Journal of Research in Personality 72 (2018) 58–63

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Replication Study

It takes two: A replication

Brett J. Grant *, Zachary Fetterman, Michelle B. Weyhaupt, Minjung Kim, Alexa M. Tullett

The University of Alabama, United States

ARTICLE INFO

Article history:
Received 28 January 2016
Revised 18 June 2016
Accepted 28 June 2016

Keywords:
Empathy
Empathic accuracy
Replication

ABSTRACT

In their original study, Zaki, Bolger, and Ochsner (2008) suggest that interpersonal factors may explain the lack of correspondence between affective empathy and empathic accuracy in previous work. Specifically, Zaki et al. found evidence that perceivers' affective empathy may only be related to empathic accuracy when the expressivity of the target is high. We attempted a high powered replication of this original study, but did not replicate the original result. In our study, empathic accuracy was not significantly predicted by either perceiver affective empathy or target expressivity, nor was it predicted by their interaction. We discuss differences in measures, sample, and stimuli that may have contributed to discrepancies between our results and those of the original study and theoretical implications.

1. Introduction

As social animals, humans possess a remarkable capacity to understand and experience other people's perspectives. This ability – known as empathy – is associated with pro-social behavior, building close relationships, and maintaining friendships (Batson & Powell, 2003; Eisenberg & Miller, 1987; Maner et al., 2002; McCullough, Worthington, & Rachal, 1997). Empathy often involves attempts to infer others' thoughts and feelings and thus heavily informs people's understanding of those around them. Without the ability to do this accurately, people would largely be at a loss in attempting to make sense of other people and effectively navigate their social environment.

Although numerous working definitions of empathy have been proposed over the years, research suggests that empathy itself is a multidimensional construct comprised of both cognitive and affective aspects (Davis, 1983; Marshall & Maric, 1996; Rogers, Dziobek, Hassenstab, Wolf, & Convit, 2007). Affective empathy refers to the tendency to feel concern and compassion for another's needs. Cognitive empathy, on the other hand, refers to a perceiver's capacity to understand a person's internal states and is often measured as the accuracy with which the perceiver can assess the thoughts and feelings a given target is experiencing (Ickes, Stinson, Bissonnette, & Garcia, 1990). These two kinds of empathy are thought to be distinct but connected, with some models proposing that affective empathy was a phylogenetic precursor of cognitive empathy (Batson, Early, & Salvarani, 1997; De Waal, 2008; Preston & de Waal, 2002).

Despite the presumed association between these two types of empathy, previous research has failed to demonstrate a consistent relationship between trait measures of affective empathy and performance measures of empathic accuracy (Hall, 1979; Ickes et al., 2000; Levenson & Ruef, 1992). Zaki, Bolger, and Ochsner (2008) suggest that these null findings may reflect a failure to take into account the interpersonal nature of empathy. That is, empathy is affected not only by a perceiver's own empathic ability but also by characteristics of the target. In their experiment, Zaki et al. (2008) found evidence that there is a relationship between a perceiver's trait affective empathy and empathic accuracy, but only when the target is high on expressivity. This finding supports the use of an "interactionist if-then approach to predicting interpersonal outcomes (p. 402)."

In this way, Zaki et al.'s (2008) approach deviates from previous perceiver-driven approaches that fail to take into account important characteristics of the target and perceiver-target relationship. Although much previous research has addressed the factors that influence empathic accuracy (Ickes, 1993; Roberts & Strayer, 1996), this work has tended to focus on the traits of the perceiver. Fleeson's (2004) work illustrates the importance of understanding both the person and the situation when examining people's behavior. For empathic accuracy, this suggests that it's important to examine not only the person's level of empathic concern (personality trait), but also the situation he or she is in (i.e. target expressivity).

Indeed, there have been several independent findings across multiple stimulus video sets and paradigms that have found a pos-

* Corresponding author at: University of Alabama, Department of Psychology, Box 870348, Tuscaloosa, AL 35487-0348, United States.

