Published in final edited form as:
J Exp Soc Psychol. 2014 January ; 50: 94–104. doi:10.1016/j.jesp.2013.08.008.

An approach to test for individual differences in the effects of situations without using moderator variables

Donna D. Whitsett[*] and **Yuichi Shoda**[*]
University of Washington, Department of Psychology, Box 351525, Seattle, WA 98195-1525, USA

Abstract

The effects of situations may vary importantly across people. If the relevant individual difference variables are known, moderation analyses can test for this possibility. But what if the moderators are not measured or are unknown? We demonstrated how a Highly-Repeated Within-Person (HRWP) design can be used to answer this question, by examining the effect of support seekers' expressions of distress separately for each participant. Although on average, participants' willingness to provide social support increased as a function of support seekers' levels of distress, the opposite was true for some participants; their willingness to provide support significantly decreased as support seekers' expressed distress increased. These findings underscore the importance of examining reliable individual differences in the effects of situations, and show that this is possible without first identifying and measuring individual difference variables that moderate the situation effects. Furthermore, the HRWP design prevents stimulus sampling problems and substantially increases statistical power.

Keywords

Within-person; Individual differences; Situation features; Helping; Emotion

Introduction

Experimental social psychology, at its heart, is an investigation of the effects of situations. In order to study the nature of the psychological processes underlying social behavior, most experiments vary situations and examine their effects on people's cognitions, affects, and behaviors. But what if the effects of the situations differ reliably from one person to another? To empirically test for such individual differences, researchers usually measure individual difference variables and examine if they moderate the effects of the situations. But if none of the individual difference variables measured in a study interact significantly with the effects of situations, what would you conclude? Supposing that the statistical power of the study was sufficiently large, one might assume that the effects do not vary significantly from one person to another. However, this conclusion may be unwarranted, because the effects of situations may significantly depend on yet to be discovered, or yet to be measured, individual difference variables.

Failing to detect individual differences in the effects of situations can have real-world consequences. Suppose an intervention is highly detrimental to a minority of individuals, but these negative effects are offset by smaller positive effects for a majority of individuals. As

[*] Corresponding authors. dwhitsett@gmail.com (D.D. Whitsett), yshoda@uw.edu (Y. Shoda)..

a result, when averaged across participants, the outcome in the treatment condition may not be different from, or may even be significantly better than, that in the control condition. It is tempting to conclude that the intervention has no effect, or possibly that it is helpful for "people in general" (see Krosnick, Visser, & Holbrook, 2002). But in the case of the above example, the conclusion is unwarranted, possibly misleading, and potentially harmful.[1] The usual caution against interpreting main effects in the presence of an interaction is just as apt, whether or not these individual difference variables that moderate the situation effect happen to be measured.

The possibility of heterogeneity in within-person treatment response and the desire to understand it further is also evident in the recent call by NIMH to go beyond the traditional "gold standard," group-based randomized clinical trials, to idiographic analyses of individual clients: "Going forward, NIMH clinical research will not only assess overall group differences, but also individualized patterns of intervention response. The goal is a personalized approach to treatment" (NIMH, 2011).

We argue that this within-person focus is fitting not just in clinical research, but in the study of social psychological phenomena as well. For example, research assessing the effects of normative messages on energy use behavior shows that normative messages increase energy usage for some individuals, but decrease energy usage for other individuals (Schultz, Nolan, Cialdini, Goldstein, and Griskevicius (2007); merely examining the effect of normative messages averaged across study participants could have resulted in the conclusion that social norms have no influence on the particular behavior that was studied.

In this article, we demonstrate an approach for assessing the variability of situation effects across individuals by testing for person × situation interactions without requiring that one first identify and measure individual difference variables. The Highly-Repeated Within-Person (HRWP) design builds on a long history of within-person methodologies and experience sampling (e.g., Baron, 2010; Bolger & Laurenceau, 2013; Walls & Schafer, 2006). In the present article, the HRWP design was applied to research in a laboratory setting to show, for each individual separately, the effect of observing others' distress on the willingness of study participants to provide social support to another person. We then discuss how this approach can be applied to investigations of a wide variety of social psychological phenomena, and how to overcome some of the limitations of within-person designs.

The effect of observing the distress of others on willingness to provide support

The receipt of social support has been shown to affect individuals' mental health, especially when this support is "invisible" to the recipient (e.g., Bolger, Zuckerman, & Kessler, 2000), but how can we predict who is more likely to receive support? It seems intuitive that people would be more willing to provide social support to an individual expressing greater distress. However, the effect of individuals' expressions of distress on their likelihood of receiving support varies from study to study.

Consistent with the intuition that people would be more willing to provide social support to an individual expressing greater distress, research has found that people who report concealing (as opposed to expressing) their distress from others receive lower levels of social support (Kahn & Hessling, 2001). On the other hand, expressions of negative emotion can also be aversive to others (Coyne, 1976; Herbert & Dunkel-Schetter, 1992), and may

[1]Of course, this is not an issue when the situation effects account for most of the behavioral variance. However, in most studies, even after statistically significant effects are found, the majority of the variance remains unaccounted for, leaving room for large differences across people in the situation effects of interest.

lead to withdrawal (Davidson, Ekman, Saron, Senulis, & Friesen, 1990), preempting the provision of support. Some studies have also shown that the more negative emotion individuals convey, the more likely listeners are to attribute undesirable personality traits to these individuals (Kornfeld, 1977). Thus, although the overt expression of distress sends an obvious signal that help is needed, those who receive these distress signals do not always respond with supportive behaviors (Silver, Wortman, & Crofton, 1990).

How might one resolve these apparently incompatible predictions? One possibility is that both predictions can be correct, in that they both identify processes that affect support provision, albeit in different directions for different individuals. It stands to reason that both processes are always potentially relevant, but the process that is more dominant may vary from one support provider (or group of support providers) to another.

The present study sought to determine the extent of such variability and demonstrate the use of the HRWP approach (Shoda, 2004; Shoda & LeeTiernan, 2002; Zayas, Whitsett, Lee, Wilson, & Shoda, 2008). Specifically, it tested the following hypothesis: there are reliable and substantial individual differences in the relationship between support seekers' expressions of distress and an observer's willingness to provide social support. We also examined whether these differences can be explained by known individual difference variables identified in the social support literature. Having measured these variables, the question still remains as to whether these individual difference variables account for all of the variations across people in the situation effect. Thus, our second hypothesis was: reliable individual differences in the effects of situations (i.e., support seekers' expressions of distress) remain even after the individual difference variables that have been identified as relevant in past social support research have been taken into account (i.e., controlled for).

Method

A Highly-Repeated Within-Person (HRWP) design

The key to the HRWP approach is that it assesses the effects of situations for each individual separately. Traditional between-person experimental designs, by definition, do not allow one to obtain information about how a particular individual would have responded in the condition to which the person was not assigned (see Fig. 1). Thus, assessment of situation effects at the individual level is not possible with a between-person design; it requires a within-person design, allowing one to compare each individual's reactions across a set of situations that differ in the situation features that characterize them. Furthermore, achieving statistical precision, while avoiding stimulus sampling problems (Wells & Windschitl, 1999),[2] calls for a *highly-repeated design*, requiring that each participant is exposed to multiple situations (i.e., different support seekers in the present study).

In the HRWP design, each individual's responses to multiple situations that differ in situation features of interest are observed. The resultant data are multilevel. At a lower, or "micro" level, is the phenomenon of how within a given person, the dependent variable (e.g., willingness to provide social support) covaries with the situation features (e.g., level of distress expressed by support seekers). At a higher, or "macro" level, is the phenomenon of how this within-person effect varies across people. This structure, with repeated observations within individuals, is exactly what a multilevel model captures. Thus, the data generated from the Highly-Repeated Within-Person design calls for analysis using

[2]The "stimulus sampling problem" refers to the ambiguity in interpreting results based on very few stimuli (e.g., two female and two male confederates to examine the effect of confederate gender), as it cannot be known whether results are due to the specific and unmeasured characteristics of the stimuli (e.g., the appearance of specific confederates), or to measured or manipulated qualities of the stimuli (e.g., confederate gender).

multilevel modeling, which allows one to characterize, for each person separately, the relationship between the situation features and each individual's response.

Fig. 2 shows the steps taken to implement the HRWP approach, as outlined by Zayas et al. (2008). Step 1 involves identifying the behavior of interest, and for this illustration, the behavior of interest was social support provision. Step 2 is to develop or obtain the stimuli that characterize the situations in which the behavior of interest occurs. When using the HRWP approach, the situations may be sampled from natural variations, or they may be created by the experimenter.

