; Nisbett, RE. *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill; 1991. McGraw-Hill series in social psychology
- Saucier G. Orthogonal markers for orthogonal factors: The case of the Big Five. *Journal of Research in Personality* 2002;36:1–31.
- Saucier, G.; Goldberg, LR. The language of personality: Lexical perspectives on the five-factor model. In: Wiggins, JS., editor. *The five-factor model of personality: Theoretical perspectives*. New York: Guilford; 1996. p. 21-50.
- Shoda Y, Mischel W, Wright JC. Intra-individual stability in the organization of and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology* 1994;67:674–687. [PubMed: 7965613]
- Steinberg KK, Smith SJ, Stroup DF, Olkin I, Lee NC, Williamson GD, Thacker SB. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. *American Journal of Epidemiology* 1997;145:917–925. [PubMed: 9149663]
- Swann WB Jr, Seyle C. Personality psychology's comeback and its emerging symbiosis with social psychology. *Personality and Social Psychology Bulletin* 2005;31:155–165. [PubMed: 15619589]
- Wiggins, JS. In defense of traits. In: Hogan, R.; Johnson, JA.; Briggs, SR., editors. *Handbook of personality psychology*. San Diego, CA, US: Academic Press; 1997. p. 95-141.
- Wu KD, Clark LA. Relations between personality traits and self-reports of daily behavior. *Journal of Research in Personality* 2003;37:231–256.
- Wallace J. An abilities conception of personality: Some implications for personality measurement. *American Psychologist* 1966;21:132–138.
- Zimmerman DW, Williams RH. Properties of the spearman correction for attenuation for normal and realistic non-normal distributions. *Applied Psychological Measurement* 1997;21:253–270.

