

CITATION

Kaiser, F. G., Otto, S., & Schuler, J. (2015) Prosocial propensity bias in experimental research on helping behavior: the proposition of a discomforting hypothesis. *Comprehensive Psychology*, 4, 11.

# Prosocial propensity bias in experimental research on helping behavior: the proposition of a discomforting hypothesis<sup>1</sup>

Florian G. Kaiser and Siegmund Otto

*Otto-von-Guericke University, Magdeburg, Germany*

Johannes Schuler

*University of Mannheim, Germany*

## Abstract

When researchers fail to control for confounding factors, the causes of behavior can be more apparent than real, even in experimental research. The current study replicates an experiment by Weinstein, Przybylski, and Ryan (2009) with the goal of demonstrating that their main finding could have resulted from differences in people's prosocial propensity. In their research, they found their hypothesized interaction effect: depending on the extent of immersion, participants presented with images of nature were found to be more prosocial in both their actions and in their declarations. Our sample of 175 adults ( $M$  age = 29.7 yr.,  $SD$  = 11.7; 97 men, 78 women) was approached personally, randomly assigned to viewing either urban or nature images, and instructed to immerse themselves in the respective images. Using two formally distinct measures of participants' prosocial propensity (i.e., before and after the intervention), the hypothesis that individual differences in people's prosocial propensity can bias conclusions about the origins of prosocial behavior in experimental research was supported. To avoid invalid conclusions, the prosocial propensity levels of research participants should be controlled for.

According to the current literature on prosocial behavior, there are apparently many ways to increase people's generosity, charitable giving, and helpfulness toward others (e.g., Greitemeyer, 2009; Leiberg, Klimecki, & Singer, 2011; Pavey, Greitemeyer, & Sparks, 2011). Conditions as varied as presenting songs with prosocial lyrics, engaging participants in short-term compassion training, or highlighting a sense of relatedness to others all significantly promote people's helpfulness and enhance individuals' prosocial propensity. Hence, it appears to be surprisingly easy to evoke people's inclination to behave prosocially, i.e., to seek equality in outcomes (i.e., disadvantages and advantages; e.g., Van Lange, Schippers, & Balliet, 2011) or to cooperate with others for the common good (e.g., Kaiser & Byrka, 2011).

From research on what is called the good-subject effect, we know that some study participants more or less intentionally try to help experimenters confirm their hypotheses (e.g., Orne, 1962; Nichols & Maner, 2008). Still others, persons with pronounced prosocial tendencies, have been found to comply more often with requests to participate in psychological research (e.g., McClintock & Allison, 1989; Kaiser & Byrka, 2011; Van Lange, *et al.*, 2011). As study participants typically have to comply with tasks in experiments beyond appearing on site, we presume that people with a more pronounced inclination to help others (including experimenters) will comply with experimental tasks more diligently in general. However, this differential tendency to comply with experimental commissions, as instigated by differences in people's generic prosocial inclination, has the potential to create "spurious relationships," particularly in helping research. Such relationships come into existence when variables are correlated with both independent and dependent variables simultaneously (e.g., Shannon, 2004). If not recognized as such, spurious relationships created by confounding factors can resemble an ostensible but nonexistent causal effect.

<sup>1</sup>Address correspondence to Florian G. Kaiser, Institute of Psychology, Otto-von-Guericke University, P.O. Box 4120, D-39016 Magdeburg, Germany or e-mail (florian.kaiser@ovgu.de).

Researchers commonly attempt to cope with confounding factors by employing experimental designs. Unfortunately, random assignment does not always solve the problem as it cannot completely control for confounding factors. For instance, confounding factors can slip into experimental research when data analysts capitalize on chance findings by exploring all possible statistical effects instead of exclusively testing the anticipated effects (e.g., Andersen, Burnham, Gould, & Cherry, 2001), but this is, by the way, a common practice in psychology (Simmons, Nelson, & Simonsohn, 2011). Moreover and as we argue, confounding factors can also slip into experimental research when the anticipated effect is an interaction, an effect moderated by a variable that allows participants to exert various degrees of diligence when complying with the requested experimental task, e.g., immersing oneself in images.

As participants typically differ in the degree to which they are inclined to help others, participants who tend to act more prosocially will assist experimenters more than less prosocial participants will. Simply by adhering to instructions or by executing assignments more conscientiously, even without recognizing the specific hypotheses and without trying to help experimenters confirm their hypotheses, study participants can unknowingly compromise even experimental research and may bias causal inferences especially in research on conditions that influence helpfulness. In this article, we argue that some of the reported effects on helping could actually be spurious rather than authentic because the critical confounding factor, a person's prosocial propensity, usually was not controlled for. To make our empirical argument, we used a previous experiment as a case study by which to illustrate our point.

*An Example: Weinstein, Przybylski, and Ryan (2009)*

In a series of four studies, Weinstein, et al. (2009; Studies 1–4) found that depending on the extent of their immersion, participants presented with images of natural environments, in contrast to participants presented with images of urban environments, scored higher on “community aspiration,” the part of Kasser and Ryan’s (1993) Aspiration Index that is meant to measure people’s inclination to behave prosocially. Simultaneously and again

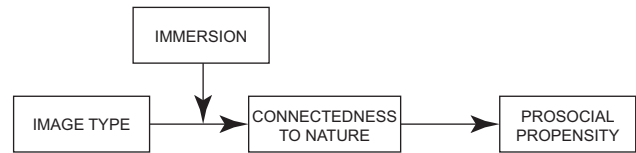


Fig. 1. According to Weinstein, et al. (2009): immersing oneself in images of nature—but not in images of urban environments—increases prosocial propensity.

depending on the extent of immersion, when playing a trust game nature viewers were also found to lend more money to others even though they had nothing to gain in return. Weinstein, et al., furthermore, found that their participants’ scores on the Connectedness to Nature Scale mediated the former two effects of the interaction between immersion and image type (see Fig. 1). On the basis of their findings, Weinstein, et al. concluded that “living in more natural surroundings may conduce to . . . greater caring for others” (p. 1327). However, Weinstein, et al. did not consider the possibility that their findings could have been brought about by a single confounding factor: people’s generic tendency to act prosocially, i.e., their propensity to help others to different degrees.

