

Do Subtle Reminders of Money Change People's Political Views?

Doug Rohrer
University of South Florida

Harold Pashler and Christine R. Harris
University of California, San Diego

A number of researchers have reported studies showing that subtle reminders of money can alter behaviors and beliefs that are seemingly unrelated to money. In 1 set of studies published in this journal, Caruso, Vohs, Baxter, and Waytz (2013) found that incidental exposures to money led subjects to indicate greater support for inequality, socioeconomic differences, group-based discrimination, and free market economies. We conducted high-powered replication attempts of these 4 money priming effects and found no evidence of priming (weighted Cohen's $d = 0.03$). We later learned that Caruso et al. also found several null effects in their line of research that were not reported in the original article. In addition, the money priming effect observed in the first study of Caruso et al. was included in the Many Labs Replication Project (Klein et al., 2014), and only 1 of the 36 labs was able to find the effect.

Keywords: social, priming, money, replication, failure

Several groups of researchers have reported studies in which subtle reminders of money produced remarkable changes in behaviors or beliefs that are ostensibly unrelated to money—a phenomenon known as money priming. Here we examine two of these papers: the seminal report of money priming effects published in *Science* by Vohs, Mead, and Goode (2006) and a follow-up paper published in this journal by Caruso, Vohs, Baxter, and Waytz (2013). We conducted additional analyses of these data sets, and we attempted to replicate the findings reported by Caruso et al.

The Reported Findings

Money priming effects were first observed in nine studies reported by Vohs et al. (2006). By random assignment, subjects in each study saw a money prime or not. The primes were subtle. In four of the studies, for instance, subjects rearranged words in order to create a short sentence (e.g., *Pick up the book*), and one half of the sentences seen by the money prime group were related to money (e.g., *We can afford it*). This money prime led subjects to

work longer on a puzzle before asking for help (Experiment 1), devote less time to helping an experimenter or a confederate (Experiments 3 and 4), and contribute less money to a charity (Experiment 6). In three other studies, the money prime was a poster or screensaver depicting U.S. currency, and this prime caused subjects to place their chairs farther from the chairs of other subjects (Experiment 7) and increased subjects' desire to be alone (Experiments 8 and 9). The effects were large. Averaged across studies, Cohen's d equaled 0.81.

In a follow-up paper published in this journal, Caruso et al. (2013) reported five studies using money primes like the ones used by Vohs et al. (2006). Two of the studies used the same sentence descramble task used in Vohs et al. (see Appendix A), and the money sentences led subjects to assert more strongly that “victims deserve their fate” (Experiment 2) and to indicate greater support for “group-based discrimination” (Experiment 3). In two other studies, subjects read instructions that were presented against a background that sometimes included an image of \$100 bills, and this prime led subjects to indicate greater support for the social structures in the United States (Experiment 1) and free market economies (Experiment 4). The effects were moderate or large, with the Cohen d values in Experiments 1 through 4 ranging from 0.44 to 0.80. (The dependent measures for Experiments 1 through 4 are given in Appendixes B, C, D, and E, respectively.) Finally, in their Experiment 5, Caruso et al. report a predicted three-way interaction that included a replication of the money priming effect observed in Experiment 4. The authors concluded, “Although the mere presence of money has been shown to result in benefits to individuals, such as enhanced goal pursuit (Vohs et al., 2006), the present work revealed that the concept of money also elicits more favorable attitudes toward existing systems that favor the socially advantaged and legitimize social inequality (p. 305).”

The effects reported by Vohs et al. (2006) and Caruso et al. (2013) were not only large but also robust. Effects were found with different priming tasks (e.g., descrambled sentences, posters, screensavers, and play money), a wide variety of measures (e.g., time spent working on a puzzle, willingness to help others, amount

Doug Rohrer, Department of Psychology, University of South Florida; Harold Pashler and Christine R. Harris, Department of Psychology, University of California, San Diego.

This research was supported by the Institute of Education Sciences, U.S. Department of Education (Grant R305A110517) to the University of South Florida, a collaborative activity grant from the James S. McDonnell Foundation, and the National Science Foundation (Grant SBE-582 0542013 to the University of California, San Diego, Temporal Dynamics of Learning Center). The opinions expressed are those of the authors and do not represent views of the funding organizations. We thank Eugene M. Caruso for providing us with the priming materials. We also thank Leif Nelson, Joseph Simmons, and Uri Simonsohn for their comments during the preparation of this paper, though we are solely responsible for the views expressed.

Correspondence concerning this article should be addressed to Doug Rohrer, University of South Florida, Psychology PCD4118G, Tampa, FL 33620 USA. E-mail: drohrer@usf.edu

of money donated, subject's placement of his or her chair, desire to be alone, and self-reported beliefs about fairness, justice, prejudice, and fair market economies), a variety of procedures (e.g., behavioral studies, online questionnaires, and paper-and-pencil questionnaires), and diverse subject populations (e.g., subjects recruited from Mechanical Turk, and college students at private, public, small, and large universities). The effects were also durable. In some of the studies in both papers, the prime and the dependent measure were separated by several minutes of intervening activities.