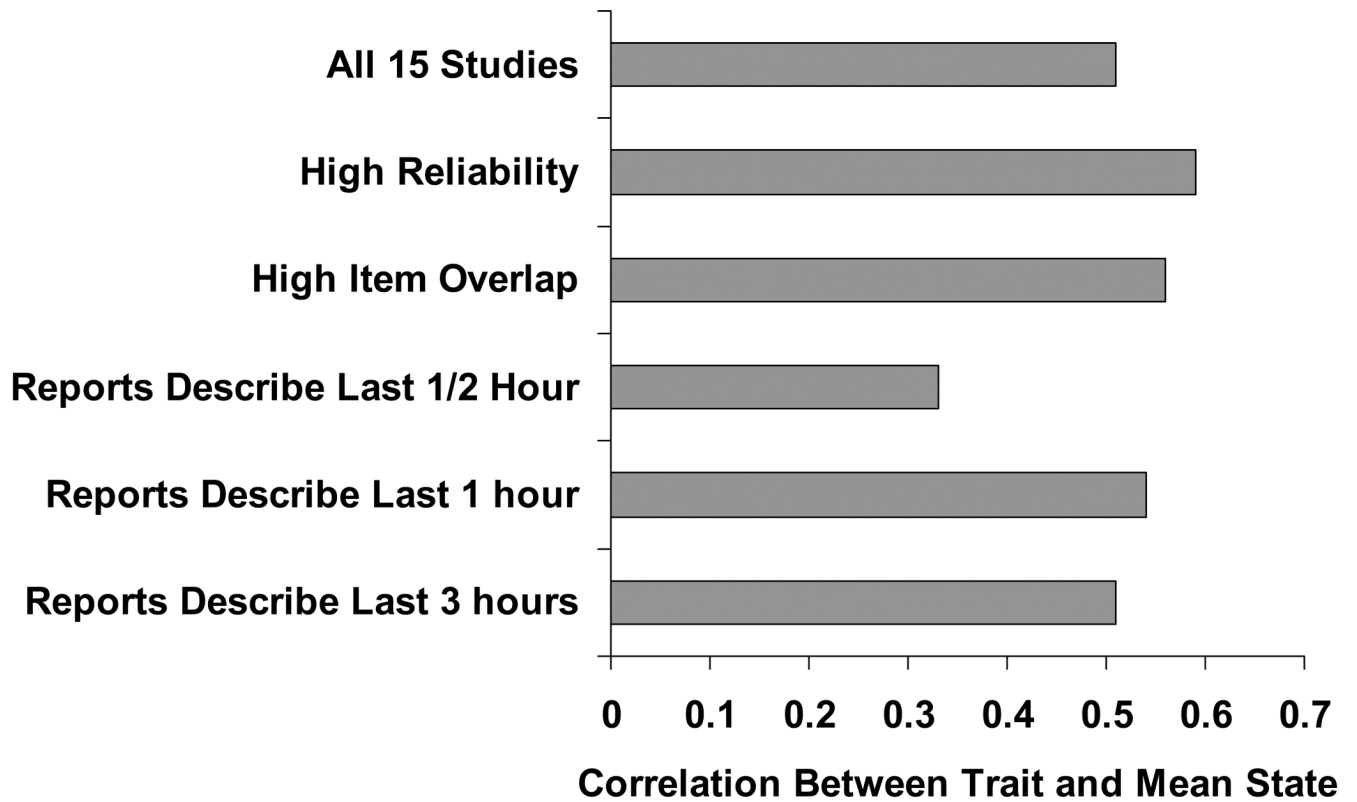


Figure 1. Effects of Study Factors on State Predictability

Bars show correlation between trait questionnaire and corresponding state average, averaged across all traits. Even when all 15 studies were included, the correlations were strong. However, when limited to studies with high reliability or high item overlap, correlations were even stronger, approaching .60.