Arguably, individual differences in their prosocial propensity make study participants differentially inclined to act like good subjects and generally help experimenters (above and beyond helping experimenters confirm their hypotheses). Depending on the extent of their prosocial propensity (the factor that involves helping the experimenter), research participants can be predicted to comply with immersion requests to different degrees. Simultaneously, people who are more prosocial can also be expected to lend comparatively more money to others in a trust game, and they necessarily score higher on people’s self-reported inclination to act prosocially (provided the measure is a valid such measure).

Hence, we anticipate that when participants in psychological experiments on helping differ in their prosocial propensity, they will not only differentially adhere to experimenters’ various requests, but they will necessarily also continue to differ on the prosocial outcome measures (see Fig. 2). Thus, a failure to control for people’s prosocial propensity levels has the potential to compromise the internal validity of experiments because the very confounding factor will lead to spurious relationships and false conclusions about the origins of

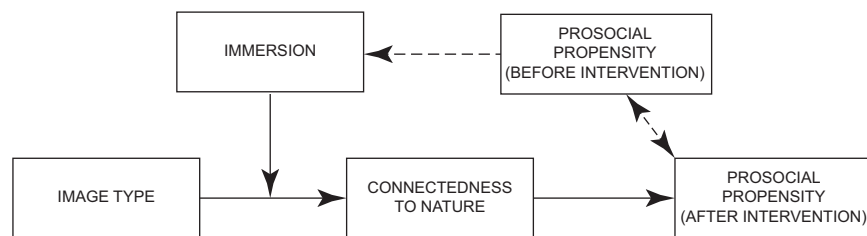


Fig. 2. Differences in individuals’ prosocial propensity levels that exist prior to an intervention continue to exist after an intervention and exert control over differences in compliance with immersion requests.

the differences in people's prosocial propensity and in people's actual helping behavior.

### Research Goals

On the basis of previous research that has demonstrated that individual differences in the propensity to generically act prosocially result in participants complying to different degrees with experimenters' requests (e.g., Kaiser & Byrka, 2011), we replicated Weinstein, et al.'s (2009) findings as depicted in Figure 1. We expected that (a) persons with a pronounced prosocial propensity would comply with instructions to immerse themselves in any kind of images more conscientiously than less prosocial persons. At the same time, this inclination was expected to remain fairly invariant over the course of our study. Thus, we expected that (b) persons who held a pronounced prosocial propensity before the intervention would continue to act prosocially and would continue to respond more affirmatively to prosocial propensity measures after our intervention (see Fig. 2). Furthermore, we expected that (c) Weinstein, et al.'s interaction depicted in Figure 1 (i.e., nature images promoting people's prosocial inclination depending on the extent of immersion) would disappear after individual differences in participants' prosocial propensity were controlled for.

What is critical for our test is not that the confounding factor (i.e., prosocial propensity before the intervention) and the dependent variable (i.e., prosocial propensity after the intervention) are conceptually distinct. Rather, it is critical that the confounding factor and the independent variable are formally and operationally distinct; i.e., that their correlation does not represent a method factor created by, e.g., indiscriminant items and overly similar item wordings. If we have two valid measures of one single concept, people's prosocial inclination—one before and one after the intervention—the two measures should even correspond to a degree that lies in the vicinity of their reliabilities and, thus, the measures should lack discriminant validity (see Campbell & Fiske, 1959).

## Method

### Participants and Recruitment

Our final sample consisted of 175 persons (97 men, 78 women). Their average age was 29.7 yr. ( $SD=11.7$ , range=15 to 72), and 136 (77.7%) of them had a university-level education.

Participants were approached either personally or through an intermediary and were asked to volunteer in an online study. As incentives, we offered compensation that consisted of €5 and feedback on our findings. Participants completed the tasks online in the lab or at a location of their choice (preferably at home). To expand the diversity of participants' prosocial propensity, we purposefully recruited convenience samples of employees from different companies, persons who could from

previous research be expected to show a greater inclination to act ecologically (e.g., vegetarians, organic food store customers, Greenpeace or Green Party members, and environmental activists), and people who could be expected to show a lower such inclination (e.g., motorsports forum members and business administration students; see Kaiser, 1998; Kaiser & Byrka, 2011, 2015; Kaiser, Woelki, & Vllasaliu, 2011). Seven persons had to be excluded prior to our analyses because they either failed to provide all the requested information ( $n=5$ ) or they expressed difficulties with immersing themselves in the presented images of nature ( $n=2$ ), stating that the images appeared unrealistic to them.

To test whether our sampling procedure led to bias in the sample composition with respect to prosocial propensity, we explored the distributional characteristics of our two measures (see below). We found no outliers in either measure (i.e., scores beyond 3 standard deviations at either end of the distribution), and both empirical distributions were also fairly normal according to Kolmogorov-Smirnov tests (both  $ps=.20$ ).

### Measures

In contrast to Weinstein, et al. (2009), we measured prosocial propensity and connectedness to nature more comprehensively, and we measured people's prosocial propensity twice: before and after the experimental intervention. The two assessments were approximately 14 to 15 min. apart: 8.5 min. to watch the images, and 5.5 to 6.5 min. to respond to 41 items (i.e., one asking about the extent of immersion and 40 about connectedness to nature).