E-mail address: bjgrant@crimson.ua.edu (B.J. Grant).

http://dx.doi.org/10.1016/j.jrp.2016.06.023
0092-6566/

itive relationship between target expressivity and perceiver empathic accuracy. In 1998, Snodgrass and colleagues found that expressivity predicted accuracy more than perceivers' "sensitivity," using an interview paradigm (Snodgrass, Hecht, & Ploutz-Snyder, 1998). In 2011 and 2012 another group in collaboration with Zaki and colleagues demonstrated that target expressivity predicted accuracy for healthy perceivers, but less so for perceivers with schizophrenia (Harvey, Zaki, Lee, Ochsner, & Green, 2013; Lee, Zaki, Harvey, Ochsner, & Green, 2011).

Given the importance of these findings, we decided to conduct a replication of Zaki et al. (2008). This replication is designed to provide an estimate of the relationship between trait measures of affective empathy and performance measures of cognitive empathy (empathic accuracy) via a pre-registered, independently conducted replication of the original study, using similar materials and a common protocol.

If our study replicates the findings reported by Zaki et al. (2008) we should find a significant effect of target expressivity such that empathic accuracy increases as targets' expressivity increases. In addition, we should find an interaction effect between targets' expressivity and perceivers' affective empathy to predict empathic accuracy. More specifically, greater target expressivity should improve the empathic accuracy of perceivers high in affective empathy more than perceivers with low affective empathy.

2. Method

The materials and procedure for a replication of the original Zaki et al. (2008) study were developed in collaboration with the lead author of the original article.

2.1. Target phase

As in Zaki et al. (2008), the study had two phases. First, in the target phase, we created a series of stimulus videos by recording 10 people (targets) as they discussed emotional events in their lives. Each target discussed 4 of the most positive and 4 of the most negative personal life events that they felt comfortable sharing while being video recorded. After discussing these events, targets used 9-point Likert scales to make summary ratings of the overall valence and arousal of the emotion they had experienced while speaking and completed the 10-item Berkeley Expressivity Questionnaire (BEQ; Gross & John, 1997) which assessed emotional expressivity (e.g. "I am an emotionally expressive person"). Finally, the targets viewed their own videos and made continuous ratings of the valence of the emotion they had felt at each moment while speaking using a sliding 9-point Likert scale (1 = *extremely negative*, 9 = *extremely positive*). A subset of stimulus videos were chosen for use in the second phase of the study. One target's videos were excluded for not following video creation instructions, leaving a total of 9 targets in the analyses. Similar to Zaki et al. (2008), of the remaining videos ($n = 72$), 48 were chosen (24 positive, 24 negative) based on comparable means and standard deviations on the summary ratings of overall arousal. We included 9 targets, as compared to the 11 used in the original study, thus we have slightly lower power to detect effects across targets.

2.2. Perceiver phase

2.2.1. Participants

Following Button et al. (2013) a direct replication of the sample size used to find the significant interaction in the Zaki et al. (2008) study ($N = 33$), which achieved nominal statistical significance ($p \sim 0.02$), would be underpowered. Our original G*Power (Faul, Buchner, Erdfelder, & Lang, 2008) analysis indicated a required sample size of 128 participants to achieve at least 80% power to detect a medium effect size ($r = 0.25$; Cohen, 2016). Therefore, we increased the number of observations and perceivers in the current study and should have much greater power to detect effects across participants.[1]

We recruited introductory psychology students from the subject pool at the University of Alabama. Participants were recruited until we reached at least the planned sample size of 128. Including participants who signed up after we reached this goal, we ended up with a final sample of 142 participants (ages 18–23).

2.2.2. Procedure

In the perceiver phase, participants (perceivers) viewed and responded to the videos created during the target phase. Perceivers first completed the 28-item Interpersonal Reactivity Index (IRI; Davis, 1983), a measure of trait empathy. This index is comprised of four separate constructs: empathic concern (e.g., "I often have tender, concerned feelings for people less fortunate than me"), perspective-taking (e.g., "I believe that there are two sides to every question and try to look at them both"), fantasy (e.g., "I really get involved with the feelings of the characters in a novel") and personal distress (e.g., "In emergency situations, I feel apprehensive and ill-at-ease"). Then, each perceiver viewed 20 target clips (randomly selected from the pool of target videos, with the limitation that each perceiver viewed 10 positive and 10 negative clips) and continuously rated how positive or negative the target was feeling using the same scale as the targets. The dependent measure, empathic accuracy, was determined by the correlation between target's ratings of their own feelings and perceiver's ratings of target's feelings.