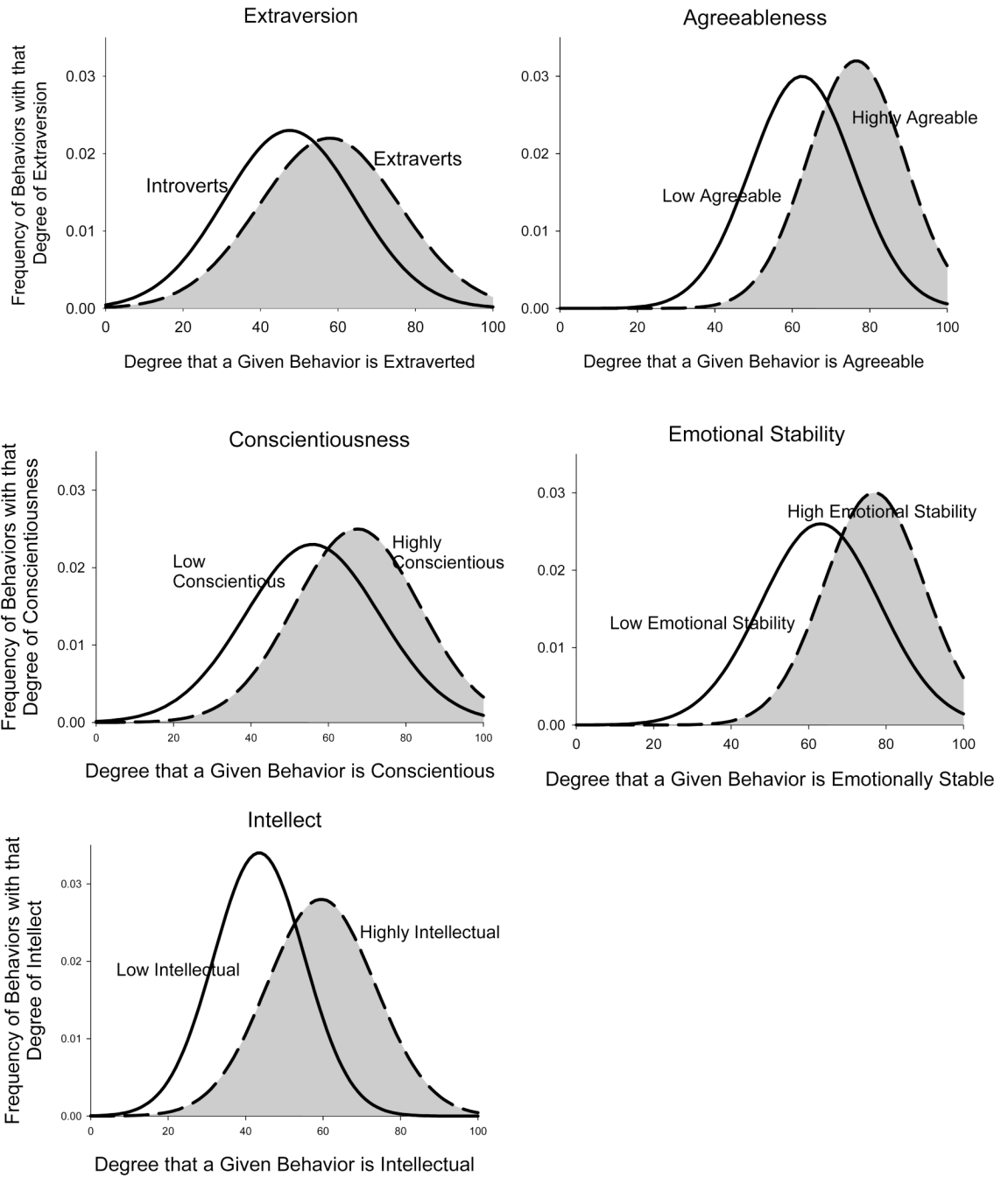


Figure 2. Distributions of actual trait-manifestation in behavior for those high and those low on a trait

Each figure shows the density distribution of states for those one SD above and those one SD below the mean on the trait. As shown, there are reliable differences in mean behavior on the trait, but there are also large overlaps in behavior. The differences between those high and those low on a trait are not so much in non-overlapping ways of acting, nor in the extremes, but rather in the frequencies with which moderately high and moderately low behaviors for the trait are enacted.

Table 1
Five basic questions about the implications of trait standing for acting in the way described by the trait

1	Are there any implications of trait standing for trait manifestation in behavior at all?
2	How strong are the implications for distributions of trait manifestation in behavior?
3	How often and how far do individuals deviate from their trait standing in their behavior?
4	Given at least some variability in individuals' behavior, what summary function across behaviors does the trait standing predict?
5	Beyond the central prediction of a particular summary function, are there additional implications of trait standing for behavior?

Table 2

Methodological details by study

Study	N	Reports per day* # of days	Mean reports per participant in analyses	When Q data were collected (in relation to ESM)	Time described in ESM reports	Reliabilities < .65		Adjectives per trait	
						Q	ESM	Q	ESM
1. ESM Situations	26	4*14	37.3	beginning	last ½ hour	E,C	A	3	3
2. ESM Standard 1	41	5*13	44.8	end	last hour	none	ES	6	4
3. ESM Standard 2	30	5*14	44.2	beginning	last hour	none	none	5	5
4. ESM Goals 1	36	5*13	33.1	beginning	last 3 hours	E,A	A,C	3	3
5. ESM Goals 2	28	7*7	32.6	beginning	last hour	none	none	6	3
6. ESM Sits. Long	47	5*35	95.6	beginning	last ½ hour	C	E,C,ES	6	3
7. ESM Adolescence	21	5*14	30.2	end	last hour	C,I	ES	4	4
8. ESM Opposites	12	5*21	45.0	2 months prior	last 3 hours	none	C	6	6
9. Weekly	63	1 per week	9.8	beginning	last week	none	none	5	5
10. ESM Extraversion	21	5*12	29.5	beginning	last 20 minutes	I	none	5	5
11. ESM Triple Length	13	5*up to 54	169.9	4-6 weeks prior	last 3 hours	none	E,ES,I	5	4
12. ESM Adulthood	62	5*10	37.8	various	last hour	none	E,C	12	3
13. ESM Defined	21	5*21	65.7	beginning	last 3 hours	I	I	5	5
14. ESM lab	44	2 per week	19.3	beginning	last 20 minutes	A,C,ES	ES	5	5
15. ESM Standard 3	30	5*21	73.2	beginning	last 3 hours	none	none	5	5

Note. Reports per day and # of days refers to the scheduled number of reports and length of the study. Q = questionnaires, and ESM = Experience sampling reports. The reliabilities column indicates which traits had reliabilities less than .65, if any.