*Prosocial Propensity Before the Intervention.*—Prosocial propensity at Time 1 (T1) was measured with the General Ecological Behavior (GEB) scale because previous studies have supplied evidence for the equivalence of these concepts. People with a prosocial inclination tend to report exhibiting more ecological behavior on an array of behavioral self-reports (Hilbig, Zettler, Moshagen, & Heydasch, 2013), and vice versa: people with a generic ecological inclination tend to report exhibiting more prosocial behavior on an array of behavioral self-reports (Kaiser, 1998). Not surprisingly, people with an ecological inclination have been simultaneously found to act more cooperatively with others and to score higher on prosocial value orientation (Kaiser & Byrka, 2011, 2015). The convergence was so close that Kaiser and Byrka (2011) concluded that ecologically engaged persons are also prosocially engaged. Accordingly, we measured the propensity to act prosocially at T1 (before the intervention) with the T1 version of the GEB, a well-established ecological behavior measure (see Kaiser, 1998; Kaiser & Wilson, 2004).

A person's GEB level was assessed with a maximum likelihood approach, which is the typical approach for determining Rasch-scale measures for persons (for more

**TABLE 1**  
Descriptive Statistics and Information on the Convergent Validity of Five Different Measures of Individual Prosocial Propensity at Time 2, After the Intervention

	N	M	SD	PP	GEB	SVO	CA	OD
Second-order prosocial propensity, PP	175	0.00	1.15	0.65 <sup>C</sup>	1.00 <sup>a</sup>	0.56	0.86	0.78
Ecological behavior, GEB	175	0.52	1.13	0.96*	0.79 <sup>R</sup>	0.45	0.56	0.52
Social value orientation, SVO	175	31.64	10.06	0.43*	0.38*	0.92 <sup>T</sup>	0.17	0.15
Community aspiration, CA	175	3.20	0.50	0.58*	0.42*	0.14	0.70 <sup>C</sup>	0.45
Overt donation, OD	175	2.71	2.45	0.63*	0.46*	0.11	0.38*	

**Note** In the section to the right of the SDs, the numbers in the diagonal cells indicate reliability estimates that are either internal consistencies (i.e., <sup>C</sup>Cronbach's  $\alpha$ ), <sup>R</sup>Rasch-model-based reliabilities, or <sup>T</sup>test-retest reliabilities. Second-order prosocial propensity represents the principal factor of four measures (i.e., GEB, SVO, CA, and OD; see Table 2). Ecological behavior (GEB) is from T2, after the intervention. Off-diagonal figures represent Pearson correlations that are either uncorrected (below the diagonal) or corrected (above the diagonal) for measurement error attenuation. A generic correction for measurement error attenuation weighs correlations against the reliabilities of the two measures involved. Widely accepted significance tests are available only for uncorrected correlations.

<sup>a</sup>Corrections resulting in implausible correlations (i.e., Heywood cases) were truncated at the highest plausible value. For overt donation, there was no reliability estimate available; thus, a correction for measurement error attenuation could be conducted only in part.

\* $p < .001$ .

details, see Embretson & Reise, 2000). People's specific GEB levels are numerically depicted in logits. Logits stand for the natural logarithm of the ratio between the probability of an affirmative and the probability of a negative answer for items that represent a specific scale. The higher a logit value, the more pronounced the particular person's level is on a scale.

*Prosocial Propensity After the Intervention.*—Prosocial propensity at Time 2 (T2), by contrast, was measured expansively as a second-order factor that was derived from four different indicators that tap different kinds of verbal and overt behavior with prosocial characteristics (see Table 1): (a) the T2 version of the GEB scale; (b) actual, overt donation (OD); (c) the Social Value Orientation (SVO) measure; and (d) the Community Aspiration (CA) Index. To enlarge its scope, to reduce measurement error from the prosocial performance measure at T2, and to prevent artificial variance shrinkage, we estimated factor scores for the second-order prosocial propensity factor by using the Bartlett method (Bartlett,

1937; also see Thompson, 2004). The results of the factor analysis are presented in Table 2.

*The General Ecological Behavior Scale.*—We used the GEB to assess *ecological behavior* twice—before and after the intervention. To do so, we employed 35 self-report items such as “I collect and recycle used paper” on each instrument. There were a total of 64 behavior items because six were used on both versions. Of these self-reports, 15 were newly created for this study (see the Appendix) and 49 were taken from Kaiser and Wilson (2004). Of the new behavior items, seven were in the T1 version and eight in the T2 version of the measure. Of the old behavior items, six were in both, 22 were exclusively in the T1 version, and 21 were exclusively in the T2 version of the measure. For 24 behaviors, engagement was coded with a yes/no format, and for 40 behaviors it was coded with a 5-point frequency scale with anchors 1: Never and 5: Always. In the two versions of the measure, 13 behaviors were probed with a dichotomous response format and 22 with a polytomous one. The responses to the polytomous behavior items were recoded into a dichotomous format by collapsing *Never*, *Seldom*, and *Occasionally* into *Unreliable ecological engagement*. *Often* and *Always* were combined into *Reliable ecological engagement*. This particular recoding of the polytomous self-report items into a dichotomous format has been shown to diminish measurement error as opposed to diminishing substantive information relevant for the valid assessment of inter-individual performance differences (see Kaiser & Wilson, 2000). *Not applicable* was a response alternative that could be chosen when an answer was not possible for any reason. Of all possible answers (i.e., two times 175 participants times 35 items), 9.9% were found to be either *Not applicable* or missing.