There are three known differences from the original study. First, we created our own target videos. This had the advantage of allowing us to test the generalizability of the original results (e.g., do the effects extend to different targets, discussing different events?). One disadvantage, however, is that we used 9 targets (versus the 11 used in the original study), and thus have slightly reduced power to detect the effects of target expressivity. Second, we increased the number of perceivers from 33 to 128, substantially increasing our power to detect main effects and interactions involving perceivers' affective empathy and empathic accuracy. Third, we used the IRI instead of the original study's Balanced Emotional Empathy Scale (BEES; Mehrabian & Epstein, 1972) to measure trait affective empathy. Because we used a different measure of affective empathy, discrepancies between our results and those of the original study may reflect this methodological change (a point to which we will return in the Discussion).

We based our predictions on the empathic concern subscale of the IRI, as it is the subscale that shares the greatest amount of conceptual overlap with the BEES (Davis, 1983), As such, based on the results of Zaki et al. (2008) we predicted that perceivers' scores on the empathic concern subscale should interact with target expressivity, with higher levels of empathic concern predicting greater empathic accuracy when target expressivity is high. We made no specific predictions about the perspective taking, fantasy, or personal distress subscales of the IRI, although we report the results for all subscales.

Overall, we included a total of 9 targets, 48 target videos (24 positive, 24 negative – with some videos repeated across conditions) and 142 perceivers in our analyses. We excluded 161 accuracy scores for participants who failed to respond over the length of the video segments. This left a total of 2457 accuracy scores.

[1] We based our original power analysis on perceivers, but we did not run a similar power analysis for targets. Thus, our decisions about power were based on maximizing the power to detect perceiver effects, but not target effects.

3. Analyses

Based on the original study by Zaki et al. (2008), before conducting the main analysis, we first transformed perceivers' and targets' affective ratings using the Cochrane-Orcutt method to adjust the model for serial correlation in the error term (Ostrom, 1990). Then, accuracy scores were *r*-to-*Z* transformed to ensure normality in the data.

To examine whether the targets' expressivity, perceivers' trait empathy or their interaction significantly predicts empathic accuracy, we employed the multilevel modeling given the nature of multilevel structured data (i.e., video observations are nested within targets and perceivers). Specifically, we adopted a cross-classified random effects model (CCREM; Goldstein, 2011; Raudenbush & Bryk, 2002) since the observations are not hierarchically nested but cross classified. CCREM allows us to decompose the random effects (variance and covariances) into within-cluster associated as well as between-cluster associated. In the current study, there were two factors of between-clusters, which were, targets and perceivers. Fig. 1 presents the example of data structure in this study.

Each observation at Level 1 (empathic accuracy) is nested in the combination of these two random effects (target and perceiver), which is at Level 2. The general CCREM model for the current study can be written as:

Level 1:

$$Y_{ijk} = \pi_{0jk} + e_{ijk}$$

where Y_{ijk} represents the empathic accuracy outcome, for participant *i* for target *j* and perceiver *k*. The intercept, π_{0jk}, represents the predicted empathic accuracy score for participants from the specific combination of target *j* and perceiver *k*. The residual e_{ijk}, represents the deviation of a participant's empathic accuracy outcome from the participant's target and perceiver predicted intercept value (assumed normally distributed with a mean zero and variance, σ^2).

Level 2:

$$\pi_{0jk} = \theta_{000} + \sum \gamma_{01j}(\text{Target\_expressivity}) + \sum \gamma_{02k}(\text{Perceiver Trait}) + b_{00j} + c_{00k} + e_{ijk}$$

The overall intercept, θ_{000}, represents the grand mean empathic accuracy outcome. The target residual, b_{00j}, represents the target effect for target *j* (averaged across targets) and is assumed to be normally distributed with a mean of zero and variance τ_{b00} (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Similarly, the perceiver residual, c_{00k}, represents the perceiver effect for perceiver *k* (averaged across perceivers) and is assumed to be normally distributed with a mean of zero and variance τ_{c00}. The random interaction effect, e_{ijk}, represents the residual beyond that predicted by the grand mean, θ_{000}, and the two random effects, b_{00j} and c_{00k}, and is assumed to be normally distributed with a mean of zero and variance τ_{d00}.