The large, robust effects reported by Vohs et al. (2006) and Caruso et al. (2013) are quite different from the well-established priming effects observed in perceptual and cognitive tasks (e.g., Meyer & Schvaneveldt, 1971). For example, in studies using the lexical decision paradigm, subjects are faster to indicate that a letter string is a word (DOCTOR . . . "yes") if the word immediately follows a related word (NURSE) rather than an unrelated word (BREAD). Although these priming effects are reliable, the effect sizes are small (Cohen's $d < 0.20$, e.g., Pashler, Coburn, & Harris, 2012). Semantic priming effects are also fleeting. For example, Becker, Moscovitch, Behrmann, and Joordens (1997) found that lexical decision priming effects disappeared if the prime and target were separated by more than 15 s, and similar finding were reported by Meyer, Schvaneveldt, and Ruddy (1972). In brief, classic priming effects are small and transient even if the prime and measure are strongly associated (e.g., NURSE-DOCTOR), whereas money priming effects are reportedly large and relatively long-lasting even when the prime and measure are seemingly unrelated (e.g., a sentence related to money and the desire to be alone). This contrast led us to look more closely at money priming effects.

Unreported Null Effects

The studies reported by Vohs et al. (2006) and Caruso et al. (2013) exhibit a salient pattern. For the effects reported in each paper, a disproportionately large number of the p values are slightly less than .05, and none are greater than .05. Put another way, the sample sizes were generally just large enough for the effect size to be statistically significant, even though sample sizes varied across studies. Thus, the studies with the smaller samples produced the largest effect sizes (Cohen's d). In fact, this inverse relation between sample size and effect size is strong enough to be visibly discernible, as shown by the scatterplots in Figure 1. For the studies in Vohs et al., the inverse relation produced a large negative correlation, $r = -0.88$, $p = .004$. For the studies in Caruso et al., the Pearson correlation was large but not statistically significant, $r = -0.89$, $p = .11$. Although the latter p value fell short of statistical significance, the Pearson r is a conservative measure because it measures the strength of a linear relationship, and the data points in the Caruso et al. scatterplot are nonlinear but clearly inversely related.

A published set of studies will exhibit scatterplots like those in Figure 1 when null effects are not reported. Two kinds of unreported null effects are considered here: publishing a set of studies without reporting that the same line of research included studies showing a null effect (unreported study), and publishing a study without reporting that the procedure included a dependent measure that was not affected in the predicted fashion (unreported mea-

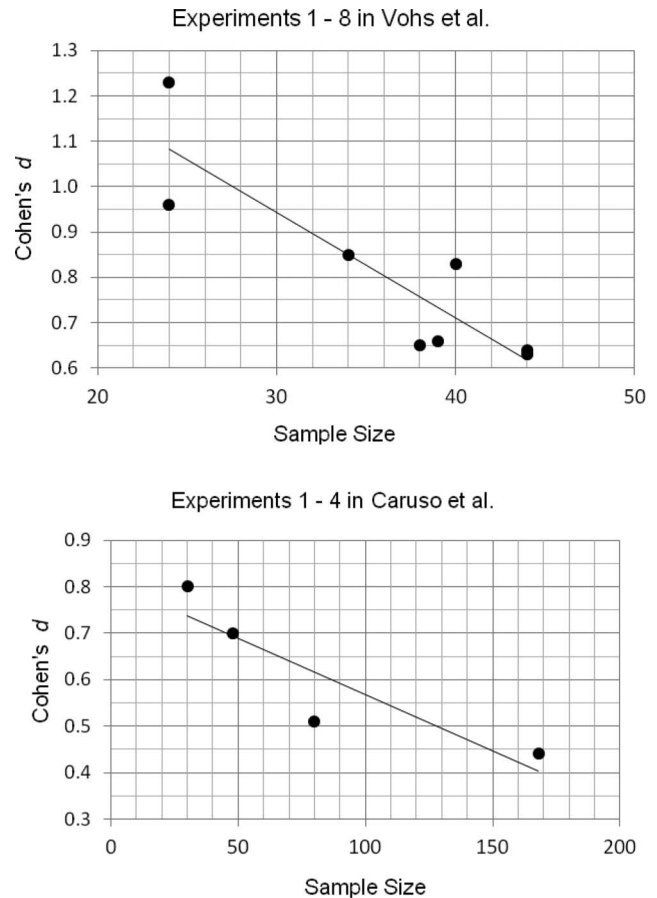


Figure 1. Relationship between effect size and sample size for studies reported by Vohs et al. (2006) and Caruso et al. (2013). Each point represents an experiment. Lines represent least-square fits. Effect sizes (Cohen's d) and sample sizes are given in Table 1. In Vohs et al. (2006), three of the studies provided two money priming effects each (which were not independent), and for each of these studies, we averaged the two d values. Vohs et al. reported a ninth experiment that could not be included here because the measure was categorical. Caruso et al. reported a fifth experiment that was excluded because we could not estimate the effect size.

sure). Unreported null effects lead to a disproportionately large number of p values that are just small enough to be statistically significant (e.g., Simmons, Nelson, & Simonsohn, 2011; Simonsohn, Nelson, & Simmons, 2014).

We wondered whether unreported null effects might explain why the studies reported by Vohs et al. (2006) and Caruso et al. (2013) produced the inverse relations shown in Figure 1. We contacted the first author of each paper and asked whether they had conducted unreported studies and whether their published studies included unreported dependent measures. We also asked each author how they determined the sample size for their studies. Both Vohs and Caruso provided detailed responses to our questions. We wrote two summaries of their responses—one for each author—and sent both summaries and an earlier version of the present paper to both authors. After some back and forth, both authors gave us permission to include our summary in the present paper.