**TABLE 2**  
Principal Axis Factor Analysis With the Four Prosocial Propensity Measures at Time 2, After the Intervention

	Prosocial Propensity	$h^2$
Ecological behavior, GEB	0.85	0.73
Social value orientation, SVO	0.35	0.12
Community aspiration, CA	0.53	0.28
Overt donation, OD	0.56	0.31
Eigenvalue	1.44	
Proportion of variance explained	36.09%	

**Note** Ecological behavior (GEB) is from T2, after the intervention.  $N = 175$ .

Only six behaviors were the same in the two versions to prevent recollection, which could in turn result in the artificial inflation of the correlation. This may seem like a small number of overlapping items, but it nevertheless allowed for a comparable assessment of individual performance differences due to what is called the “scale-freeness” of Rasch-model-based measures (see Michell, 1986). With Rasch scales, a successful calibration requires that person scores, except for measurement error, are independent of the particular selection of items employed as long as these items validly represent the array of possible items (which, in this case, are self-reports of ecological behaviors; see Bond & Fox, 2007). In other words, a complete overlap of behavioral self-reports was not mandatory for the ecological performance measure as long as the selected behaviors formally represented a single class and reflected extensive coverage of the behavioral class in question. We were able to jointly calibrate the two versions with the classical Rasch model and thus ensured that the selected behaviors formally represented a homogeneous class.

Apart from presenting items before and after the intervention, we coded, recoded, and calibrated the GEB in line with standard practices (e.g., Kaiser & Wilson, 2004). All 64 self-reports were found to fit the model to an acceptable degree (more details on fit statistics are available upon request). The Rasch-model-based reliability estimate of the instrument with data from T1 and T2 was  $rel = .79$  ( $N = 350$ ), and its test-retest reliability was  $r_{tt} = .82$ .

*Overt Donation.*—Similar to the method used by Weinstein, et al. (2009), we offered participants the opportunity to overtly act in a prosocial manner. Specifically, they were given the opportunity to donate part or all of the €5 compensation to charity. Donation was defined as the exact amount contributed. Even before they attended the experiment, 31 members of three environmental organizations decided to donate their compensation to their organizations. Although originally unanticipated, these donations were regarded as full donations as they were later confirmed by the organizations.

*Social Value Orientation (SVO).*—We used the basic SVO slider measure consisting of six items (Murphy, Ackermann, & Handgraaf, 2011). The slider measure assesses SVO by asking people to allocate money to a fictitious other, and by doing so, to simultaneously deny themselves this money. In other words, a person's prosocial propensity is measurable as the average proportion allocated to the fictitious other. The test-retest reliability of the SVO measure was reported to be  $r_{tt} = .92$  (see Murphy, et al., 2011). As we measured SVO only once in our study, we were unable to determine a test-retest reliability estimate of our own.

*The Community Aspiration Index.*—The Community Aspiration Index (Kasser & Ryan, 1993) was the standard outcome measure in Weinstein, et al.'s (2009) research. The scale consists of five general prosocial as-

pirations in life such as “contributing to the betterment of society” or “helping people in need.” For each aspiration, participants rated its importance on a scale with anchors 1: Extremely unimportant and 5: Extremely important. Prosocial propensity was assessed by averaging the importance ratings across the five aspirations. Overall, the five verdicts showed an acceptable degree of internal consistency with a Cronbach's  $\alpha$  of .70.

*Connectedness to Nature.*—As the anticipated mediator between the intervention and the outcome (see Fig. 1), a measure of attitude toward nature that is not yet widely used was chosen. This measure has previously been found to be technically superior to more conventional such instruments and to overlap ( $r = .71$  when corrected for measurement error attenuation; see Brügger, Kaiser, & Roczen, 2011) with Mayer and Frantz's (2004) instrument, which was the one used by Weinstein, et al. (2009). Although this measure is conceptually more accurately labeled as a measure of attitude toward nature, we refer to it as a measure of connectedness to nature, as did Weinstein, et al. As in the original study by Brügger, et al., 40 connectedness items were employed with either a dichotomous response format (i.e., 23 items), coded as 0 (No/Disagree) or 1 (Yes/Agree), or with a polytomous response format (i.e., 17 items), with anchors 1: Never and 5: Very often. The latter responses were subsequently recoded from the 5-point to a 3-point format by collapsing *Seldom* and *Occasionally* as well as *Often* and *Very often*; *Never* was retained as *Never*. Again, the recoding of the polytomous items into a trichotomous format was applied to diminish measurement error. Thus far, the recoding has been found to have a positive effect on the valid assessment of inter-individual differences in people's connectedness to nature (see Brügger, et al., 2011). Typical item examples for the connectedness to nature instrument are “I collect mushrooms or berries,” “The croaking of frogs is comforting,” and “If one of my plants dies, I reproach myself.” Once again, *Not applicable* was a response alternative that could be chosen when an answer was not possible. Of all statements, 4.4% were found to be either *Not applicable* or missing. We employed the partial-credit Rasch model, which also was the one that had been used by Brügger, et al. All 40 items were found to fit the model to an acceptable degree (more details on fit statistics are available upon request), and the Rasch-model-based reliability of the scale was  $rel = .88$  ( $N = 175$ ). Again, as with the other Rasch scale, individual levels of connectedness were also assessed with a maximum likelihood approach. And as with the GEB, people's connectedness was also numerically depicted in logits.

### Design and Procedure

Immersion and image type were considered the between-subjects factors. In accordance with Weinstein, et al. (2009), we did not assign participants to different lev-

els of immersion; rather, we randomly assigned participants to view images of either nature ( $n=89$ ) or urban scenes ( $n=86$ ). To corroborate the idea that immersing oneself in images of natural scenery increases one's prosocial inclination, we had to develop our own images. To do so, and analogous to Weinstein, *et al.*, three raters compared 60 images on authenticity, color, complexity, and illumination. The raters consensually selected six comparable pairs (in terms of layout and perspective) of urban and nature images (copies of the images are available as supplementary materials with the online version of this article). For example, a street image with buildings on both sides was matched with a dirt road framed by trees. Moreover, and to facilitate immersion, we presented images with complementary soundscapes: street noises with the urban images, and nature noises (e.g., birds chirping, wind swooshing, and foliage rustling) with the nature images.

First, and in both conditions, prosocial propensity before the intervention was assessed before the six images were presented, each image for 80 sec. Before each presentation, participants were requested to immerse themselves in the images. Like Weinstein, *et al.* (2009), we asked participants to imagine being in the presented scene and to consciously recognize the colors, sounds, and features of each scene.