In the present study, the situations consisted of 65 stimuli, each of which was a video clip of about 25 seconds long. In each video clip, a different support seeker was shown being interviewed, describing their anticipation of a forthcoming distress-provoking situation[3] in his or her own words. As described later, each participant indicated his or her willingness to provide social support to each of the 65 support seekers.

Characterizing the stimulus set

Step 3 in implementing the HRWP design is to identify situation features likely to influence support provision. For this study, we were particularly interested in characterizing the level of distress apparent in each of the support-seeker videos. While the situation features of interest along which the stimuli are coded may be either continuous or discrete (e.g., gender), the current research used continuous ratings of distress. The features may be identified “bottom-up” (e.g., from open-ended descriptions of situations) or “top down” (e.g., characteristics of situations identified as important in a previous theory) (see Zayas et al., 2008). The present study included both types. For example, the “bottom-up” features we examined included support seekers’ fidgeting, forced smile, and disgust, which were identified in a previous study of situation features that people attend to when deciding whether to provide support (Whitsett, Almvig, & Shoda, 2010). In addition, we also examined emotion regulation strategies and positive/negative and verbal/nonverbal expressions of emotion identified in theories of emotion regulation (Folkman & Lazarus, 1985; Gross & John, 1998; Gross & John, 2003; Nolen-Hoeksema & Davis, 1999).

The situation features upon which the researcher chooses to characterize the stimuli can either be naturally occurring in the situations (e.g., in the way each support seeker expressed distress in his or her own way), or they can be situation characteristics that are intentionally manipulated by the researcher. In that case, there may not be any need for coding of the stimuli; rather, the stimuli that were manipulated to represent the experimental condition may be coded 1, while the control condition is coded 0.

In the present study, for Step 4, a separate group of raters viewed each video and indicated the extent to which each feature applied. In order to make ratings of each feature independent of each other, each rater was responsible for only one feature, and rated each clip on that feature. Video clips were presented in random order. Raters were also randomly assigned to the feature that they were asked to rate, and so the number of raters assigned to each feature varied from 8 to 16 (mode = 10, mean = 7.7). The raters used a 1 (“not at all”) to 7 (“very much”) Likert scale to indicate the presence of their assigned feature in each of the clips. These ratings were next standardized within each rater to control for individual differences in use of the response scale (e.g., using only the middle of the scale, vs. the

[3]All support seekers’ distress was induced by leading them to believe that they were in a taste-testing study in which they would eat what most people in the U.S. find unappetizing and unpleasant. The video clips showed these individuals describing how they felt about the taste test. After describing how they felt on video, they were informed that they did not actually have to taste anything (see Zayas et al., 2008 for more details).

entire range). After removing outliers (those whose ratings did not correlate substantially with any other raters), the within-rater standardized ratings were averaged across all of the raters for that feature to index the degree to which each feature was present in each video clip. The average Cronbach's alpha, assessing inter-rater reliability, was .90.

Factors characterizing support seekers' expressions

Within this set of video clips, many of the situation features were highly correlated with each other (i.e., videos that were rated high on one feature were also rated high on other features). Thus, after averaging all ratings available for each feature for each situation, the resultant situation × features matrix was submitted to a principal components analysis with Varimax rotation to reduce the number of variables characterizing the videos. A scree plot indicated that three orthogonal factors, accounting for 82% of the total variance, represented a reasonable solution. Situation features with the five highest loadings on each of the three Varimax rotated factors are shown in Table 1. We labeled these three factors "Overall Distress," "Emotional Restraint," and "Visible Anxiety." The "Overall Distress" factor included high loadings for features such as verbally expressing negative expectations and showing disgust. "Emotional Restraint" comprised features such as emotional suppression and appearing timid. "Visible Anxiety" included nonverbal cues such as looking down or fidgeting.

Participants (support providers)

Participants were 48 female and 32 male psychology students at the University of Washington who received extra course credit. Ages ranged from 18 to 23 (M = 19.0, SD = 1.1). Fifty-one self-identified as White, twenty-six as Asian, and three reported more than one ethnicity.

Procedure overview

For Step 5 of the HRWP approach, each participant ("support-provider") viewed, in random order, 65 short video clips that depicted support seekers disclosing how they felt about the distress-provoking situation they believed they would encounter shortly. After viewing each video, participants indicated the degree to which they were willing to help each support seeker. "Willing to help" was defined in the following way:

> "By 'willing to help,' we mean that you would be willing to provide some comforting words of support. An example of this type of help is someone saying, 'I know how you feel' to the person."

Ratings ranged from 1 (not at all) to 7 (very much), with a mean of 4.28 (SD = 1.58).[4]

In addition to rating their willingness to provide support to the person shown in each video, participants also completed the Personal Distress and Empathic Concern scales of the Interpersonal Reactivity Index (Davis, 1980, 1983b). Empathic concern involves feelings of warmth and concern for another person (Davis, 1983a, 1983b). Personal distress refers to the tendency to become distressed oneself upon witnessing another person in distress (Davis, 1983a).

[4]See Whitsett (2007) for a discussion of how the pattern of variation in each participant's willingness to provide support, across support seekers, was stable over time when measured one week later.

Results

The effect of observing others' distress

For Step 6, willingness to provide support to each of the 65 support seekers was examined within participants, using multilevel modeling with HLM (Raudenbush & Bryk, 2002). When used in an HRWP design, the level-1 model in HLM is a within-person analysis. Within each person, their willingness to help a particular support seeker was modeled as the sum of the "personal average" for each participant, plus how his or her willingness to provide support varies from the personal average as a function of each support seeker's score on the three support-seeker factors: Overall Distress, Emotional Restraint, and Visible Anxiety (i.e., the situation features).

Fig. 3 shows two scatterplots representing the level-1 (i.e., within-person) relationships, each for a different support provider. Each of the 65 points on the scatterplots represents the individual's willingness to provide support to each of the 65 support seekers. For participant 74, one can see that the more "Overall Distress" displayed by support seekers, the more this person was willing to provide support. In contrast, for participant 40, the more a support seeker displayed Overall Distress, the less this participant was willing to provide support. These relationships are in opposite directions, yet both are highly statistically significant at $p < .001$.

The results of the HLM analyses are summarized in Table 2. The second column shows the slope for predicting participants' willingness to provide support from the three support-seeker factors. All three factors were significant predictors of willingness to provide support, as shown in the third column. The slopes for support seekers' "Overall Distress" and "Visible Anxiety" were positive, and highly statistically significant (with p-values being virtually zero). That is, the more "Overall Distress" and "Visible Anxiety" a support-seeker conveyed, the more participants were willing to provide social support. On the other hand, the slope for "Emotional Restraint" was negative, indicating that participants' willingness to provide support decreased in response to greater emotion suppression on the part of the support seekers.

Individual differences in the effect of observing others' distress

The results summarized thus far for Table 2 are average "main effects," or slopes for "an average person." What about the differences in these slopes across individuals? Fig. 4 shows 80 regression lines predicting participants' willingness to provide support from the support seekers' displays of "Overall Distress." Each line corresponds to one of the 80 support providers. This figure shows a great deal of variation in slopes for the relationship between support seekers' "Overall Distress" and participants' willingness to provide support. Some slopes are positive, indicating that as "Overall Distress" of the support seeker increased, willingness to provide support increased. Some slopes are negative, indicating that as support seekers' "Overall Distress" increased, willingness to provide support decreased.

As recommended by Fleeson (2007), in order to determine whether the slopes were significantly different between individuals, we next conducted statistical significance tests for the standard deviations across the slopes (Raudenbush & Bryk, 2002, p. 78–79). The null hypothesis is that all individuals have identical slopes, except for variations in slope due to random fluctuations associated with observations of each data point, indicating that the effects of situation features are the same for all individuals. A significant result means that we can reject this null hypothesis, in favor of the alternative hypothesis that there are reliable individual differences in slope; that is, there is an interaction between characteristics

of the individual support providers, which may be yet to be discovered or measured, and the effect of the situation features on willingness to provide support.

When used in the HRWP approach, the level 2 model in HLM addresses how the individuals' level-1 coefficients vary from one person to another. Specifically, each person's level-1 slope coefficients are modeled as the average for the entire sample of participants, plus how a particular individual's coefficient varies from the group average. The statistical significance tests discussed above (found in the "random effects" section of HLM outputs) examine whether or not the variation of the participants' level-1 slopes from the group average is greater than is expected by chance. The results in the fourth and fifth columns of Table 2 show that the slopes predicting support providers' willingness to help from the three situation factors varied reliably from one support provider to another, rejecting the null hypothesis at $p < .001$ for the first two factors and at $p < .05$ for the third.