For the studies in Vohs et al. (2006), K. Vohs approved the following summary of her responses to our questions (personal communications, June 15, 2013; November 1, 2013):

(1) With regard to sample sizes and termination rules, although sample sizes varied across studies, sample sizes were not based on any form of preliminary data analyses. Instead, the variation in sample size reflected various logistical constraints—not the cessation of data collection once statistical significance was achieved.

(2) With regard to unreported studies, the authors conducted two additional money priming studies that showed no effects, the details of which were shared with us.

(3) With regard to unreported measures, Experiments 1 through 6 included measures of potential mediators, specifically, the self-construal scale (Singelis, 1994); power motivation scale (Schmidt & Frieze, 1997); leadership motivation subscale from the power motivation scale (Schmidt & Frieze, 1997); and an author-developed measure of power. A measure of chair positioning was included in Experiment 7. Of these 19 additional outcome measures, none were statistically affected by the manipulation (although 4 of the 19 approached significance, all in the predicted direction). In other words, Vohs et al. reported nine dependent measures that were statistically affected by the manipulation in the predicted direction (one in each experiment) but did not report 19 additional measures that were statistically unchanged.

In brief, Vohs et al. observed null effects in their two unpublished studies, and they observed 19 null effects in their nine published studies, though 18 of the 19 measures were mediators. None of the null effects were reported in the published paper or the supplementary online materials.

For the studies in Caruso et al. (2013), E. Caruso approved the following summary of his responses to our questions (personal communications, May 14, 2013; October 10, 2013):

With regard to termination rules, Caruso and his collaborators acknowledged that they did not have well-defined stopping rules in place for all studies. With regard to unreported studies, the authors conducted four money priming studies that were not included in the paper: a study similar to Experiment 5 that produced mixed results, a study similar to Experiment 4 with a result that was nearly statistically significant ($p = .10$), a study similar to Experiment 4 that did not find an effect, and a study using a manipulation unlike any of those used in any of the studies published in Caruso et al. With regard to unreported measures, two studies included a measure that the authors did not mention in their published report. In Experiment 1, the reported measure was followed by a 3-item measure of belief in social Darwinism (none of the three items was affected). In Experiment 2, the reported measure was preceded by an 8-item measure of beliefs about taxes (five of the eight items were unaffected, but a composite of the other three showed an effect in the predicted direction).

In brief, Caruso et al. conducted four studies showing a null effect that were not included in the paper, and two of their five reported studies included one or more unreported dependent measure that, based on their hypothesis, should have been affected by the money priming manipulation.

To summarize, both Vohs et al. (2006) and Caruso et al. (2013) observed multiple null effects but did not report any. To be fair, these omissions did not violate the then-policy of the journals in which these papers were published. Still, we report these omissions because these numerous null effects increase the chance that

the money priming phenomena reported by these authors are not real.

Previous Replication Attempts

After we submitted the second version of the present paper to this journal, we learned that the money priming effect found in Experiment 1 of Caruso et al. (2013) was one of the findings included in a massive replication attempt known as the Many Labs Replication Project (Klein et al., 2014). In this project, each of 36 laboratories (total $n = 6344$) tried to replicate each of 13 findings from cognitive or social psychology. Averaged across laboratories, only 2 of the 13 effects did not replicate, and the weaker of the 2 null effects was the money priming effect, Cohen's $d = -0.02$, 95% CI = $(-0.08, 0.02)$. In the original study by Caruso et al. ($n = 30$), Cohen's $d = 0.80$, 95% CI = $(0.05, 1.54)$.

The Present Studies

Here we present direct replication attempts of four money priming studies reported by Caruso et al. (2013). We considered beginning with the money priming studies reported by Vohs et al. (2006), but the measures in these studies required social interaction between the subject and experimenter, making it difficult to know whether the replication attempt was sufficiently similar to the original study. By contrast, the first four studies reported by Caruso et al. used simple, easily reproduced priming tasks and questionnaire measures involving no social interaction. In fact, two of the studies were conducted online.

We attempted to replicate Experiments 1 through 4 in Caruso et al. (2013). Our experiments are numbered in correspondence with the original studies—for example, our Experiment 1 is a replication attempt of the Experiment 1 in Caruso et al., and so forth. As in each of the original experiments, subjects in the present studies were randomly assigned to the money prime group or the control group. Caruso et al. also reported a fifth experiment in which they replicated the money priming effect observed in their Experiment 4 (same priming manipulation, same dependent measure) as part of a predicted three-way interaction. We did not attempt to replicate this fifth study, partly because it involved the same priming effect observed in Experiment 4, and partly because of logistical constraints. (The study included author-created measures and took place in a museum.)

We report our findings with full disclosure. We report the results of Experiments 1–4 and our first two versions of Experiment 2, and these six studies are the only money priming studies that any of us has conducted as of the present date (December 2014). We also report all conditions and all measures for each study, and we describe the procedure in full. The number of subjects in the final version of Experiment 2 was selected in advance (as detailed in the introduction to that study). For each of the other studies, we chose in advance an approximate sample size that far exceeded the number of subjects in the corresponding original study. For these studies, the research assistants who collected the data periodically informed one of us how many subjects had been tested, and data collection was terminated once the sample size exceeded our planned sample size. (In each study, as many as two dozen subjects were tested each day.) Neither we nor our research assistants examined the effect sizes, formally or informally, until after we terminated data collection.