After the presentation, the extent to which people immersed themselves was assessed in two ways: (a) by measuring the time participants spent looking at the six images and (b) by asking participants how intensely they had been able to immerse themselves in the presented scenes. Responses were given on a 7-point scale with anchors 1: Not at all and 7: Very well. Due to the skewed distribution of self-reported immersion (skewness =  $-1.05$ ) with very few respondents choosing the four lowest response categories, we had to choose between two strategies: collapsing categories or transforming the data. As both strategies have limitations (see MacCallum, Zhang, Preacher, & Rucker, 2002; Tabachnick & Fidell, 2007), we decided to reduce the skewness by calculating the logarithm of the difference between a constant (i.e., maximum immersion score plus 1) minus the specific immersion score (for more details, see Tabachnick & Fidell, 2007). To recapture its original orientation, we additionally multiplied the transformed immersion scores by  $-1$ ; this new variable was no longer skewed (skewness =  $-0.007$ ).

Subsequently, participants completed a measure of connectedness to nature. Last, the four distinct indicators of prosocial propensity after the intervention were assessed. Upon request, participants received written information about the purpose of our research.

## Results

Our findings are reported in two sections. In the first section, we repeated Weinstein, *et al.*'s (2009) core analysis to

determine whether we could corroborate the finding that an increase in people's prosocial propensity was apparently related to the degree to which they immersed themselves in images of nature. In line with Weinstein, *et al.*, we then tested whether the conditional effect of images of nature, depending on the immersion score, was mediated by participants' connectedness to nature. In the second section, we tested an alternative hypothesis. It predicted that participants' prosocial propensity levels before the intervention would account for the effect of the interaction between immersion and images of nature on people's prosocial propensity after the intervention.

### Mediated Moderation

By using multiple-regression analyses and in accordance with Weinstein, *et al.* (2009), we corroborated their core finding that image type, immersion, and the interaction<sup>2</sup> of image type and immersion jointly and significantly accounted for people's prosocial propensity at T2 ( $F_{3,171}=4.74, p<.01, \eta^2=6.1\%$ ). Unlike Weinstein, *et al.*, we did not find that image type predicted prosocial propensity at T2 (after the intervention;  $\beta=-0.03, t_{171}=-0.36, p=.72$ ). We found that people's prosocial propensity at T2 increased with the extent to which they immersed themselves in either of the two image types ( $\beta=0.27, t_{171}=3.15, p<.01$ ). In contrast to Weinstein, *et al.*, again, the interaction of image type and immersion was not significantly related to people's prosocial propensity at T2 ( $\beta=0.14, t_{171}=1.64, p=.10$ ).

When we exclusively tested the anticipated effect (i.e., the interaction of image type and immersion; see Fig. 1) by excluding the non-significant effect of image type and the unanticipated immersion effect from the statistical model as recommended (e.g., Andersen, *et al.*, 2001), we found that the anticipated interaction was significant ( $F_{1,173}=4.32, p<.05, \eta^2=1.9\%$ ). Tests of simple main effects indicated that immersion in images of nature was linked with people's prosocial propensity at T2 ( $\beta=0.38, t_{87}=3.79, p<.001$ ), whereas immersion in urban images was not related to participants' prosocial propensity at T2 ( $\beta=0.10, t_{84}=0.94, p=.35$ ).

In line with Weinstein, *et al.* (2009), we additionally conducted a mediation test following the logic proposed by Kenny and colleagues (Baron & Kenny, 1986; Kenny, Kashy, & Bolger, 1998). This logic involves a statistically significant reduction in the strength of a relation after the mediator has been included. When the non-significant image type and the unanticipated immersion effect were again excluded from the statistical model (see Fig. 1), connectedness to nature mediated the interaction of image type and immersion ( $F_{2,172}=38.81, p<.001, \eta^2=30.3\%$ ). There was (a) a

<sup>2</sup>In order to reduce multicollinearity between the interactions and their components in our research, we mean-centered immersion before computing the interaction between immersion and image type.

significant effect on prosocial propensity at T2 from the interaction of image type and immersion—when not corrected for the mediator (i.e., connectedness to nature;  $\beta=0.18$ ,  $t_{172}=2.08$ ,  $p<.05$ ); (b) a positive association between the image-type-immersion-interaction and connectedness to nature ( $\beta=0.20$ ,  $t_{172}=2.68$ ,  $p<.01$ ); and (c) a positive association between participants' connectedness to nature and prosocial propensity at T2 ( $\beta=0.60$ ,  $t_{172}=8.46$ ,  $p<.001$ ). (d) Finally, when corrected for the mediator (i.e., connectedness to nature), the effect on prosocial propensity at T2 from the interaction of image type and immersion was no longer significant ( $\beta=0.05$ ,  $t_{172}=0.73$ ,  $p=.47$ ). A Sobel test confirmed the significance of the decrease in the effect of the interaction between image type and immersion on prosocial propensity at T2 when controlling for connectedness to nature ( $z=-2.56$ ,  $p<.05$ ).

### Confounding Factor Hypothesis

As we had expected (see Fig. 2), prosocial propensity at T1 (before the intervention) was linearly related to self-reported compliance with the immersion instructions ( $\beta=0.23$ ,  $t_{172}=3.04$ ,  $p<.01$ ). And given the high level of convergent validity between the two propensity measures as reflected in a correlation of  $r=.82$ , people's prosocial propensity at T2 (after the intervention) was also linearly related to self-reported compliance with the immersion instructions ( $\beta=0.25$ ,  $t_{173}=3.36$ ,  $p<.01$ ). With increasing levels of prosocial propensity before and after the intervention, participants claimed that they immersed themselves in the images to greater degrees, regardless of the type of image.