Combined model:

$$Y_{ijk} = \theta_{000} + e_{ijk} + b_{00j} + c_{00k}$$

This model demonstrates the portioning of the variability in participant's empathic accuracy outcomes into the component between target, b_{00j}, the component between perceiver, c_{00k}, and the remaining variability between participants within cells, e_{ijk}.

Variance components can be used to describe the proportions of variability at each level, which is called intraclass correlation coefficient (ICC; Raudenbush & Bryk, 2002). The ICC for empathic accuracy outcomes for the same targets, *j*, for different perceivers, *k* and *k*′ (where $k \neq k'$), can be estimated using:

$$PY_{ijk}, Y_{i'jk'} = \tau_{b00}/(\tau_{b00} + \tau_{c00} + \sigma^2).$$

The corresponding correlation for empathic accuracy outcomes for the same perceivers, *k*, for different targets, *j* and *j*′, would be:

$$PY_{ijk}, Y_{i'j'k} = \tau_{c00}/(\tau_{b00} + \tau_{c00} + \sigma^2).$$

The proportion of the total variability that is within cells can be found by estimating:

$$PY_{ijk}, Y_{i'jk} = \sigma^2/(\tau_{b00} + \tau_{c00} + \sigma^2).$$

The total variance can be found by summing together the denominator in the three equations ($\tau_{b00} + \tau_{c00} + \sigma^2$).

First, we analyzed an unconditional model, which has no predictors in the model, to assess the (ICC). Next, we examined the effect of the target's expressivity (BEQ) and perceiver's trait affective empathy (IRI subscales) sequentially. Finally, we calculated the interaction between target's expressivity (BEQ) and perceiver's trait affective empathy (IRI subscales). Analyses for a series of CCREM were performed using HLM 7 Student Version (Scientific Software International, 2005–2016).

4. Results

Overall, perceivers were quite accurate at assessing the affect of the targets (the mean raw correlation between targets' and perceivers' ratings was $r = 0.68$, $SD = 0.33$). Accuracy differed significantly based on the valance of the video such that perceivers were less accurate for negative videos ($r = 0.65$, $SD = 0.36$) compared to positive videos ($r = 0.71$, $SD = 0.30$), $t(38) = -4.16$, $p < 0.001$. Accuracy scores varied from −0.98 to 1. Generally, expressivity had a significant relationship to both the intensity of affect reported by the targets ($r = 0.043$, $p = 0.033$) and their arousal ratings ($r = 0.084$, $p < 0.001$). This suggests that people who experience emotions more strongly may also see themselves as more expressive. Descriptive statistics for key variables are as follows: Target BEQ (M = 4.76, $SD = 0.75$, range = 3.94–5.94), IRI – Fantasy (M = 3.56, $SD = 0.76$), IRI – Perspective Taking (M = 3.56, $SD = 0.63$), IRI – Empathic Concern (M = 3.8, $SD = 0.62$), IRI – Personal Distress (M = 2.55, $SD = 0.68$).[2]

Before conducting the main analysis, we used a series of data transformation procedures to make the accuracy variable. Perceivers' and targets' affective ratings were first transformed using the Cochrane-Orcutt method to adjust the model for serial correlation in the error term (Ostrom, 1990). A new variable, *accuracy*, was created by computing correlations between perceivers' and targets' ratings for each perceiver-video combination. Next, accuracy scores were *r*-to-*Z* transformed to ensure normality in the data. We ran a Kolmogorov-Smirnov Test to assess normality and found that our data was not normal and skewed (D(2457) = 0.168, $p < 0.001$, skewness = −2.478). Given that the accuracy measures were negatively skewed, the inverse and reflect method was adopted to ensure the normality of the data (Tabachnick & Fidell, 1989) and to proceed with the main data analysis ($M = 1.45$, $SD = 0.18$, Range = 1.00–1.84). Transformations and correlations were performed using IBM SPSS Version 20.0 (IBM Corp. Released, 2011). Correlations between the IRI subscales and the transformed accuracy scores can be viewed in Table 1.

4.1. Unconditional model

After ensuring the normality assumption, we further examined the relationship among the targets' expressivity, perceivers' trait empathy, and empathic accuracy using CCREM. First, we analyzed

[2] Gender information was not collected, thus gender effects are not reported.