An informal Monte Carlo simulation of the test of significance for the person-to-person variations in intra-individual slope

Because this significance test is still relatively new to social psychology, to further illustrate its meaning, we conducted an informal Monte Carlo simulation. First, analogous to the 80 support providers in the present study, a sample of 80 individuals were created, such that for each individual, there were 65 observations, analogous to the 65 video clips in the present study. Each observation was characterized by a value on two variables, X and Y. Analogous to the level of willingness to help participants indicated for each video clip, variable Y served as the outcome variable of interest; analogous to the level of "Overall Distress" characterizing each of the 65 video clips in the present study, variable X served as the predictor of Y. X was randomly sampled from a standard normal distribution (i.e., a normal distribution with a mean of 0 and SD of 1). Then the 65 values of Y were determined by the following equation:

$$\mathrm{Y}=\mathrm{X}+\mathrm{error},$$

where "error" was sampled from a standard normal distribution independently for each of the 65 observations, for each simulated support-provider. Thus, the slope of the within-person scatterplot of Y against X was 1.0 for all simulated support providers, except for chance associations between X and random error. Data generated this way therefore represented the null hypothesis that all individuals had identical (i.e., 1.0) slopes predicting Y from X, except for some variations from individual to individual due to random error contained in each observation.

Data generated with the simulation process described above were then analyzed using HLM, and the *p* values were recorded for the significance test examining the null hypothesis that all individuals had the same slope, except for random error. This process of data generation and HLM analysis was repeated 100 times. Due to chance, using the alpha level of 0.05, of these 100 simulated runs, the results should be statistically significant (i.e., $p < .05$) roughly 5 times — even though there are no genuine individual differences in slope. Our simulated runs resulted in significant *p*-values exactly 5 of the 100 times. That is, the *p*-values were not significant in 95 out of 100 simulations. While 100 simulated runs is extremely small for a Monte Carlo simulation, the results are just as expected, and support the validity and accuracy of this significance test provided by HLM.

In addition, we also ran simulations in which the data were sampled from a population embodying an alternative hypothesis. This population was created by the following equation for each fictitious participant:

$$Y = X + (0.1 * Z) * X + error,$$

where Z represents this individual's value on an individual difference variable. For each individual, one value of Z was randomly sampled from a normal distribution with a mean of 0 and *SD* of 1. Thus, analogous to the individual's score on the "Personal Distress" scale, different individuals were characterized by generally different values of Z. Each individual's value of Z was assumed to be constant as the individual encountered different situations. For example, for an individual for whom Z was 0.15, the equation creating the 65 observations for this individual was:

$$Y = X + 0.015 * X + error = 1.015 * X + error.$$

For a person for whom Z was −0.05, the equation creating the 65 observations for this individual was:

$$Y = X - 0.005 * X + error = 0.995 * X + error.$$

In other words, the simulation represented sampling from a population in which there were genuine differences from person to person in the "effect" of X on Y, reflected in person-to-person variations in the slope predicting Y from X, in addition to variations due to random error. When the data generated through this process were analyzed with HLM, the result of the significance test for the *SD* of slopes across people was usually significant ($p < .05$ in 87 of 100 simulations).

In short, the informal Monte Carlo simulation described above showed that when the data generation process assumed that there were no genuine person-to-person variations in slope, the type-I error committed by HLM's significance test was in fact low, and in this case, identical to the specified alpha level (of 0.05). In contrast, when the data generation process assumed genuine variations across people in the slope predicting Y from X, HLM usually produced significant results, indicating substantial statistical power.

Interactions with measured variables: The moderating effect of Personal Distress and Empathic Concern

Although the HRWP approach does not require measuring individual difference variables, doing so can help researchers predict the effects of situations as a function of known individual difference variables. Because empathy is related to helping (Batson, Duncan, Ackerman, Buckley, & Birch, 1981), we examined individual differences in two components of empathy (Empathic Concern and Personal Distress). This step is shown as Step 7 in Fig. 2.

As before, the three situation factors were entered simultaneously as predictors at level 1, with willingness to provide support as the outcome. This time, however, the level-2 model predicted each participant's intercept and slope coefficient from the measured individual difference variables, scores on the Empathic Concern and Personal Distress scales. The level-2 predictors, Empathic Concern and Personal Distress, were standardized. The results are shown in Table 3. As shown, individual differences in the Empathic Concern scale did not predict the relationships (i.e., the level-1 slope coefficients) between the three situation factors and willingness to provide support.[5]

On the other hand, the effects of individual differences in the Personal Distress scale highly significantly moderated the relationships between the situation factors and willingness to provide support. Specifically, individuals higher in Personal Distress had a substantially *more positive* slope predicting their willingness to provide support from support seekers' Overall Distress. As illustrated in Fig. 5, a one standard deviation increase in the Personal Distress scale predicted an increase of .28 in the slope for the relationship between willingness to provide support and the Overall Distress situation factor. Conversely, individuals lower in Personal Distress had a substantially *less positive* relationship between support seekers' Overall Distress and willingness to provide support. In fact, for 24 of the 80 participants, their willingness to provide support *decreased*, rather than increased, as the level of Overall Distress expressed by the support seeker increased. The average Personal Distress score of these participants was 2.35, significantly lower ($t = -2.96$; $p < .01$) than that of the other participants ($M = 2.78$).

Recall that in the earlier analysis reported in Table 2, significant individual-to-individual variation in the slope was found (shown in the 4th and 5th columns), predicting that there exist some individual difference characteristics that moderated the relationship between Overall Distress and willingness to provide support. The fact that Personal Distress was found to moderate the effect of Overall Distress on willingness to provide support, as reported in Table 3 and illustrated in Fig. 5, confirms this prediction.

Individual differences in Personal Distress also significantly moderated the relationship between the Emotional Restraint situation factor and willingness to provide support. Participants became generally less willing to help support seekers as they exhibited greater Emotional Restraint (the average slope for this relationship was — .12). However, for those who were high in Personal Distress, this slope was not as negative (e.g., for those who were one standard deviation above the mean in Personal Distress, the slope was only — .08); in contrast, the slope was more negative for those who were low in Personal Distress (e.g., for those who were one standard deviation below the mean in Personal Distress, the slope was — .16).

Remaining individual differences in slope: Interactions with unmeasured individual difference characteristics

Do participants' Personal Distress scores explain all of the variation in slopes for the relationship between support seekers' Overall Distress and participants' willingness to provide support? Results of multilevel modeling that included Personal Distress as a level-2 predictor showed that the residual variation of the slopes remained almost as large, and as highly significant ($p < .001$), as when the variance in slopes attributable to participants' Personal Distress was not accounted for. The standard deviation of the slopes predicting willingness to provide support from Emotional Restraint also remained significant after accounting for Personal Distress ($p < .01$). If Personal Distress was all that mattered, these standard deviations would no longer have been significant after taking Personal Distress into account. The fact that they remained highly statistically significant suggests that there remain other unmeasured (or even unidentified) individual differences that influence the effects of the situation factors (i.e., Overall Distress and Emotional Restraint).

[5]Empathic concern also had no effect on the level-1 intercept (or each support provider's "personal average" level of willingness to provide support), which indicates that the overall willingness to provide support across situations was unrelated to Empathic Concern. However, individual differences in the Personal Distress scale were negatively related to overall willingness to provide support, as reflected in the intercept; for a support seeker with an average level of Overall Distress, for every one standard deviation increase in Personal Distress, a decrease of .15 (out of a 1 to 7 scale) was predicted in their willingness to provide support ($p < .05$). This is consistent with the prediction by Davis (1983a) that people experiencing high levels of Personal Distress are less likely to offer help.

What if the study had not used the HRWP approach? A demonstration of what is gained by the HRWP approach

The results so far have shown that with the HRWP design, it was possible to detect the effects of situation features, such as the support seekers' overall distress, with ease (with p values smaller than .001 even for the situation feature with the smallest effect size, "Visible Anxiety"). In addition, because the situation features of interest are represented by a large number of stimuli (65 to be exact), it is unlikely that the results reflect "stimulus sampling" issues (Wells & Windschitl, 1999), which occur when spurious findings are produced due to stimulus idiosyncrasies. In this section, we illustrate the extent to which the HRWP approach facilitated detection of the effects and overcame stimulus sampling problems. Specifically, we chose "Emotional Restraint" (ER) as the situation characteristic of interest, because among the three situation characteristics we examined, its effect size was medium, not as powerful as "Overall Distress", but stronger than "Visible Anxiety."