Given that the instructions had said to watch the entire presentation, the time that participants actually spent looking at the six images (excluding participants with implausibly excessive inspection times of  $\geq 10$  min.)<sup>3</sup> provided another, more objective measure of how diligently participants followed the immersion instructions. Compliance here basically meant that participants did not fast forward through the presentation. Similar to what we found for self-reported immersion, participants with higher prosocial propensity levels at T1 looked at images for longer periods of time (regardless of image type;  $\beta=0.16$ ,  $t_{172}=2.12$ ,  $p<.05$ ). After we excluded eight participants with implausibly excessive inspection times, the effect remained statistically unaffected and simply appeared a bit more pronounced in a mere descriptive sense ( $\beta=0.22$ ,  $t_{164}=2.90$ ,  $p<.01$ ). This time, however, the relation with actual inspection times held exclusively for prosocial propensity at T1.

The inspection-time effect on prosocial propensity at T2 was statistically non-significant ( $\beta=0.06$ ,  $t_{173}=0.81$ ,  $p=.42$ ) and remained non-significant after excluding participants with implausibly excessive inspection

times ( $\beta=0.11$ ,  $t_{165}=1.43$ ,  $p=.15$ ). Specifically, we even found that the length of time spent paying attention to urban images was significantly correlated with prosocial propensity at T1 ( $r=0.26$ ,  $p=.02$ ), whereas the length of time spent paying attention to images of nature was not ( $r=0.02$ ,  $p=.83$ ). In contrast to prosocial propensity at T1, the length of time spent paying attention to urban images ( $r=.16$ ,  $p=.14$ ) or images of nature ( $r=-.07$ ,  $p=.49$ ) was not significantly related to prosocial propensity at T2.

In contrast to the self-reports on immersion, when we considered the actual inspection times, we found that participants who had a more pronounced inclination to help others before we started our intervention (i.e., prosocial propensity at T1) complied better with our instructions to immerse themselves regardless of image type. This finding obviously held even for urban images, despite the fact that images of nature are known for their tendency to spontaneously hold people's attention to a somewhat greater degree (see Hartig, Korpela, Evans, & Gärling, 1997). As such, these differential findings for the T1 measure before the intervention and the T2 measure after the intervention support the differential sensitivity and, thus, the formal distinctiveness of the two prosocial propensity measures in our research.

Moreover, participants' prosocial propensity at T1 was significantly correlated with their prosocial propensity at T2 ( $r=.82$ ,  $p<.001$ ; see Fig. 2). Remember that a person's prosocial propensity was estimated before the intervention at T1 as a logit score on the GEB measure. After the intervention at T2, prosocial propensity was estimated as a second-order factor score derived from the common variance of four indicators [i.e., GEB, SVO, CA, including a person's overt donation (OD); see Table 2].

The various results so far speak of prosocial propensity as representing a confounding third variable, i.e., a variable that is correlated with one of the two independent variables as well as with the dependent variable (see Fig. 2). The correlation between the confounding factor and the dependent variable was not much of a surprise as the two represent a single concept (also see Table 1): people's inclination to act prosocially (although measured with different content and by employing different measurement models for the two instruments). However, the correlation between the confounding factor and immersion (one of the independent variables) was, by contrast, a genuine empirical test.

Predictably, we found support for our main hypothesis depicted in Fig. 2. When controlling for people's prosocial propensity levels at T1, the overall model—involving image type, immersion, the interaction between image type and immersion, and prosocial propensity at T1—significantly accounted for people's prosocial propensity at T2 ( $F_{4,169}=86.07$ ,  $p<.001$ ,  $\eta^2=67.1\%$ ). However, none of the three specific effects maintained significance: the in-

<sup>3</sup>Due to the fact that some participants answered the questionnaires in their homes, we were unable to oversee all our participants.

teraction ( $\beta=0.02$ ,  $t_{169}=0.42$ ,  $p=.67$ ), immersion ( $\beta=0.07$ ,  $t_{169}=1.44$ ,  $p=.15$ ), or image type ( $\beta=-0.01$ ,  $t_{169}=-0.16$ ,  $p=.87$ ). Only people's prosocial propensity at T1 significantly accounted for prosocial propensity at T2 ( $\beta=0.80$ ,  $t_{169}=17.46$ ,  $p<.001$ ).

With regard to Weinstein, *et al.*'s (2009) second core finding in which connectedness to nature mediated the interaction between images of nature and immersion, we also found that the effect of the interaction between images of nature and immersion on connectedness was attenuated by people's prosocial propensity at T1. When controlling for prosocial propensity, the interaction between image type and immersion decreased to a non-significant effect on connectedness to nature (from  $\beta=0.20$ ,  $t_{172}=2.68$ ,  $p<.01$ , to  $\beta=0.14$ ,  $t_{170}=1.93$ ,  $p=.06$ ). Obviously, people immerse themselves in images of nature to greater or lesser degrees because of their differential compliance with immersion requests. Nevertheless, with an increasing connectedness to nature, participants immerse themselves in images of nature to a greater degree, or at least they claim to be immersing themselves ( $r=.48$ ,  $p<.001$ ). No such relation was found for urban images ( $r=.09$ ,  $p=.43$ ).

## Discussion

This study tested the hypothesis that inter-individual differences in a person's propensity to act prosocially can compromise psychological research, particularly research on helping and helpfulness. Whereas we were able to replicate Weinstein, *et al.*'s (2009) core finding (i.e., the interaction between immersion and image type), we did not replicate the effect of image type. Rather, we found that, regardless of image type, the degree of immersion accounted for people's prosocial propensity at T2. In other words, the immersion effect turned out to be unconditional and, thus, even more comprehensive than what we had originally suspected. Jointly, the unexpected immersion effect and—when tested exclusively—the anticipated interaction between immersion and image type provide illustrative evidence for the suspected spurious causal effect that arises when research participants differ substantially in their readiness to help others.