Experiment 1

The first study was an attempt to replicate the first experiment in Caruso et al. (2013). Subjects read instructions that were set against a background image that displayed either U.S. currency (money prime) or a degraded version of the same image (control), as shown in Figure 2. Subjects then completed a questionnaire designed to assess their support for the economic and social structures in the United States—a construct known as “system justification” (see Appendix B). Caruso et al. found that the money prime increased scores on this measure.

Method

Participants. We tested 136 subjects. This sample is about 4.5 times as large as the sample of 30 used in the original study reported by Caruso et al. (2013). The present sample was slightly older (35 vs. 32) and slightly more female (70% vs. 57%) than the original sample. Subjects in the present study were drawn from an

online pool maintained by the University of California, San Diego (UCSD), Learning, Attention, and Perception lab. Subjects in the pool have passed an English proficiency test, and the sample for this experiment was further restricted to U.S. citizens. Each received \$1. Subjects in the original study also were drawn from a university study pool, and each of those subjects participated in exchange for a chance to win \$25.

Design. Each subject was randomly assigned to either the money condition ($n = 69$) or the control condition ($n = 67$).

Procedure. Subjects completed the study online. Subjects first read general instructions (71 words) displayed in a blue font against a background image of \$100 bills (money priming) or a degraded version of the same image (control), as shown in Figure 2. Both images are identical to the ones used in Caruso et al. (2013). Subjects then answered three short demographic questions (age, gender, and ethnicity). Immediately afterward, subjects completed the 8-item system justification scale (see Appendix B) created by Kay and Jost (2003). Although Kay and Jost used a 9-point scale, we used a 7-point scale in accordance with the original study reported by Caruso et al. Finally, subjects answered 3 debrief questions. (Do you remember the image that was displayed on the instructions page? If so, please describe it in a phrase. While you were doing this experiment, did you have any particular thoughts about the purpose of this study? If so, please elaborate. Please make any final comments you would like to make, if any, about this very brief study.) The debrief questions appeared one at a time, and subjects could not alter their answers to a previous question.

After we completed the experiment and sent a demonstration version of our study to the first author of Caruso et al. (2013), he pointed out two procedural differences between the original study and the present study (E. Caruso, personal communication, October 31, 2012). First, in the original study, the instructions were placed within a textbox, whereas the instructions in the present study were displayed directly against the image. Second, in the original study, the instructions and three demographic questions appeared on the same screen, whereas subjects in the present study saw the three demographic questions on a separate screen. (In the original paper, the authors wrote, “Following the instructions screen, participants indicated their age, gender, and ethnicity,” p. 302.)

Results and Discussion

Whereas the money prime increased mean score in the original study, the present study found a small, unreliable effect in the opposite direction, $t(134) = -0.43$, $p = .67$ (Table 1, Figure 3). Cohen’s $d = -0.07$ and its 95% CI ($-0.41, 0.27$) excludes the effect size of $d = 0.80$ observed in the original study (see Figure 4). No subjects indicated in their answers to the debrief questions that the aim of the study was to assess the effect of the money prime on their survey responses.

Experiment 2

The second study was an attempt to replicate Experiment 2 in Caruso et al. (2013). Subjects desrambled sets of words to create sentences that were sometimes related to money (money priming) or never related to money (control). Immediately afterward, sub-



Figure 2. The images used in Experiments 1 and 4. Subjects saw either the money image (money prime) or the neutral image (control).

Table 1
Results of Original Studies by Caruso et al. (2013) and Replication Attempts

	Condition	<i>M</i>	<i>SD</i>	<i>N</i>	Power ^a	Cronbach's α	Cohen's <i>d</i>
Experiment 1							
Original	Money	4.96	1.27	30	0.56	0.88	0.80
	Control	3.99	1.19				
Replication	Money	3.76	0.97	136	0.99	0.80	-0.07
	Control	3.83	1.02				
Experiment 2							
Original	Money	2.53	0.52	168	0.81	0.68	0.44
	Control	2.32	0.44				
Replication	Money	2.59	0.50	420	0.99	0.76	0.06
	Control	2.56	0.55				
Experiment 3							
Original	Money	3.20	1.32	80	0.62	0.92	0.51
	Control	2.56	1.24				
Replication	Money	2.10	0.89	156	0.89	0.90	-0.06
	Control	2.15	0.88				
Experiment 4							
Original	Money	1.10	1.74	48	0.66	—	0.70 ^b
	Control	-0.02	1.43				
Replication	Money	0.68	1.38	116	0.96	0.91	0.14
	Control	0.48	1.55				

^a Power analyses assumed two-tailed tests and $\alpha = .05$, and were based on the effect size observed in the original experiment (Faul, 2009). ^b Caruso et al. did not report an effect size for Experiment 4. We computed this value by assuming that the money and control groups had equal sample sizes.

jects completed a questionnaire designed to measure the extent to which they believe society is just (*Belief in a Just World*, Appendix C). Caruso et al. found that money priming increased scores on this measure.