We asked: what would have been the result if the effect of ER was investigated by using a traditional between-subjects design? In such a study, two conditions, one representing high ER and the other representing low ER would be created. Analogous to a study in which the effect of the confederate type is manipulated by creating a condition with one type of confederate (e.g., female) and another condition with another type confederate (e.g., male), we created a hypothetical study in which half of the participants were exposed to a support seeker who was one of the highest in ER, and the other half, to a support seeker who was one of the lowest in ER. We simulated such a between-subjects study simply by designating half of the participants as subjects in the "high ER" condition, and the other half, in the "low ER" condition. To manipulate the level of ER, we then selected one support seeker high in ER and examined willingness to provide support to this support seeker among participants in the "high ER" condition, and we selected a support seeker low in ER for the "low ER" condition. Simulating the outcome of such a study amounts to simply looking up in our data set how each of these participants responded to the support seeker they are exposed to in this simulated between-subjects study.

Specifically, out of the 65 support seekers, we selected 16 with the highest ratings of ER, and 16 with the lowest ratings of ER, representing the top quartile and bottom quartile in the extent to which a support seeker was characterized by ER (e.g., appearing timid and not showing emotions). We then randomly chose one of the 16 "high ER" support seekers and considered him or her as the situation operationalizing the "high ER" condition, and one of the 16 "low ER" support seekers as the situation for the "low ER" condition. Finally, we randomly assigned half of our 80 participants to the "high ER" condition, and the other half, to the "low ER" condition, looked up in our data set each participant's response to the support seeker to which they are exposed in this hypothetical between-subjects study, and conducted a *t*-test comparing the two groups. We repeated this 10,000 times, each time randomly selecting a "high ER" and a "low ER" support seeker from the top and bottom quartiles of the ER distribution, respectively, and randomly assigning half of the 80 participants to each condition.

Of the 10,000 simulated between-subjects studies conducted as explained above, 34.3% resulted in a significant t-test, correctly rejecting the null hypothesis at $p < .05$, with the mean differences consistent with the effect that was found with the entire set of data (i.e., shown in Table 2). We will refer to this as the "correct discovery."

Were all the remaining 65.7% of simulated studies a "miss," in that no significant result was found? No, because some of these studies, 10.1% of the 10,000 simulated studies to be exact, resulted in a signifi-cant finding in the direction *opposite* from what we know to be true from the entire data set. Such a "false discovery" is the result of stimulus sampling

errors at work. For example, the “high ER” support seeker sampled in a given simulated study could have also happened to have other, idiosyncratic characteristics that made participants respond in ways more typical of their reactions to “low ER” support seekers, and/ or the “low ER” support seeker sampled happened to be one to whom participants responded in ways similar to their reactions to “high ER” support seekers. In sum, among these simulated between-subjects studies, with one situation each representing high vs. low ER situations, only a modest proportion (36.2%) resulted in a correct discovery, while an alarmingly large proportion (10.1%) of the studies resulted in a false discovery — a statistically significant result opposite from what is representative of the population.

In the above simulated between-subjects studies, only one support seeker was sampled per condition. What if two support seekers were sampled per condition, and participants in that condition was exposed to either of them? This is analogous to a study in which two confederates each were used to represent a condition (e.g., two female confederates for the “female” condition, and two male confederates for the “male” condition). Of 10,000 simulated studies with this design, 28.2% resulted in a “correct discovery,” and 3.9% resulted in a “false discovery.” In other words, sampling two support seekers per condition reduced the false discovery rate substantially, although it also reduced the correct discovery rate somewhat. The latter is because the probability of randomly choosing two support seekers that drew extreme responses (including those in the “correct” direction) from participants is lower, compared to when only one such support seeker was chosen. We ran similar simulated studies for when four support seekers per condition were sampled, 8 support seekers per condition were sampled, as well as 16 support seekers per condition were sampled. The results are shown in the top left panel of Fig. 6, indicating that false discovery rates declined and approached 0% as the number of support seekers sampled per condition increased. The correct discovery rates also declined but remained above 15%.

Representing a somewhat unusual variation of the traditional between-subjects design, we also ran a similar set of simulations in which instead of participants being assigned to one of the support seekers sampled for a condition, they were exposed to all of them, and their responses to them are averaged across all the support seekers to whom they were exposed. The results are shown in the top right panel of Fig. 6. Similar to the first set of simulations, the false discovery rate again approaches 0% as the number of support seekers sampled increases. Unlike the previous set of simulations, however, the rate of correct discovery increased slightly. This is because averaging each participant's responses to multiple support seekers reduces the random noise associated with any given response.

Finally, we ran the same sets of simulations as described above, but this time, assuming that the study was run within-subjects. Unlike between-subjects studies in which a given participant is exposed to support seekers representing only the “high ER” or “low ER” condition, here, they were exposed to both conditions. Thus statistical power is higher because participants can serve as their own “controls,” reducing the error variance. This resulted in generally higher correct discovery rates, while the false discovery rate again approached 0 as the number of situations sampled increases. The lower left panel of Fig. 6 displays the results from simulations in which participants were exposed to only one support seeker per condition. The lower right panel displays the results from simulations in which participants were exposed to all support seekers representing each condition. When 16 situations were sampled, this design is similar to the HRWP design, in that all participants were exposed to many support seekers, including those clearly high in ER or clearly low in ER.

In summary, if the present study had been done using a between-subjects design, with one stimulus per condition, the result would have been a much reduced ability to discover the

phenomenon that was found with the HRWP design, but also a significant number of false discoveries, or significant results in the opposite direction of the results obtained when all the stimuli were used, due to stimulus sampling errors. Increasing the number of stimuli sampled is effective in reducing false discoveries. But the proportion of correct discoveries with between-subjects designs is far lower than what is possible with a within-subjects design. With a within-subjects design in which participants were exposed to all stimuli in each condition, which is highly similar to the HRWP design, the proportion of correct discoveries approached 100%.

Note, however, although such a within-subject design is powerful both with regard to maximizing correct discoveries and minimizing false discoveries, it still lacks the ability of the HRWP paradigm, which includes the use of multilevel modeling, to identify individual differences in the effects of situations. In the case of the present data, as shown in the right columns of Table 2 and Table 3, the HRWP approach allows us to discover that while ER generally reduces people's willingness to provide support, there were significant individual differences in this effect. In fact, the results shown in the "*SD* of slopes" column of Table 2 shows that for those participants who are greater than 1.2 *SD* above average in the effects of ER, the effect is positive (i.e., the opposite of the general trend).

Discussion

When the effect of a situation variable is significantly moderated by an individual difference variable, most researchers generally refrain from interpreting the situation "main effect." This caution should apply whether or not the individual difference variable is identified and available for statistical analysis. But when it is not identified, which we suspect may often be the case, what can one do? The present research demonstrates how a Highly-Repeated Within-Person (HRWP) design can be used to reveal heterogeneity of situation effects across individuals, without having to first identify and measure moderating variables. The HRWP design achieved these goals by determining the situation effect *for each individual support provider separately*.

Specifically, we applied the HRWP design to examine the following question: are people more willing to provide support to those who express more distress compared to those who express less distress? The results indicate that the answer is a resounding "it depends." The HRWP analysis demonstrated that participants reliably differed in the effect of observing others' distress on their willingness to provide support. Even more importantly than statistical significance, the individual-to-individual heterogeneity in the magnitude and direction of the situation effects were substantial; while for many individuals, willingness to help others clearly increased as others' expression of distress increased, for 24 participants (30% of the sample), willingness to provide support *decreased* as support seekers expressed more distress.

Applicability to a variety of social psychological phenomena

The HRWP approach can be applied to a variety of social psychological phenomena, including, but clearly not limited to, the effect of encountering people from a certain group (e.g., characterized by culture, race, gender, disability, sexual orientation, religion), being the target of certain types of behavior by others (e.g., prejudicial remarks, claims of discrimination, challenges to one's identity), and experiencing certain types of stressors (e.g., time pressure, financial hardship, negative performance feedback, interpersonal rejection) that trigger characteristic cognitive, affective, and behavioral dynamics (e.g., the attachment system). For example, the HRWP approach has been used to examine the conditions in which racial microaggressions are hurtful to members of racial minorities (Wang, Leu, & Shoda, 2011), to identify psychological features of potential dating partners that men with a

history of perpetrating psychological abuse are attracted to (Zayas & Shoda, 2007), as well as for identifying the aspects of facial expressions that underlie judgments about a person's sexual orientation (Tabak, 2012). It has also been incorporated into a clinical intervention, in which the features of social situations that trigger maladaptive responses are identified, and targeted for a focused intervention (Shoda, Wilson, Chen, Gilmore, & Smith, 2013).