As anticipated, a confounding third variable (i.e., prosocial propensity) was substantively correlated with one of the two independent variables (i.e., the extent of immersion). This association was  $\beta=0.23$ , when measured by self-reports of immersion, and  $\beta=0.16$  or  $0.22$ , when measured by inspection times (with or without implausible values, respectively). Participants with higher prosocial propensity levels looked for longer periods of time at images regardless of image type (including the comparatively less appealing urban images; cf. Hartig, *et al.*, 1997), and they also acknowledged that their immersion was comparatively more intense.

Again, as anticipated, when we controlled for people's prosocial propensity measured before the intervention, the immersion effect and the interaction effect, when tested exclusively, essentially disappeared. Evidently, people's prosocial propensity can be held statistically accountable for either the immersion effect or the interaction effect in our research. Thus, we continue to propose the general hypothesis that the more prosocial people are, the more closely they will adhere to experimenters' instructions and the more diligently they will execute requests, possibly including the exaggeration of self-reports on the extent of their immersion.

The present findings also suggest that the existence of the image effect that Weinstein, *et al.* (2009) reported should be questioned. Compared with Weinstein, *et al.*'s study, for which the power was approximately 0.50, the present study had superior statistical power of 0.77, which is close to the conventional standard of 0.80. With this superior power, we had a better chance of detecting even small effects (for more details on power analyses, see Faul, Erdfelder, Buchner, & Lang, 2009). Thus, our failure to find an image-type effect was presumably not due to a lack of power. Our non-significant finding rather indicates that the image-type effect that Weinstein, *et al.* reported with North American students might not generalize to other samples. Thus, it appears that delving into images of nature does not generally instigate more benevolence and prosocial behavior.

Alternatively, the image-type effect could have been attenuated in our research because of the particulars of the connectedness to nature measure that we used. Our measure of 40 items was longer than the one used in Weinstein, *et al.* (2009; 14 items). Our measure also focused primarily on past behavior and customs rather than on impressions at the particular moment of the assessment. This diversion from the actual situation could have weakened the influence of the presented images. At the very least and again, our failure to replicate Weinstein, *et al.*'s image-type effect leaves its universal validity dubious.

We were able to support our contention that the T2 criterion and the T1 covariate were measures of a single concept: the prosocial propensity of individuals. A correlation coefficient between prosocial propensity before and after the intervention as high as  $r=.82$ , in combination with a reliability coefficient of a comparable magnitude (i.e.,  $rel=.79$ ; see Table 1), indicates that the two measures represent the same concept (see Campbell & Fiske, 1959). Despite their expected convergent validity, the two measures were nevertheless sensitive to experimental influences that were apparent in some differential findings with the two estimates of the extent of immersion (i.e., measured with self-reports and inspection times).

As the two measures (i.e., the criterion and the covariate in our research) were formally distinct, their con-



vergence is not trivial from a methodological point of view. The T1 propensity measure consisted exclusively of self-reported ecological behavior items (i.e., no items referred directly to prosocial behavior). The T2 propensity measure, by contrast, involved, in addition to self-reported ecological behavior items, several instruments specifically designed to measure people's inclination to act prosocially. Also included in the T2 measure was a measure of real, overt helping performance. With an overlap of only six out of 64 ecological behaviors, the T2 criterion and the T1 covariate consisted of largely independent sets of items.

Moreover, the inclusion of ecological behavior in a more comprehensive prosocial propensity measure showed its expected convergence with other commonly used prosocial propensity measures—including the one used by Weinstein, et al. (2009). Hence, our finding replicated previous research in which the same or rather similar self-report measures of past ecological behavior provided substantial overlap with prosocial behavior (see Kaiser, 1998; Kaiser & Byrka, 2011) and with general cooperativeness or helpfulness (see Hilbig, et al., 2013).

Two features of this study can be criticized. The first shortcoming is our clustered convenience sample, which could be considered an oversampling of persons who potentially represented the two extremes of the inclination in question. Whereas samples consisting of oversampled extremes carry the risk of artificially exaggerating the strengths of effects, convenience samples, by contrast, can create non-generalizable findings.

The second shortcoming is that the present study was not an exact replication of Weinstein, et al.'s (2009) research. Specifically, our measures of the dependent variable (i.e., people's prosocial propensity) and the presumed mediator (i.e., people's connectedness to nature) differed from those used in Weinstein, et al.'s experiment. Because our goal was to pose an evidence-based generic challenge to certain findings from research on helping, an exact replication was not critical for meeting our objective. We believe that our study remains a tentative empirical illustration of our presumption that people's prosocial propensity might represent a confounding factor that deserves further exploration.

Asking participants with differential propensities to act prosocially may result in spurious findings and flawed conclusions about experimental effects if participants end up exaggerating their self-reports, adhering to instructions, or complying with commissions differentially. To avoid compromising the internal and external validity of psychological experiments, a change in the research practice of asking participants to adhere to instructions or provide self-reports (both potential prosocial performances) might be necessary: The first stage of such change would be for researchers to control for individual differences in people's prosocial propensity.

## Acknowledgment

We wish to thank Jane Zagorski for her language support, and Samuel Gosling, Terry Hartig, and Christoph Klauer for their comments on earlier versions of this article.