As noted in the introduction, we conducted three versions of Experiment 2. The third version is reported here, and the first two versions are reported in Appendix F. In the first attempt, we used

a Likert scale for the dependent measure that differed from the one used in the original study (partly because of an error in the description of the original study, and partly because of an error on our part). After we sent a demo and the results to D. Caruso, he informed us of both differences (personal communication, October 31, 2012.) We then conducted a second version using the corrected scale. The second version was included in an earlier version of this

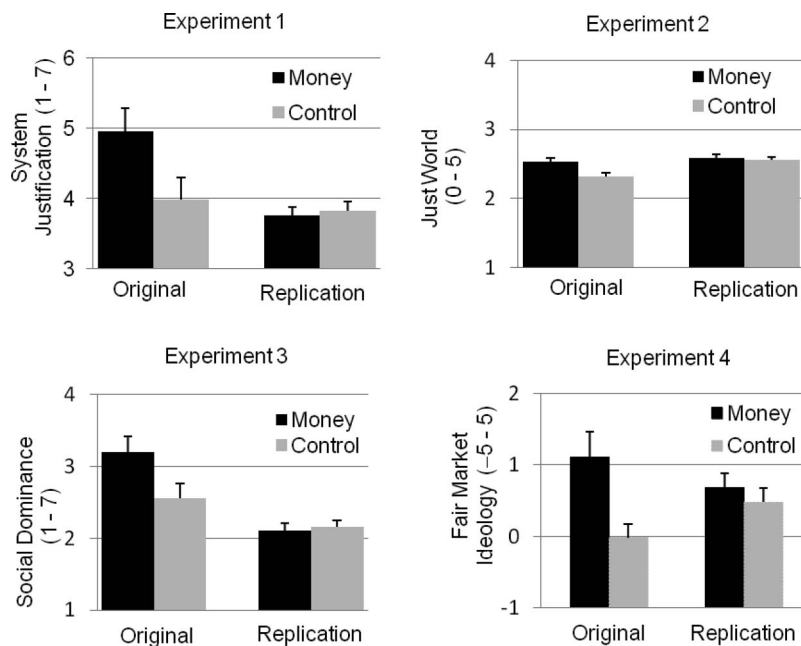


Figure 3. Results of the original studies (Caruso et al., 2013) and the replication attempts presented here. Priming effects were observed in each of the original studies but in none of the replication attempts. Error bars represent one standard error.

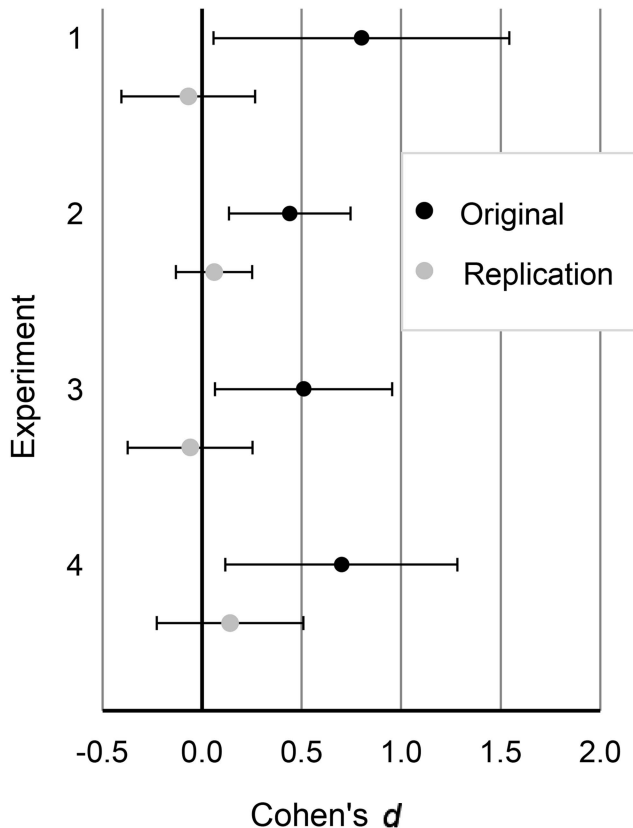


Figure 4. Effect sizes and 95% confidence intervals for the original studies and the replication attempts reported here. The effect size in each of the original studies lies outside the confidence interval of the effect size observed in the replication attempt.

paper, and a reviewer asked that we repeat the study with a sample size that equaled 2.5 times the sample size used in the original, which we did. Neither of the first two versions produced a statistically significant effect.

Method

Participants. We tested 420 subjects, and this sample is 2.5 times as large as the sample of 168 used in the original study reported by Caruso et al. (2013). The present sample was older (36 vs. 21) and more female (57% vs. 47%) than the original. We recruited subjects from Amazon's Mechanical Turk who were U.S. residents with a minimum approval rate of at least 95%. Each passed an online English proficiency test, and those completing the study were paid \$0.40. The original sample was drawn from a pool of college students.

Design. Each subject was randomly assigned to either the money condition ($n = 211$) or the control condition ($n = 209$).