Table 3

Density Distributions for all Studies

Study	Extraversion		Agreeableness		Conscientiousness		E. Stability		Intellect	
	M	SD	M	SD	M	SD	M	SD	M	SD
1.	45.80	22.61	69.12	15.87	58.90	23.13				
2.	50.31	18.17	66.60	12.77	63.00	15.97	72.25	13.53	45.33	14.63
3.	57.05	15.77	69.91	11.90	64.92	16.35	69.60	14.21	52.49	14.75
4.	53.56	17.83	68.32	13.49	50.35	20.72	70.09	16.45	46.90	13.85
5.	50.69	20.40			54.32	25.54	79.64	16.66		
6.	42.62	21.86			65.05	20.29	70.66	15.20		
7.	54.81	17.90	61.04	15.41	62.02	14.43	68.03	14.49	57.60	15.06
8.	56.26	15.69	72.31	9.66	66.97	11.12	71.64	13.22	58.56	10.78
9.	62.83	9.98	72.16	9.06	61.78	9.80	60.79	14.62	45.56	10.24
10.	47.34	20.12								
11.	52.54	14.24	72.94	10.30	69.54	12.30	73.93	12.77	50.34	11.82
12.	46.94	23.54	68.66	19.96	65.79	23.31				
13.	58.03	11.65	67.25	9.82	62.39	10.07	64.19	13.90	57.72	10.09
14.	55.37	16.52	74.71	10.89	58.55	12.21	80.78	11.82	56.72	15.72
15.	57.47	14.11	68.50	10.60	63.78	12.13	63.85	11.46	58.32	10.03
<i>Mega-Analysis</i>										
% within	78%		63%		75%		62%		49%	
% between	22%		37%		25%		38%		51%	

Note. *M* and *SD* refer to the typical individual's mean and standard deviation of the given state. The mega-analysis line shows what the distribution of a typical individual is like, characterized by the five dimensions of the Big 5 (with between-person standard deviations in parentheses). The average person's mean is close to the midpoint for extraversion and intellect, but much closer to the agreeable, conscientious, and emotionally stable end than to the disagreeable, unconscientious, and neurotic ends, respectively. The average person's *SD* is high for each of the traits, especially for extraversion and conscientiousness, meaning that the typical individual routinely and regularly expressed most levels of most states. The % within and % between lines show that the amount of variance in states that is within people is greater than the amount of variance that is between people.

Table 4
Results of the Mega-Analysis: Relationship of Big-Five Trait Standing to Multiple Behavior Summaries

Behavior Summary	Extraversion	Agreeableness	Conscientious	Emot. Stability	Intellect
Single State	.18***	.34***	.24***	.31***	.37***
Mean	.42***	.55***	.48***	.53***	.56***
Median	.41***	.52***	.45***	.51***	.55***
Mode	.28***	.45***	.33***	.42***	.48***
Semi-Partial Mean					
	.35***	.43***	.38***	.46***	.51***
Maximum	.34***	.42***	.38***	.42***	.54***
Minimum	.22*	.33***	.32***	.37***	.27***
Variability	.09*	-.07	-.08	-.23***	.23***
Quadratic Trait predicting					
Variability	.15**	.06	.05	.05	.03
Skew	-.24***	-.25***	-.18***	-.28***	-.20***
Kurtosis	.08	.19***	.13**	.27***	-.02

Note. Table entries are correlations between trait standing and acting the way described by the trait. Each row refers to a different way of summarizing states. Correlations were calculated across up to 15 experience sampling studies combined into one data file (N = 15 studies, 495 participants, and 21,871 occasions). Questionnaire scores, parameters, and single states were centered within study before computing the respective correlations, to eliminate artificial inflation of correlations due to between-study differences. The Mean row shows that traits predict average states stronger than .3 or even .4, ranging from .42 for extraversion to .56 for intellect.