Overcoming limitations of within-subjects designs

Because the HRWP approach requires observing the same individual in multiple situations, one must consider the possible limitations of any within-subject design. Chief among them are reactivity and carry-over; the effect of a given condition may depend on the condition to which participants were previously exposed. However, there are a variety of strategies to minimize or mitigate these concerns (e.g., "camouflaging" treatments, by using stimuli that vary in salient but unimportant ways so that the specific situation variable(s) of interest is/ are not apparent; see Greenwald, 1976). Furthermore, the effects of situations examined in traditional, between-subject experiments have their own context effect (e.g., Greenwald, 1976, p. 315). Within-subjects designs also take advantage of habituation to the novelty of the experimental setup, after the first few trials. Furthermore, by counter balancing or randomizing the sequence of stimuli, as was the case in the present study, one is focusing on those situation effects that are relatively independent of sequencing. In fact, such a strategy is routinely used in experiments employing a large number of stimuli, such as those examining the effects of priming on reaction time or on event-related potentials (ERP). Finally, many types of social stimuli in fact occur repeatedly in one's life. For example, bank tellers, restaurant servers, and bus drivers encounter a stream of customers in a given day, but the potential effect of customer characteristics (e.g., their race, gender, and age) on their behaviors are no less real and socially meaningful.

A potential logistical limitation of the HRWP design is the time it takes to expose each participant to a large number of stimuli. This requires devising ways in which key information about a situation can be presented and processed quickly, for example in the form of visual stimuli that convey a complex array of social information at a glance. The applicability of this design is further enhanced by overall experimental contexts (e.g., a "cover story") that allow presentation of a succession of stimuli. An example of such a strategy is the present study, in which it was relatively natural for participants to frame the experiment as deciding how willing they would be to provide support to each of the support seekers, using stimuli that can be presented relatively quickly.

The stimuli may be created by researchers to operationalize a certain variable (e.g., photos of individuals manipulated by computer software to express a certain emotion, such as smiling; Tabak, 2012), or they may be samples of situations that occur frequently in the course of individuals' daily lives. As discussed later, the former will overcome the stimulus sampling problem present in many laboratory studies, while the latter will allow researchers to be a step closer to an ultimate goal of understanding people's responses to social situations they encounter in their lives, rather than only in a laboratory.

Benefits of the HRWP approach

The HRWP analysis is not limited to the analysis of the role of unidentified or unmeasured individual difference variables; it also allows researchers to examine the role of known individual difference variables. In the present study, the HRWP analysis discovered that, compared to those who are low in Personal Distress, those high in Personal Distress were *more* responsive to the support seekers' expressions of distress, increasing their willingness to help at a much higher rate in response to increased expressions of distress by support seekers. At the same time, as predicted by Davis (1983a), with regard to the overall levels,

averaged across support seekers, people high in the Personal Distress component were also significantly *less* willing to provide support (see footnote 5). Had we not employed an HRWP design, but approached the role of Personal Distress more traditionally by relating it to people's average behavioral tendencies, we would not have discovered this important juxtaposition that characterizes those high in Personal Distress: lower overall willingness to provide support but greater responsiveness to support seekers' level of distress.

The HRWP approach also facilitates the search for new individual difference variables that predict how the situation of interest affects individuals, ultimately identifying and characterizing groups of people for whom the effects of situations are relatively homogeneous and predictable. Such efforts to group persons according to the direction and magnitude of the effects of situation features can have significant real-world applications. For example, in the present research, one may form three groups of individuals: those who provide more support in response to support-seekers's expressions of distress. This may then be followed by an effort to identify ways in which the three groups differ with regard to their cognitive and affective dynamics.

Finally, it should be noted that the HRWP approach employs not only a within-person design, but also a *highly* repeated one. This is critical, because if we had exposed participants to only one "high distress expression" support seeker and one "low distress expression" support seeker, we would not have been able to rule out the possibility that the effect is due to factors other than support seekers' distress levels. When a participant is exposed to as many as 65 support seekers who varied in the degree to which they expressed distress, however, and when there is a definite trend such that high distress expression predicted greater willingness to provide support, it becomes clear that the relationship observed within a particular participant is not due to chance (as seen in participant 74, Fig. 3, top panel, for whom the relationship between distress expression and willingness to help was $r = .84$, with *n of situations* = 65, $p < .0001$). And when there are other individuals for whom such relationships are in the opposite direction (e.g., participant 40, Fig. 3, bottom panel, for whom $r = -.73$, *n of situations* = 65, $p < .0001$; and many others, shown in declining lines in Fig. 4), it becomes clear that the individual differences in these within-person relationships go well beyond random errors associated with each observation. Furthermore, as shown in the simulation, the HRWP design greatly increases the chance of identifying an effect where one exists.

Testing, rather than assuming, the homogeneity of situation effects across individuals

Rather than *assuming* that a situation effect is widely applicable across individuals, an HRWP design empirically *tests* if it is homogeneous. Confidence in the homogeneity of the main effect across individuals is increased when the HRWP approach shows that the effect of a situation variable of interest does not depend highly on individual difference variables, measured, or unmeasured.[6] If, on the other hand, the results indicate that the situation effect differs greatly and reliably from person to person, the scientific community will gain valuable information, prompting us to interpret the general findings with caution, and more importantly, to launch an investigation to identify and measure some of the relevant individual difference variables. The goal of any science of course is to account for complex phenomena parsimoniously. Systematically documenting meaningful and reliable exceptions from the general trend, we believe, is a step toward ultimately achieving such a goal, just as quantum mechanics was born of researchers documenting reliable exceptions from Newtonian physics.

[6]For reasons discussed by Cohen (1994) and others, the practical significance of the individual differences in slopes should be guided by an examination of data plots such as Fig. 5.

The HRWP approach assesses the extent of the heterogeneity of a situation effect even after taking a known individual difference moderator into account. In the present study, if after taking Personal Distress into account, the standard deviation of the slopes had no longer been significant, then one could have been reasonably certain (assuming that the study had adequate power) that Personal Distress is probably the primary individual difference moderator, and it would not likely be fruitful to seek further individual difference moderators. However, the present study found that the standard deviations in slope remained large and significant even after Personal Distress was taken into account. That is, much of the "if... then..." relations for support provision remain unaccounted for by Personal Distress, indicating that future research efforts to identify other potential moderator variables to explain additional variance in the slopes would be fruitful.

There is good reason to believe that the effects of situations may vary critically from one person to another (e.g., Fleeson & Noftle, 2009; Fournier, Moskowitz, & Zuroff, 2009; Shoda, Mischel, & Wright, 1994; Van Mechelen, 2009). This is particularly notable in research that utilizes experience sampling (e.g., Koval & Kuppens, 2012; Timmermans, Van Mechelen, & Kuppens, 2010). The present findings provide a demonstration of a laboratory paradigm to examine individual-to-individual variations in the effects of situations. Of course, we would be remiss if we do not emphasize that the findings of the present study are specific to the sample of participants and the range of distress situations examined. Nonetheless, the findings suggest the need for a cumulative effort in the scientific community to examine a variety of participant populations, and to identify individual difference moderators that underlie the variations across individuals. In doing so, we believe the present study serves as an example of approaches that can move us closer to a more collective, cumulative science by directly examining the heterogeneity of situation effects across people.

Acknowledgments

This article is based on the first author's dissertation, which was supported in part by NIMH Grant MH39349. We are grateful to members of the Shoda lab at the University of Washington and to Walter Mischel for his generous comments on earlier versions of this manuscript. We also wish to thank the action editor and two anonymous reviewers for their valuable comments and suggestions.

References

Baron J. Looking at individual subjects in research on judgment and decision making (or anything). Acta Psychologica Sinica. 2010; 42:1–11.

Batson CD, Duncan BD, Ackerman P, Buckley T, Birch K. Is empathic emotion a source of altruistic motivation? Journal of Personality and Social Psychology. 1981; 40:290–302.