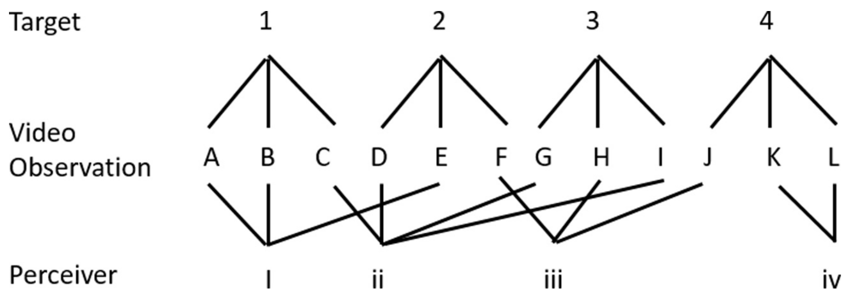

Fig. 1. Network graph of cross-classified data structured with video observations grouped within cross-classifications of target and perceiver.

Table 1
Correlations.

	Empathic accuracy	IRI - PT	IRI - FS	IRI - EC	IRI - PD
Empathic accuracy		−0.042*	0.037	−0.026	−0.001
IRI - PT			0.109**	0.333**	−0.059**
IRI - FS				0.284**	0.105**
IRI - EC					0.284**
IRI - PD					

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

a random intercept-only model in which no predictors were included in the model and intercepts are allowed to vary across two between-clusters. The model can be written as:

$$\text{Level 1}: \quad Y_{ijk} = \pi_{0jk} + e_{ijk},$$

$$\text{Level 2}: \quad \pi_{0jk} = \theta_{000} + b_{00j} + c_{00k},$$

where Y_{ijk} denotes the empathic accuracy measure for i-th observation (i = 1, …, 2457), j-th target (j = 1, …, 9) and k-th perceiver (k = 1,…, 142); π_{0jk} represents the average accuracy score for the; e_{ijk} denotes the residual variance of within-cluster observations; θ_{000} presents the grand-mean of accuracy measure for all targets and perceivers; b_{00j} and c_{00k} indicates the random effects associated with targets and perceivers, respectively.

Table 2 presents the series of CCREM models. The first CCREM column shows the results of the intercept-only model (Uncond.). As shown in the table, the mean of the accuracy measure (θ_{000}) was 1.44 (SE = 0.02). The between-target variance (b_{00j}) was 0.002 (SD = 0.05) and the between-perceiver variance (c_{00k}) was 0.002 (SD = 0.05), which are statistically significant. We calculated the ICC based on the intercept-only model to examine the ratio of between group/subject variance over the total variance. We found the total variance to be 0.035 (0.002 + 0.002 + 0.029). The proportion of the total variability that is within cells was found to be 0.87 (0.029/0.033). The ICC for the target clustering effect and perceiver effect was 0.06 (0.002/0.033) and 0.069 (0.002/0.033), respectively. They indicate that the 6% of the total variance of the accuracy measure is associated with target clustering. Similarly, 7% of the total variance is associated with perceiver clustering.

As in the original study, we next ran analyses to see if perceiver's trait affective empathy or targets' expressivity predicted empathic accuracy given that we found some significant amount of variances are associated with both targets and perceivers.

4.2. Target effects on empathic accuracy

First, we examined the effect of target's expressivity on empathic accuracy by adding the targets' expressivity measure (BEQ) into the model at level 2, where γ_{01} represents the effect of target's expressivity on the empathic accuracy measure. The second column in Table 1 presents the result of this model (Model 1). Targets' expressivity was not significantly associated with empathic accuracy (γ_{01} = −0.002, p = 0.926). Given that no significant effect is observed, all the random effects are not accounted for by adding a predictor.

4.3. Perceiver effects on empathic accuracy

Next we examined the effect of perceiver's trait empathy (IRI) on empathic accuracy by adding the IRI subscales individually at level 2. The individual results for Perspective Taking, Fantasy Seeking, Empathic Concern and Personal Distress are reported in Table 1 (Models 2–5, respectively). In these models, γ_{02} represents the effect of the respective IRI subscale score on the empathic accuracy measure. As shown in Table 1, perceivers' empathic concern (γ_{02} = 0.03, p = 0.989), perspective taking (γ_{02} = 0.038, p = 0.232), fantasy (γ_{02} = 0.024, p = 0.072), and personal distress (γ_{02} = 0.033, p = 0.550) scores were not significantly associated with empathic accuracy. Given that no significant effect is observed, all the random effects are not accounted for by adding a predictor.