Procedure. Subjects first read instructions and reported their age, gender, political ideology (7-point scale with endpoints *Republican/Conservative* and *Democrat/Liberal*), religiosity (7-point scale with endpoints *Not at all* and *Extremely*), and wealth (7-point scale with endpoints *Not at all* and *Extremely*). They then completed the sentence descrambling tasks (see Appendix A). For each

of 30 sentences, the subject selected and reordered 4 of 5 presented words to create a sentence or phrase. For example, the words “we cup afford can it” yields the sentence “we can afford it.” Subjects provided their answer by clicking on the 4 words in the correct order, and they could restart each sentence before submitting each answer. In the money condition, 15 of the 30 descrambled sentences were related to money.

Immediately after the descramble task, subjects completed the 20-item belief in a just world scale (Appendix C; Rubin & Peplau, 1975). Finally, subjects received 3 debrief questions. (At the beginning of the study, you solved some puzzles. Did you notice any common theme to these puzzles? While you were doing this experiment, did you have any particular thoughts about the purpose of this study? If so, please elaborate. Please make any final comments you would like to make, if any, about this very brief study.) The debrief questions appeared one at a time, and subjects could not alter their answers to a previous question.

We know of two procedural differences between the present study and the original study. Subjects in the present study answered all questions on a computer whereas subjects in the original study completed the study on paper, and subjects in the present study answered debrief questions after they had completed the study.

Results and Discussion

The money prime produced no effect, $t(418) = 0.58$, $p = .56$ (Table 1, Figure 3).

We observed a Cohen's d of 0.06, and its 95% CI ($-0.14, 0.25$) excludes the effect size of 0.44 observed in the original study (see Figure 4). In their responses to the debrief questions, three subjects indicated that they believed that the purpose of the study (debrief question 2) was to assess the influence of money on their attitudes or beliefs (e.g., “I thought the puzzles about people having large sums of money were supposed to influence our thoughts.”).

Experiment 3

In the third study, we attempted to replicate Experiment 3 in Caruso et al. (2013). Subjects completed the same sentence descrambling task (money or control) used in Experiment 2 and then completed a measure of “social dominance” that reportedly measures support for social hierarchy and group-based discrimination (see Appendix D). Caruso et al. found that money priming increased scores on this measure.

Method

Participants. We tested 156 adults, which is about twice as large as the sample of 80 subjects used in the original study reported by Caruso et al. (2013). The present sample was about the same age (21 vs. 20), but more female (83% vs. 50%) than the original. Subjects in the present study were drawn from the UCSD psychology subject pool, and each passed an online English proficiency test before beginning the study. Each received partial course credit.

In the original study, adults in a university dining hall participated in exchange for a candy bar.

Design. Each subject was randomly assigned to either the money condition ($n = 78$) or the control condition ($n = 78$).

Procedure. Each subject completed the study on a computer while sitting alone in a small laboratory room. Each subject first read instructions and then completed the sentence descrambling task used in Experiment 2 (see Appendix A). Subjects then indicated their gender, age, and ethnicity before completing the 16-item social dominance scale (Appendix D; Pratto, Sidanius, Stallworth, & Malle, 1994). Finally, subjects answered the same debrief questions used in Experiment 2.

We know of two procedural differences between the present study and the original study in Caruso et al. Subjects in the present study completed the task on a computer while sitting alone in a laboratory room while subjects in the original study completed a paper-and-pencil questionnaire in a university dining hall. Second, whereas the demographic questions appeared immediately *before* the descramble task in the original study, the demographic questions (age, gender, and ethnicity) in the present study were presented immediately *after* the descramble task because of an inadvertent error. However, the three demographic questions required no more than about 10 seconds to answer (the answers to the gender and ethnicity questions were chosen from a drop-down menu), and this additional duration of time was much smaller than the time devoted to either the manipulation (a 30-sentence descramble task requiring more than 5 min) and the dependent measure (the 16-item questionnaire).

Results and Discussion

The priming effect was not statistically significant and in the opposite direction of the effect observed in the original study, $t(154) = -0.36, p = .72$ (Table 1, Figure 3). Cohen's $d = -0.06$ and its 95% CI = $(-0.37, 0.26)$ excludes the effect size of 0.51 observed in the original study (see Figure 4). In their answers to the debrief questions, no subjects indicated that the aim of the study was to assess the effect of the money prime on their survey responses.

Experiment 4

Our final study was a replication attempt of Experiment 4 in Caruso et al. (2013). Subjects saw either the money image or the neutral image used in Experiment 1 (see Figure 2) and then immediately completed a measure of their support for free market economies (see Appendix E).

Unlike in the previous studies, both the present study and the corresponding original study included both U.S. and non-U.S. residents. Caruso et al. found a significant money priming effect for U.S. subjects ($d = 0.70$), but not for non-U.S. subjects ($d = 0.12$).

Because Caruso et al. (2013) found a money priming effect with U.S. subjects only, that finding is the only one that is of interest here. Therefore, our replication attempt included far more U.S. subjects than in the original study. We also tested non-U.S. subjects simply to verify that non-U.S. subjects are not affected, as observed by Caruso et al.

Method

Participants and design. We tested 116 U.S. residents (mean age = 37.3, 56% women) and 228 non-U.S. residents (mean age =

31.0, 37% women). The original study by Caruso et al. (2013) included 275 adults (mean age = 31.5, 38% women). Caruso et al. did not describe the U.S. and non-U.S. samples separately, but the statistical data allowed us to deduce that their sample included 48 U.S. adults and 227 non-U.S. adults. Thus, our U.S. sample was about 2.5 times as large as the U.S. sample in the original study. The two studies included nearly the same number of non-U.S. students. Subjects in both studies were recruited from Amazon's Mechanical Turk and paid \$0.25. We also required that each of our subjects pass an online English proficiency test.