Bolger, N.; Laurenceau, J. Intensive longitudinal methods: An introduction to diary and experience sampling research. The Guilford Press; New York: 2013.

Bolger N, Zuckerman A, Kessler RC. Invisible support and adjustment to stress. Journal of Personality and Social Psychology. 2000; 79:953–961. [PubMed: 11138764]

Cohen J. The Earth is round (p b .05). American Psychologist. 1994; 49:997–1003.

Coyne JC. Depression and the response of others. Journal of Abnormal Psychology. 1976; 85:186–193. [PubMed: 1254779]

Davidson RJ, Ekman P, Saron CD, Senulis JA, Friesen WV. Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology I. Journal of Personality and Social Psychology. 1990; 58:330–341. [PubMed: 2319445]

Davis MH. A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology. 1980; 10:85.

Davis MH. The effects of dispositional empathy on emotional reactions and helping: A multidimensional approach. Journal of Personality. 1983a; 51(2):167–184.

Davis MH. Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of Personality and Social Psychology. 1983b; 44(1):113–126.

Fleeson, W. Studying personality processes. Explaining change in between-persons longitudinal and within-person multilevel models.. In: Robins, RW.; Fraley, RC.; Krueger, RF., editors. Analyzing and interpreting personality data. Guilford Press; New York, NY: 2007. p. 523-542.

Fleeson W, Noftle EE. In favor of the synthetic resolution to the person–situation debate. Journal of Research in Personality. 2009; 43:150–154.

Folkman S, Lazarus RS. If it changes it must be a process: Study of emotion and coping during three stages of a college examination. Journal of Personality and Social Psychology. 1985; 48:150–170. [PubMed: 2980281]

Fournier MA, Moskowitz DS, Zuroff DC. The interpersonal signature. Journal of Research in Personality. 2009; 43:155–162.

Greenwald AG. Within-subjects designs: To use or not to use? Psychological Bulletin. 1976; 83:314–320.

Gross JJ, John OP. Mapping the domain of emotional expressivity: Multi-method evidence for a hierarchical model. Journal of Personality and Social Psychology. 1998; 74:170–191. [PubMed: 9457781]

Gross JJ, John OP. Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. Journal of Personality and Social Psychology. 2003; 85(2):348–362. [PubMed: 12916575]

Herbert, TB.; Dunkel-Schetter, C. Negative social reactions to victims: An overview of responses and their determinants.. In: Montada, L.; Filipp, S.; Lerner, MJ., editors. Life crises and experiences of loss in adulthood. Lawrence Erlbaum Associates, Inc.; Hillsdale, NJ: 1992. p. 497-518.

Kahn JH, Hessling RM. Measuring the tendency to conceal versus disclose psychological distress. Journal of Social and Clinical Psychology. 2001; 20(1):41–65.

Kornfeld AD. The favorability of person-perception and the R-S scale. Journal of Clinical Psychology. 1977; 33(2):444–448.

Koval P, Kuppens P. Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. Emotion. 2012; 12(2):256–267. [PubMed: 21787072]

Krosnick, JA.; Visser, PS.; Holbrook, AL. Social psychology under the microscope: Do classic experiments replicate when participants are representative of the general public rather than convenience samples of college students?. Paper presented at the annual meeting of the Society of Experimental Social Psychology; Columbus, Ohio. 2002.

National Institutes of Mental Health (NIMH). [October 1, 2011] The National Institute of Mental Health Strategic Plan. 2011. See. http://menanddepression.nimh.nih.gov/research-funding/research-priorities/nimh-strategic-research-priorities.shtml

Nolen-Hoeksema S, Davis CG. “Thanks for sharing that”: Ruminators and their social support networks. Journal of Personality and Social Psychology. 1999; 77(4):801–814. [PubMed: 10531672]

Raudenbush, SW.; Bryk, AS. Hierarchical linear models. Applications and data analysis methods. 2nd ed.. Sage Publications; Thousand Oaks: 2002.

Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. Psychological Science. 2007; 18(5):429–434. [PubMed: 17576283]

Shoda, Y. Individual differences in social psychology: Understanding situations to understand people, understanding people to understand situations.. In: Sansone, C.; Morf, CC.; Panter, AT., editors. The SAGE handbook of methods in social psychology. Sage Publications; Thousand Oaks, CA: 2004. p. 117-141.

Shoda, Y.; LeeTiernan, S. What remains invariant? Finding order within a person's thoughts, feelings, and behaviors across situations.. In: Cervone, D.; Mischel, W., editors. Advances in personality science. Guilford Press; New York: 2002. p. 241-270.

Shoda Y, Mischel W, Wright JC. Intraindividual stability in the organization 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]

Shoda, Y.; Wilson, NL.; Chen, A.; Gilmore, A.; Smith, RE. Cognitive-Affective Processing System analysis of intra-individual dynamics in collaborative therapeutic assessment: Translating basic theory and research into clinical applications.. Journal of Personality. 2013. http://onlinelibrary.wiley.com/doi/10.1111/jopy.12015/pdf (Advance online publication)

Silver, RC.; Wortman, CB.; Crofton, C. The role of coping in support provision: The self-presentational dilemma of victims of life crises.. In: Sarason, IG.; Sarason, BR.; Pierce, GR., editors. Social support: An interactional view. Wiley; New York: 1990. p. 397-426.

Tabak, J. Doctoral dissertation. University of Washington; Seattle, WA: 2012. You are what you emote: Emotional facial expressions impact sexual orientation judgments..

Timmermans T, Van Mechelen I, Kuppens P. The relationship between individual differences in intraindividual variability in core affect and interpersonal behaviour. European Journal of Personality. 2010; 24:623–638.

Van Mechelen I. A royal road to understanding the mechanisms underlying person-in-context behavior. Journal of Research in Personality. 2009; 43:179–186.

Walls, TA.; Schafer, JL., editors. Multilevel models for intensive longitudinal data. Oxford University Press, Inc.; New York: 2006.

Wang J, Leu J, Shoda Y. When the seemingly innocuous “stings”: Racial microaggressions and their emotional consequences. Personality and Social Psychology Bulletin. 2011; 37:1666–1678. [PubMed: 21885859]

Wells GL, Windschitl PD. Stimulus sampling and social psychological experimentation. Personality and Social Psychology Bulletin. 1999; 25(9):1115–1125.

Whitsett, DD. Doctoral dissertation. University of Washington; Seattle, WA: 2007. Identifying the distress cues social-support providers use when making support-relevant judgments: A highly-repeated within-subjects approach..

Whitsett DD, Almvig T, Shoda Y. Identifying the distress cues that influence support provision: A paired comparison approach. The Journal of Social Psychology. 2010; 150(5):503–519. [PubMed: 21058577]

Zayas V, Shoda Y. Predicting preferences for dating partners from past experiences of psychological abuse: Identifying the “psychological ingredients” of situations. Personality and Social Psychology Bulletin. 2007; 33:123–138. [PubMed: 17178935]

Zayas, V.; Whitsett, DD.; Lee, JJY.; Wilson, N.; Shoda, Y. From situation assessment to personality: Building a social-cognitive model of a person.. In: Boyle, G.; Matthews, G.; Saklofske, D., editors. Handbook of personality theory and testing. SAGE; Newbury Park: 2008. p. 377-401.

HIGHLIGHTS

- Revealed person × situation interactions without identifying moderator variables
- For some, support provision increased as support seekers expressed more distress.
- For others, support provision decreased as support seekers expressed more distress.
- The HRWP design greatly increases the chance of identifying an effect where one exists.