4.4. Interaction between perceiver and target effects on accuracy

Our final analyses examined the interaction between perceiver's trait affective empathy (IRI) and target's expressivity on empathic accuracy. The individual interaction results between the IRI subscales (empathic concern, perspective taking, fantasy, and personal distress) and the BEQ are reported in Table 1 (Models 6–9, respectively). In these models, γ_{02} represents the effect of the respective IRI subscale score on the empathic accuracy measure, γ_{01} represents the effect of target's expressivity on the empathic accuracy measure, and γ_{03} represents the interaction term between the respective IRI subscale score and the BEQ. As shown in Table 1, there were no significant interactions between empathic concern (γ_{03} = −0.001, p = 0.564), perspective taking (γ_{03} = −0.0008, p = 0.692), fantasy (γ_{03} = 0.012, p = 0.173), or personal distress (γ_{03} = −0.0008, p = 0.628) and the BEQ. It is also worth noting that in Model 8, we found that empathic concern had a significant effect on empathic accuracy (γ_{03} = 0.045, p = 0.003). This effect, however, may be spurious due to high multicollinearity (i.e., moderate to high correlation between predictors which can lead to a multitude of exacerbated results depending on which predictors are included in the model; BEQ and BEQ ∗ EC interaction, r = 0.689, p < 0.001; EC and BEQ ∗ EC interaction, r = 0.711, p < 0.001; EC and BEQ, r = −0.007, p = 0.72; EC and BEQ).

5. Discussion

Zaki et al. (2008) found that perceiver's trait affective empathy predicted empathic accuracy for expressive targets. In our replication, we did not find evidence that empathic accuracy was pre-

Table 2
Some of the parameters for the models estimated with the two-level cross-classified dataset.

		CCREM									
		Uncond.	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Fixed effects											
Intercept	θ_{000}	1.447**	1.447**	1.447**	1.448**	1.447**	1.448**	1.284**	1.436**	1.298**	1.486**
BEQ	γ_{01}		−0.002					0.003	−0.012	0.0007	−0.004
PT	γ_{02}			0.038				0.044			
BEQ * PT	γ_{03}							−0.0008			
FS	γ_{02}				0.024				0.01		
BEQ * FS	γ_{03}								0.002		
EC	γ_{02}					0.03				0.045*	
BEQ * EC	γ_{03}									−0.001	
PD	γ_{02}						0.033				−0.005
BEQ * PD	γ_{03}										−0.0008
Random effects											
Participant	e_{ijk}	0.029	0.17	0.17	0.029	0.03	0.03	0.17	0.17	0.17	0.17
Target	b_{00j}	0.002**	0.045**	0.002**	0.001**	0.002**	0.002**	0.029**	0.044**	0.024**	0.045**
Perceiver	c_{00k}	0.002**	0.048**	0.002**	0.002**	0.002**	0.002**	0.048**	0.047**	0.048**	0.048**
Model deviance		−1581	−1581	−1582	−1584	−1580	−1580	−1587	−1582	−1591	−1580

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

dicted by target expressivity, perceivers' empathic concern, or their interaction. These results suggest that in our study, target expressivity and perceiver trait empathy were not predictive of empathic accuracy.

The observed null effect of expressivity runs counter to a set of findings in the literature (Harvey et al., 2013; Lee et al., 2011; Snodgrass et al., 1998; Zaki et al., 2008). Therefore, it is important to consider methodological differences between the original study and our replication as these may help to account for discrepancies between the original results and those reported here. One possible concern is differences in statistical power. A limitation for the current study is a relatively small sample size for the target factor. Although we increased the number of perceivers in the current study, we did not increase the number of targets, which would be necessary to increase the power to detect the significance of target's expressivity. Given the relatively small sample size for the targets (i.e., 9), the standard error for the target's expressivity might be overestimated, which leads to a decrease in power. Nevertheless, the fixed effect (regression weight) for the target's expressivity should be robust to the small sample size.