Design. Each subject was randomly assigned to either the money condition ($n = 50$ U.S. and 117 non-U.S.) or the control condition ($n = 66$ U.S. and 111 non-U.S.).

Procedure. Subjects completed the study online. They first saw general instructions with one of the two background images used in Experiment 1 (see Figure 2). Next, they reported their age, gender, and nationality. Subjects then completed the 25-item fair market ideology scale (Appendix E; Jost, Blount, Pfeffer, & Hunyady, 2003). The study concluded with the debrief questions used in Experiment 1.

Results and Discussion

The U.S. residents did not show a money priming effect, $t(114) = 0.76, p = .46$ (Table 1, Figure 3). The effect size (d) was 0.14, and its 95% CI $(-0.23, 0.50)$ excludes the effect size of 0.70 observed in the original study (see Figure 4). Responses to debrief questions showed no evidence that subjects inferred that the aim of the study was to assess the effect of the money prime on their survey responses. Less importantly, and as in the original study, the non-U.S. sample also showed no evidence of money priming, $t(226) = 0.37, p = .71, d = .06, 95\% \text{ CI} = (-0.20, 0.32)$. For these non-U.S. subjects, the money priming group averaged 0.84 ($SD = 0.83$), and the control group averaged 0.79 ($SD = 0.96$). In brief, the present study did not find a money priming effect with either U.S. residents or non-U.S. residents.

General Discussion

We report two lines of evidence that raise doubts about the phenomenon of money priming. First, we learned that the Vohs et al. (2006) and Caruso et al. (2013) observed multiple null effects in both their reported and unreported studies that were not mentioned in their papers. Second, we conducted high-powered replication attempts of the four money priming effects found by Caruso et al. and found no evidence of priming (Figures 3 and 4).

Why Did the Replication Attempts Fail?

Direct replication attempts sometimes do not find an effect even when the effect is real, and we evaluate five conventional explanations of why this might happen.

1. Random measurement error? A failure to replicate should be interpreted with caution, and this is partly because random measurement error can cause two identical experiments to produce different results (e.g., Stanley & Spence, 2014). In the present case, however, we report not one but four replication failures. Furthermore, and as noted in the introduction, the money priming effect observed in Experiment 1 in Caruso et al. (2013)

was one of the findings included in the Many Labs Replication Project (Klein et al., 2014), and only 1 of the 36 labs found an effect – a ratio that is less than the alpha level of $.05 = 1/20$. So many null effects cannot be plausibly attributed to noise.

2. Too little power? Because of chance and insufficient power, replication failures can sometimes represent a Type II error (an erroneous failure to reject a null hypothesis when an effect exists). However, this also is an unlikely explanation of the failed replications reported here. The sample size for each of the replication attempts was two to four times as large as the sample size of the corresponding original study, and the power of each of our studies to find the effect reported by Caruso et al. (2013) ranged between 0.89 and 0.99 (see Table 1).

3 Demand effects? We initially wondered if money priming effects might reflect demand characteristics, wherein subjects infer the experimenters' hypothesis and play along. For this reason, we added three debrief questions at the end of each of the present studies, and subjects were asked to speculate on the purpose of the experiment. However, on the basis of their responses, less than 1% of the subjects in the money priming conditions correctly surmised the purpose of the study. We conclude that these money manipulations are indeed as subtle and rarely noticed as they were designed to be, and we see no reason to suspect that the subjects tested by Caruso et al. would have noticed them any more often than did our subjects. Thus, demand effects seem to be an unlikely explanation of the conflicting results.

4. Procedural differences? Replication failures also can be caused by procedural differences between the original study and a replication attempt (Simons, 2014). However, the procedures of the replication attempts reported here closely mirror those of the original studies. (Each of these differences is detailed in the method section of each study.) Furthermore, it is hard to understand how minor procedural differences would eliminate the money priming effect if the effect is as robust as suggested by the results of Caruso et al. (2013). These authors found a money priming effect with two kinds of manipulations, four different measures, several kinds of procedures (including online and paper-and-pencil questionnaires), and a variety of subject populations (including university students and Mechanical Turk). Given this absence of boundary conditions, it seems rather implausible that minor procedural differences would change the effect size. Moreover, why would procedural differences always eliminate the effect rather than occasionally decrease or increase the effect size?

5. Type I error in original studies? Another possibility is that the original findings might have represented false positives. A Type I error is a perfectly plausible explanation whenever a single result cannot be replicated, but the present case is rather different. Caruso et al. (2013) found consistently medium-to-large effects in all four of the studies that we failed to replicate. Furthermore, in their Experiment 5, these authors replicated the money priming effect found in Experiment 4 as part of significant three-way interaction that conformed qualitatively to the authors' predictions (and given the fact that interactions can take many forms, this would be expected to rarely occur by chance). If money priming does not affect attitudes about social inequality in the way the authors conclude, the chance of this broad and impressive pattern of large effects occurring by chance in all five studies is extremely remote. Even when accounting for the six null effects found but not reported by Caruso et al. (see the introduction), these authors