## References

- Andersen, D. R., Burnham, K. P., Gould, W. R., & Cherry, S. (2001) Concerns about finding effects that are actually spurious. *Wildlife Society Bulletin*, 29, 311-316.
- Baron, R. M., & Kenny, D. A. (1986) The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bartlett, M. S. (1937) The statistical conception of mental factors. *British Journal of Psychology*, 28, 97-104.
- Bond, T. G., & Fox, C. M. (2007) *Applying the Rasch model: fundamental measurement in the human sciences*. (2nd ed.) Mahwah, NJ: Erlbaum.
- Brügger, A., Kaiser, F. G., & Roczen, N. (2011) One for all? Connectedness to nature, inclusion of nature, environmental identity, and implicit association with nature. *European Psychologist*, 16, 324-333.
- Campbell, D. T., & Fiske, D. W. (1959) Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Embretson, S. E., & Reise, S. P. (2000) *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009) Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160.
- Greitemeyer, T. (2009) Effects of songs with prosocial lyrics on prosocial behavior: further evidence and a mediation mechanism. *Personality and Social Psychology Bulletin*, 35, 1500-1511.
- Hartig, T., Korpela, K., Evans, G. W., & Gärling, T. (1997) A measure of restorative quality in environments. *Scandinavian Housing and Planning Research*, 14, 175-194.
- Hilbig, B. E., Zettler, I., Moshagen, M., & Heydasch, T. (2013) Tracing the path from personality—via cooperativeness—to conservation. *European Journal of Personality*, 27, 319-327.
- Kaiser, F. G. (1998) A general measure of ecological behavior. *Journal of Applied Social Psychology*, 28, 395-422.
- Kaiser, F. G., & Byrka, K. (2011) Environmentalism as a trait: gauging people's prosocial personality in terms of environmental engagement. *International Journal of Psychology*, 46, 71-79.
- Kaiser, F. G., & Byrka, K. (2015) The Campbell paradigm as a conceptual alternative to the expectation of hypocrisy in contemporary attitude research. *The Journal of Social Psychology*, 155, 12-29.
- Kaiser, F. G., & Wilson, M. (2000) Assessing people's general ecological behavior: a cross-cultural measure. *Journal of Applied Social Psychology*, 30, 952-978.
- Kaiser, F. G., & Wilson, M. (2004) Goal-directed conservation behavior: the specific composition of a general performance. *Personality and Individual Differences*, 36, 1531-1544.
- Kaiser, F. G., Woelki, D., & Villasaliu, L. (2011) Partizipative Interventionsmaßnahmen und partizipatives umweltpolitisches Handeln: Ausdruck individueller Umweltmotivation, nicht deren Ursache [Participative interventions and participatory political activism for the environment: indicators not causes of intrinsic environmental motivation]. *Umweltpsychologie*, 15(2), 77-92.

- Kasser, T., & Ryan, R. M. (1993) A dark side of the American Dream: correlates of financial success as a central life aspiration. *Journal of Personality and Social Psychology*, 65, 410-422.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998) Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology*. (Vol. 1, 4th ed.) New York: McGraw-Hill. Pp. 233-265.
- Leiberg, S., Klimecki, O., & Singer, T. (2011) Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS One*, 6(3), e17798.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002) On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19-40.
- Mayer, F. S., & Frantz, C. M. (2004) The connectedness to nature scale: a measure of individuals' feeling in community with nature. *Journal of Environmental Psychology*, 24, 503-515.
- McClintock, C. G., & Allison, S. T. (1989) Social value orientation and helping behavior. *Journal of Applied Social Psychology*, 19, 353-362.
- Michell, J. (1986) Measurement scales and statistics: a clash of paradigms. *Psychological Bulletin*, 100, 398-407.
- Murphy, R. O., Ackermann, K., & Handgraaf, M. J. J. (2011) Measuring social value orientation. *Judgment and Decision Making*, 6, 771-781.
- Nichols, A. L., & Maner, J. K. (2008) The good-subject effect: investigating participant demand characteristics. *The Journal of General Psychology*, 135, 151-165.
- Orne, M. T. (1962) On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
- Pavey, L., Greitemeyer, T., & Sparks, P. (2011) Highlighting relatedness promotes prosocial motives and behavior. *Personality and Social Psychology Bulletin*, 37, 905-917.
- Shannon, M. L. (2004) Spurious relationships. In M. S. Lewis-Beck, A. Brymann, & T. Futing Liao (Eds.), *The Sage encyclopedia of social science research methods*. Thousand Oaks, CA: Sage. Pp. 1063-1064.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.
- Tabachnick, B. G., & Fidell, L. S. (2007) *Using multivariate statistics*. (5th ed.) Boston, MA: Allyn & Bacon.
- Thompson, B. (2004) *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, DC: American Psychological Association.
- Van Lange, P. A. M., Schippers, M., & Balliet, D. (2011) Who volunteers in psychological experiments? An empirical review of prosocial motivation in volunteering. *Personality and Individual Differences*, 51, 279-284.
- Weinstein, N., Przybylski, A. K., & Ryan, R. M. (2009) Can nature make us more caring? Effects of immersion in nature on intrinsic aspiration and generosity. *Personality and Social Psychology Bulletin*, 35, 1315-1329.

## Appendix

### 15 Newly Developed Ecological Behavior Items

1. Before I leave for holidays, I turn off the fridge.<sup>T1</sup>
2. When I take a break for more than 10 min., I turn off my computer screen.<sup>T1</sup>
3. I replace burnt-out light bulbs with energy-saving bulbs.<sup>T1</sup>
4. *I own an uplighter for illumination.*<sup>T1</sup>
5. I have activated the power-management feature on my computer.<sup>T1</sup>
6. I defrost my fridge whenever there is a layer of ice in the freezer compartment.<sup>T1</sup>
7. I use a water kettle.<sup>T1</sup>
8. I discuss energy-saving issues with friends and acquaintances.
9. When cooking, I put lids on the pots.
10. *After charging my mobile phone, I leave the charger plugged in to the socket.*
11. I wait until I have a full load before starting the dishwasher.
12. I let hot food cool down before putting it in the fridge.
13. I turn the lights off when I leave a room.
14. I am a vegetarian.
15. I invested in the thermal insulation of my flat or house.

**Note** <sup>T1</sup>Behavior items in the T1 version of the measure. The other behavior items were in the T2 version of the measure. *Items in italics* represent unecological activities. Such behavior items were reverse-coded.