Another important difference is that we used a different measure of affective empathy. The original study used the BEES while our replication used the IRI. Davis (1983) developed the IRI under the assumption that empathy is best measured using a multidimensional approach. Instead of measuring either cognitive or emotional empathy individually, he argues that we may glean more information by distinguishing between different types of reactions to others (Hall, 1979; Ickes et al., 1990; Levenson & Ruef, 1992). In his analyses, Davis (1983) found that the empathic concern and fantasy subscales of the IRI were most strongly related to the BEES, whereas the perspective taking and personal distress subscales were less strongly related. Although our predictions, which were based on the empathic concern subscale, were not supported, we did observe a marginally significant relationship between the fantasy subscale and empathic accuracy ($\gamma_{02} = 0.024$, $p = 0.072$). This may suggest that the original relationship between the BEES and empathic accuracy reflects a tendency to adopt the feelings and perspectives of characters in novels and movies, a result that is consistent with the fact that participants are responding to videos of other people. More broadly, inconsistencies between our results and those reported in the original study may reflect differences between the BEES and the IRI.

Our population of perceivers and targets also differed from those used in the original paper. Our participants were undergraduates in Alabama whereas participants in the Zaki et al. (2008) were undergraduates in New York. It is possible that cultural differences in expressivity or trait affective empathy may contribute to differences in our results. Furthermore, the content and subject matter of the videos created may have also impacted discrepancies. Thus, the present results may suggest limitations to the generalizability of the original results. Perhaps expressivity and trait empathy only contribute to empathic accuracy for certain populations (e.g., only those that have substantial insight into their own trait empathy or expressivity) or for certain types of content (e.g., only content that is highly relatable). Future studies that systematically manipulate these variables may shed more light on these possibilities.

Another possibility is that the effect of target expressivity and perceiver trait affective empathy on empathic accuracy may not be as strong as the original study suggests. Perhaps, in general, people are not very good at self-reporting their expressivity or how skilled they are at understanding the emotional state of others (Bargh & Chartrand, 1999; Wilson & Dunn, 2004; Wilson & Gilbert, 2003). There are two possible reasons that these trait measures (the BEQ, the BEES, and the IRI) might show weak or nonexistent relationships with behavioral measures of empathic accuracy. First, trait empathy measures might assess something different than people's perceptions of their own empathic accuracy. If this were the case, questions that specifically ask, for instance, "How accurate are you at perceiving the emotions of others?" might show a more robust relationship with empathic accuracy. Alternatively, it might be the case that people have poor insight into their own expressivity or empathic accuracy. For example, previous research by Barr and Kleck (1995) found—across multiple experiments—that participants reported greater intensity in their own facial expressions than was actually observed by independent judges. In the current study, one finding that may shed light on this possibility is the observation that, trait expressivity ratings were related to the intensity and arousal ratings reported by the targets. Perhaps people rate their own expressivity by introspecting about the intensity of their emotions, but fail to take into account whether others would be able to perceive that intensity. Taken into context, these results highlight the importance of using empathic accuracy paradigms like that employed by Zaki et al. (2008). As was suggested by our results, self-reports of expressivity and trait empathy may miss important elements of interpersonal interactions that are captured by empathic accuracy measures.

Finally, it is important to note that our replication is not the first replication of the original study. A partial replication of the original Zaki et al. (2008) study found effects similar to the original study. aan het Rot and Hogenelst (2014) attempted to replicate the Zaki et al. finding that perceiver's trait affective empathy and empathic accuracy was moderated by emotional expressivity of the target. They found that higher perceiver affective empathy and more expressive targets both contributed to perceiver's empathic accuracy independently. However, they did not replicate the interaction effect of the original study. This may be due to the fact that, like our current study, they increased power with respect to perceivers (N = 100), but not to targets (N = 11). Between this and the original study, there is data to suggest that affective empathic and target expressivity can impact empathic accuracy.

Although our results are inconsistent with the results of Zaki et al. (2008), we hope other researchers will conduct further replications of this study. Important considerations for future replications include increasing the number of target videos and maintaining consistency across stimuli and measures. A greater understanding of the interpersonal nature of empathic accuracy has the potential to not only shed light on theoretically interesting questions (i.e., What makes someone accurate? Are some people harder to read than others?), but also to inform methodological concerns in empathy research (i.e., Can people accurately self-report their own empathic abilities?). Thus, the Zaki et al. (2008) study continues to provide an important foundation for future research examining the process by which people come to understand the thoughts and feelings of others.