 p < .001

**
 p < .01

*
 p < .05

Table 5
Relationship of Big Five Trait Scores to Single States in Each Study

Study	Extraversion	Agreeableness	Conscient.	Emot. Stability	Intellect
1.	.00	.30	.06		
2.	.22	.37	.33	.40	.48
3.	.17	.25	.25	.26	.46
4.	.25	.15	.21	.41	.19
5.	.15		.09	.07	
6.	.02		-.01	.30	
7.	.26	.58	.43	.52	.47
8.	.28	.24	.46	.29	.27
9.	.49	.53	.47	.46	.58
10.	-.18				
11.	.12	.48	.11	.21	.12
12.	.12	.31	.21		
13.	.46	.63	.60	.48	.36
14.	.32	.28	.23	.26	.19
15.	.20	.42	.37	.36	.44
<i>Meta-Analysis</i>	.21***	.34***	.24***	.32***	.37***
<i>Mega-Analysis</i>	.18***	.34***	.24***	.31***	.37***

Note. Each line shows correlations from one study, except for the last two lines which show the results of the meta-analysis and mega-analysis. Each correlation is the correlation between scores from a standard questionnaire assessment of the trait and a single randomly selected state from each individual. It can be interpreted as the correlation between how individuals report acting in general on a questionnaire and how they act at any given moment. N's varied across studies from 12 to 63 participants and 9.8 to 169.9 reports per participant.

Table 6
Relationship of Big Five Trait Scores to Mean Big Five States in Each Study

Study	Extraversion	Agreeableness	Conscient.	Emot. Stability	Intellect
1.	.20	.52**	.18		
2.	.40**	.62***	.65***	.74***	.66***
3.	.45*	.50**	.33	.41*	.63**
4.	.46**	.13	.55**	.66**	.28
5.	.07		.55**	.07	
6.	.02		.25	.42**	
7.	.75**	.73**	.65**	.76**	.75**
8.	.33	.51	.68*	-.01	.56
9.	.63**	.62*	.59**	.63**	.71**
10.	-.21				
11.	.55	.72**	.20	.33	.32
12.	.29*	.52**	.42**		
13.	.70**	.73**	.73**	.63**	.59**
14.	.59**	.38*	.33*	.41**	.32*
15.	.49**	.60**	.60**	.58**	.62**
<i>Meta-Analysis</i>					
	.38***	.50***	.44***	.49***	.53***
<i>Mega-Analysis</i>					
	.42***	.54***	.48***	.53***	.56***

Note. Each line shows correlations from one study, except for the last two lines which show the results of the meta-analysis and mega-analysis. Each correlation is the correlation between scores from a standard questionnaire assessment of the trait and the means of individuals' density distributions of corresponding states in everyday behavior. It can be interpreted as the correlation between how individuals report acting in general on a questionnaire and how they act on average in a few days of their everyday lives. N's varied across studies from 12 to 63 participants and 9.8 to 169.9 reports per participant.

Table 7
Relationship of Big Five Trait Scores to Maximum State in Each Study

Study	Extraversion	Agreeableness	Conscientiousness	E. Stability	Intellect
1.	-.04	.35	.45*		
2.	.22	.45**	.31*	.66***	.61***
3.	.12	.52**	.23	.38*	.54**
4.	.34*	-.25	.31	.26	.24
5.	.04		.62**		
6.	.16		.04	.20	
7.	.70**	.39	.30	.79**	.59**
8.	.06	.48	-.30	.04	.61*
9.	.47**	.64**	.58**	.61**	.75**
10.	.35				
11.	.09	.62*	.16	-.10	.41
12.	.42**	.28*	.25		
13.	.48*	.56**	.71**	.25	.43
14.	.62**	.38*	.16	.28	.40**
15.	.54**	.56**	.58**	.45*	.64**
<i>Meta-Analysis</i>					
	.33***	.38***	.33***	.39***	.51***
<i>Mega-Analysis</i>					
	.34***	.42***	.38***	.42***	.54***

Note. Each line shows correlations from one study, except for the last two lines which show the results of the meta-analysis and mega-analysis.