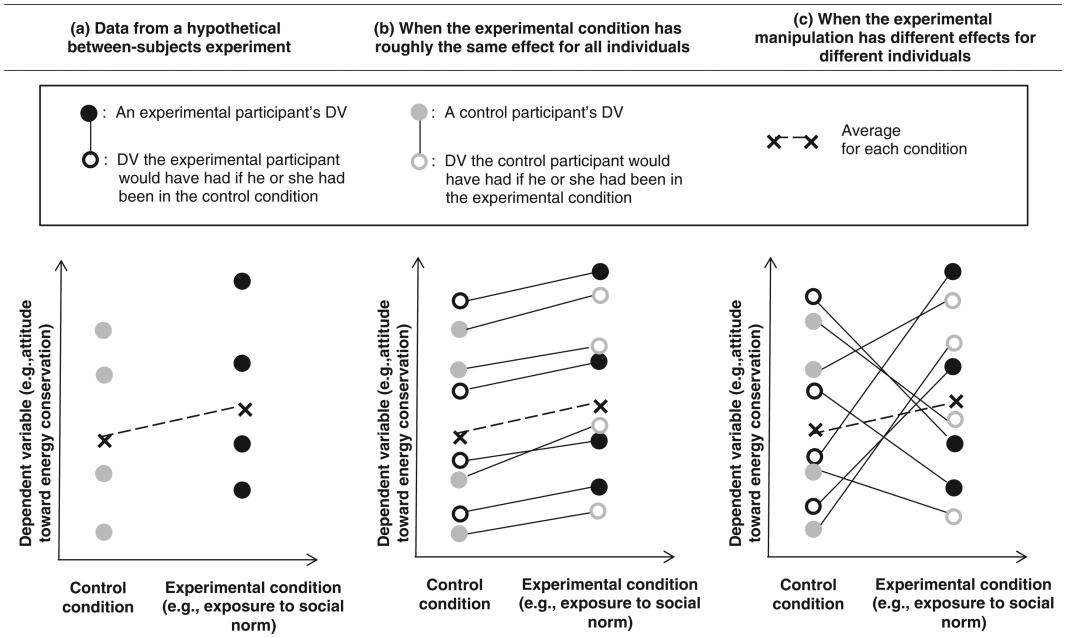

Fig. 1.

The limitations of a between-person design. In a typical between-person experiment, by definition a participant is only in one condition. Thus, it is not known how the participant would have responded to the other experimental condition(s). Panel (a) depicts possible results of a between-subjects experiment. Panels (b) and (c) depict two hypothetical scenarios with regard to what would have been obtained if each participant were observed in the other condition. Panel (b) represents a possibility in which the effect of the manipulation is relatively uniform across all participants. In contrast, Panel (c) represents a possibility in which the manipulation has relatively large positive effects for some individuals while it has relatively large negative effects for the others. Note that the two possibilities are identical in the observed data; an independent groups t-test would yield exactly the same positive effect of the manipulation for the two cases. Thus, an experiment using a between-subjects design is not able to distinguish (c) from (b).

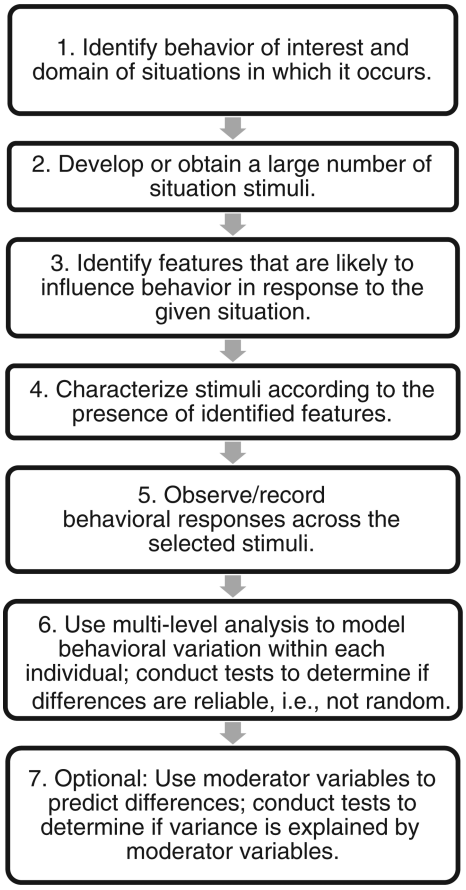

Fig. 2.
Steps involved in the Highly-Repeated Within-Person approach.

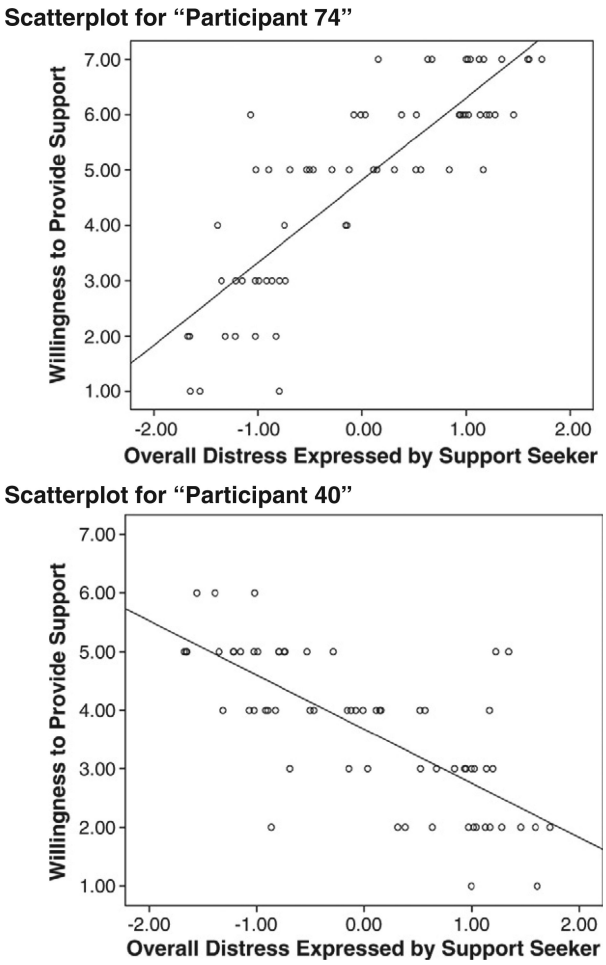

Fig. 3.

Individual scatterplots for the relationship between support seekers' levels of overall distress and support providers' willingness to provide support. Each scatterplot represents a single participant's response to each of the 65 support seekers. For participant 74, $r = .841$ ($p < .0001$). For participant 40, $r = -.731$ ($p < .0001$). Participant 74's Personal Distress score was 3.14 (0.84 *SD* above the mean for the sample of participants). Participant 40's Personal Distress score was −1.86 (1.29 *SD* below the mean for the sample of participants).

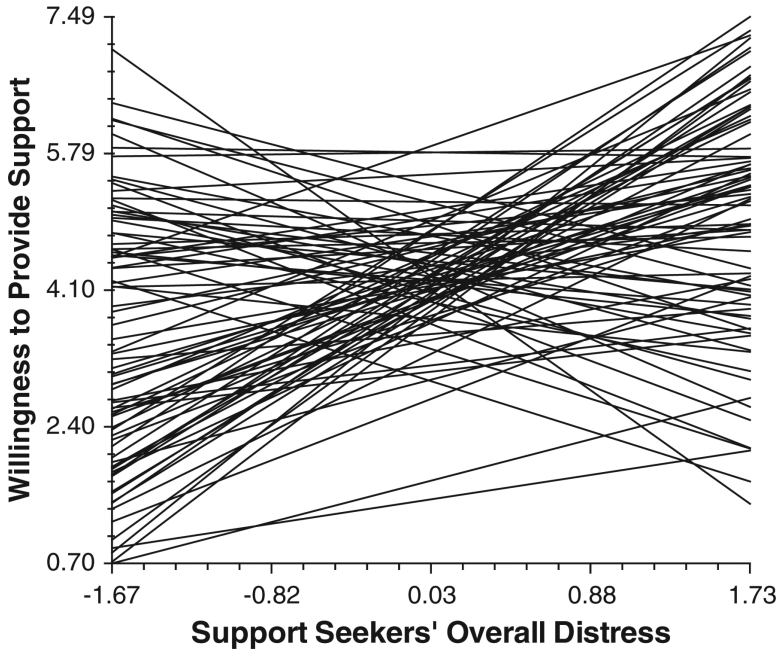

Fig. 4.

Regression lines for 80 participants showing each person's relationship between support seekers' Overall Distress and participants' willingness to provide support. Each line represents a regression line drawn for each participant, based on the 65 observations available per participant. While many of the slopes are positive, some are more positive than others. Furthermore, some slopes are negative, meaning that these participants' willingness to provide support decreased as support seekers' distress increased.