References

aan het Rot, M., & Hogenelst, K. (2014). The influence of affective empathy and autism spectrum traits on empathic accuracy. *PLoS One, 9*(6), e98436.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7), 462.

Barr, C. L., & Kleck, R. E. (1995). Self-other perception of the intensity of facial expressions of emotion: Do we know what we show? *Journal of Personality and Social Psychology, 68*(4), 608.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin, 23*(7), 751–758.

Batson, C. D., & Powell, A. A. (2003). Altruism and prosocial behavior. *Handbook of Psychology, 3*(19), 463–484.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376.

Cohen, J. (2016). A power primer. In A. E. Kazdin & A. E. Kazdin (Eds.), *Methodological issues and strategies in clinical research* (4th ed., pp. 279–284). Washington, DC, US: American Psychological Association. http://dx.doi.org/10.1037/14805-018.

Davis, M. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113–126.

De Waal, F. B. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279–300.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin, 101*(1), 91.

Faul, F., Buchner, A., Erdfelder, E., & Lang, A. G. (2008). *GPOWER version 3.0. 10*. Germany: Universitat Kiel.

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science, 13*(2), 83–87. http://dx.doi.org/10.1111/j.0963-7214.2004.00280.x.

Goldstein, H. (2011). Multilevel statistical models (Vol. 922). John Wiley & Sons.

Gross, J. J., & John, O. P. (1997). Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior. *Journal of Personality and Social Psychology, 72*, 435–448.

Hall, J. A. (1979). Gender, gender roles, and nonverbal communication skills. In R. Rosenthal (Ed.), *Skills in nonverbal communication* (pp. 32–67). Cambridge, MA: Oelgeschlager, Gunn, & Hain.

Harvey, P. O., Zaki, J., Lee, J., Ochsner, K., & Green, M. F. (2013). Neural substrates of empathic accuracy in people with schizophrenia. *Schizophrenia Bulletin, 39*(3), 617–628.

IBM Corp. Released (2011). *IBM SPSS statistics for windows, version 20.0*. Armonk, NY: IBM Corp..

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61*(4), 587–610.

Ickes, W., Buysse, A., Pham, H., Rivers, K., Erickson, J., Hancock, M., et al. (2000). On the difficulty of distinguishing "good" and "poor" perceivers: A social relations analysis of empathic accuracy data. *Personal Relationships, 7*, 219–234.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology, 59*, 730–742.

Lee, J., Zaki, J., Harvey, P. O., Ochsner, K., & Green, M. F. (2011). Schizophrenia patients are impaired in empathic accuracy. *Psychological Medicine, 41*(11), 2297.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*, 234–246.

Maner, J. K., Luce, C. L., Neuberg, S. L., Cialdini, R. B., Brown, S., & Sagarin, B. J. (2002). The effects of perspective taking on motivations for helping: Still no evidence for altruism. *Personality and Social Psychology Bulletin, 28*(11), 1601–1610.

Marshall, W. L., & Maric, A. (1996). Cognitive and emotional components of generalized empathy deficits in child molesters. *Journal of Child Sexual Abuse, 5*(2), 101–110.

McCullough, M. E., Worthington, E. L., Jr., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology, 73*(2), 321.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*(4), 525–543.

Ostrom, C. W. (1990). *Time series analysis: Regression techniques* (Vol. 9) : . Sage.

Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1–71.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1) : . Sage.

Roberts, W., & Strayer, J. (1996). Empathy, emotional expressiveness, and prosocial behavior. *Child Development, 67*(2), 449–470.

Rogers, K., Dziobek, I., Hassenstab, J., Wolf, O. T., & Convit, A. (2007). Who cares? Revisiting empathy in Asperger syndrome. *Journal of Autism and Developmental Disorders, 37*(4), 709–715.

Scientific Software International Inc. (2005–2016). Hierarchical linear and nonlinear modeling student edition for Windows, version 7.01. Skokie, IL.

Snijders, T. B., & Bosker, R. R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publishers.

Snodgrass, S. E., Hecht, M. A., & Ploutz-Snyder, R. (1998). Interpersonal sensitivity: Expressivity or perceptivity? *Journal of Personality and Social Psychology, 74*(1), 238.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. New York: Harper & Row.

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

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology, 35*, 345–411.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two the interpersonal nature of empathic accuracy. *Psychological Science, 19*(4), 399–404.