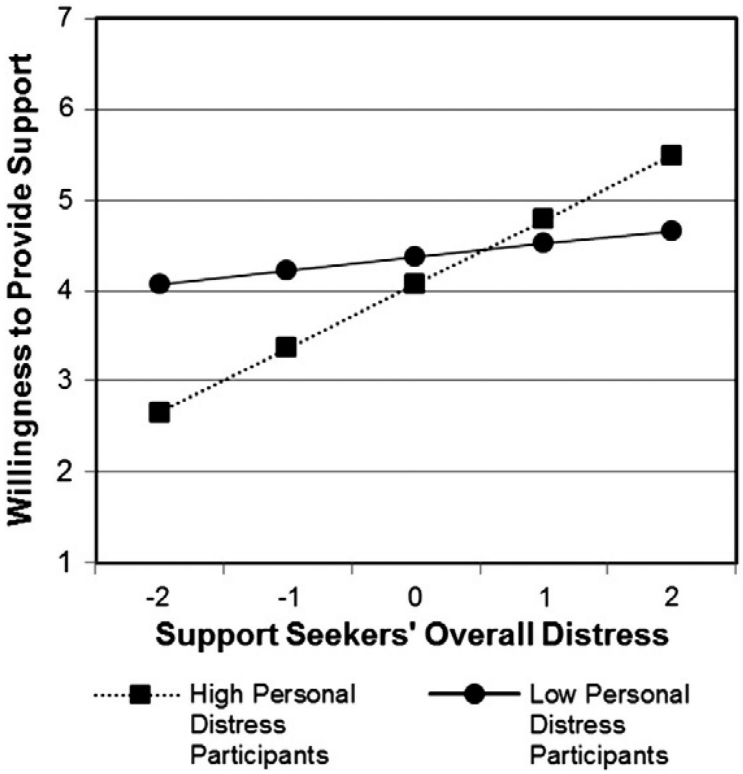

Fig. 5.

Average regression lines for the relationship between support seekers' Overall Distress and willingness to provide support, for support providers high in Personal Distress and those low in Personal Distress. High Personal Distress = 1 *SD* above the sample mean, and low Personal Distress = 1 *SD* below the sample mean.

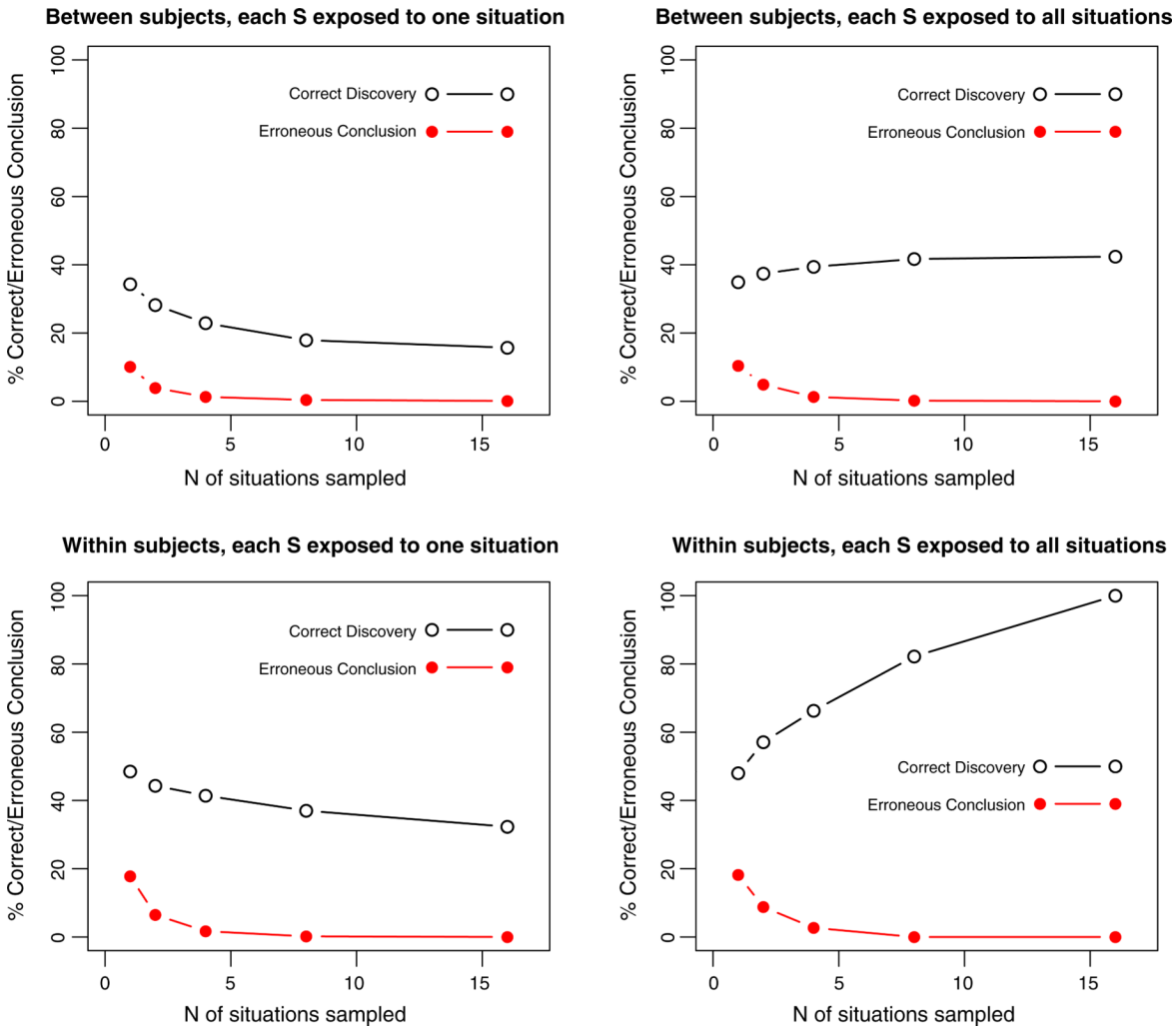

Fig. 6.

The effects of experimental designs on correct discovery and false discovery rates. In each panel, the Y-axis represents the proportions of simulated experiments producing statistically significant results in the correct direction (labeled “Correct Discovery”) and statistically significant results opposite the correct direction (labeled “False Discovery”). The top panels show results of simulated experiments that used a between-subjects design, and the bottom panels show a within-subjects design. The left panels show simulated experiments in which each participant was exposed to only one version of the experimental condition(s) to which they were assigned (referred to as “situation(s)”), and the right panels are for designs in which each participant was exposed to all versions of the conditions to which they were assigned. The X-axis represents the number of versions of each experimental condition included in a study.

Table 1

Features of situations (i.e., support seekers) with five highest loadings for each factor describing the stimulus set.

	Factor 1 "Overall Distress"	Factor 2 "Emotional Restraint"	Factor 3 "Visible Anxiety"
Negative expectations about what is to come next	.980	.021	.120
Disgust	.976	−.016	.067
Positive attitude	−.974	−.071	−.098
Focuses on negative aspects of the task	.974	−.015	.118
Lack of positive (reverse-rated)	−.973	−.074	−.096
Wants to talk to someone about how they are feeling	−.070	−.842	−.084
Suppression	−.198	.828	−.006
Timid	.269	.783	.115
Quiet voice	.211	.777	−.071
Use of humor	−.369	−.631	.211
Looks down	.079	.192	.830
Looks at food	−.082	.137	.760
Fidgety eyes	.268	−.065	.773
Engages in fidgeting/squirming	.215	−.327	.577
Conveys a forced/fake smile	.009	−.463	.560

Table 2

Predicting willingness to provide support from the three situation factors.

Level-1 predictor (situation factors)	Average support provider's level-1 slope coefficient	*p* (H_0: average slope = 0)	*SD* of slopes across support providers	*p* (H_0: *SD* of slopes = 0)
"Overall Distress"	.43	<.001	.79	<.001
"Emotional Restraint"	–.12	<.001	.10	<.001
"Visible Anxiety"	.06	<.001	.08	.017

Note. The level-1 predictors are factor scores. Thus, they are standardized, with a mean of zero and a standard deviation of one.

Table 3

Predicting individuals' level-1 slopes from individual difference variables.

Level-1 predictor (situation factors)	Average support provider's level-1 slope coefficient	Level-2 moderator coefficient (change in level-1 slope predicted for 1 *SD* increase in individual difference variables)		*SD* of slopes across support providers remaining after taking both level-2 variables into account (H_0: *SD* of slopes = 0)
		Empathic Concern (*df* = 78)	Personal Distress (*df* = 78)	
"Overall Distress"	.43***	.00	.28***	.75***
"Emotional Restraint"	–.12***	.00	.04*	.10**
"Visible Anxiety"	.06***	.01	.00	.08*

Note. The level-2 predictors, Empathic Concern and Personal Distress, are standardized, and they were entered into the model separately rather than simultaneously. To compute the *SD* of slopes remaining after accounting for both level-2 variables, a model was run in which both level-2 variables were entered simultaneously.

* $p < .05$.

** $p < .01$.

*** $p < .001$.