found a significant effect ($p < .05$) in 5 of 11 attempts. When $p = .05$ and the effect is not real, the chance of observing 5 or more successes in 11 attempts is about 1 in 10,000. This value equals the cumulative binomial probability of at least 5 successes in 11 trials, or

$$\sum_{k=5}^{11} \binom{11}{k} (.05)^k (.95)^{11-k}.$$

Other Reports of Money Priming Effects

Although we focus here on the seminal report by Vohs et al. (2006) and the follow-up studies by Caruso et al. (2013), money priming effects have been reported by other researchers. Pfeffer and DeVoe (2009) found that subjects who descrambled money sentences (rather than neutral sentences) were less willing to volunteer ($d = 0.34$ by our calculation). Mogilner (2010) gave the sentence descramble task to café customers and later surreptitiously observed that money-primed subjects spent more time reading or working on a computer than did the controls ($d = 0.75$ by our calculation). In another study using the sentence descramble task, Molinsky, Grant, and Margolis (2012) reported that the money prime reduced subjects' compassion and empathy ($d = 0.58-0.64$, by our calculation). In a study by Roberts and Roberts (2012), middle school students completed a questionnaire that sometimes included a picture of a \$100 bill, and this money prime significantly decreased one of the two measures of subjects' self-reported willingness to donate ($d = 0.37$, by our estimation). Finally, Chatterjee, Rose, and Sinha (2013) report three money priming studies showing astonishingly large effects. In two of the studies, for example, subjects descrambled sentences that were related to cash or credit cards, and this manipulation altered subjects' willingness to donate money or time to a charitable cause (Cohen's $d = 2.19$ and 3.46 , respectively, by our calculation). We do not know of any attempts to replicate any of these priming effects, nearly all of which were obtained with the same priming manipulations used by Vohs et al. (2006) and Caruso et al. (2013).

Conclusion

We have been unable to find evidence for the money priming effects reported by Caruso et al. (2013). Each of the four high-powered replication attempts reported here found an effect size that was not statistically different from zero (weighted Cohen's $d = 0.03$), and the 95% confidence interval of each effect excludes the effect size observed in the corresponding original study (Table 1, Figure 3 and 4). The null effects supplement the recent finding that the money priming effect reported in Experiment 1 of Caruso et al. produced no effect (Cohen's $d = -0.02$) in the Many Labs Replication Project described in the introduction (Klein et al., 2014). Although replication failures should be interpreted with caution, the sheer number of so many high-powered replication failures cast doubt on the money priming effects found by Caruso et al. Although we did not attempt to replicate any of the findings reported by Vohs et al. (2006), nearly all of their reported effects were found with priming manipulations that were identical to or similar to the manipulations that were used by Caruso et al.

We cannot explain why Caruso et al. (2013) found so many money priming effects. One possibility is that all five of their

effects are Type I errors, but this appears to be extremely unlikely, even when accounting for the six null effects they also found. As explained above, the chance of finding five Type I errors in 11 attempts is about 1 in 10,000.

We doubt that the many failures to replicate the findings of Caruso et al. (2013) can be attributed to procedural differences between the replication attempts and the original studies. Each replication attempt and the corresponding original study relied on the same priming manipulation and the same dependent measure, and the subject samples were drawn from similar or identical populations. There were minor procedural differences between the replication attempts and the original studies, but it is implausible that such minor differences would entirely eliminate effects that are reportedly so robust. As detailed in the introduction, Caruso et al. found medium-to-large effects with a variety of methods, including two different manipulations, four dependent measures, several kinds of procedures (online and paper-and-pencil), and different subject populations. Given this paucity of boundary conditions, we find it implausible that procedural differences, minor or otherwise, would always eliminate the effect rather than occasionally weaken the effect or, for that matter, strengthen it.

Finally, we have included here all of the stimuli and measures used in our studies, and we urge other researchers to do the same. Many published reports of money priming effects exclude some of the most basic information about the methodology, and we encourage these authors to make publicly available both their methodology and results so that other researchers can conduct direct replications. Direct replication is the final arbiter.

References

- Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1059–1082. <http://dx.doi.org/10.1037/0278-7393.23.5.1059>
- Caruso, E. M., Vohs, K. D., Baxter, B., & Waytz, A. (2013). Mere exposure to money increases endorsement of free-market systems and social inequality. *Journal of Experimental Psychology: General*, *142*, 301–306. <http://dx.doi.org/10.1037/a0029288>
- Chatterjee, P., Rose, R. L., & Sinha, J. (2013). Why money meanings matter in decisions to donate time and money. *Marketing Letters*, *24*, 109–118. <http://dx.doi.org/10.1007/s11002-012-9215-0>
- Faul, F. (2009). G*Power (Version 3.1.2) [Computer software]. Retrieved from www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/
- Jost, J. T., Blount, S., Pfeffer, J., & Hunyady, G. (2003). Fair market ideology: Its cognitive-motivational underpinnings. *Research in Organizational Behavior*, *25*, 53–91. [http://dx.doi.org/10.1016/S0191-3085\(03\)25002-4](http://dx.doi.org/10.1016/S0191-3085(03)25002-4)
- Kay, A. C., & Jost, J. T. (2003). Complementary justice: Effects of “poor but happy” and “poor but honest” stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology*, *85*, 823–837. <http://dx.doi.org/10.1037/0022-3514.85.5.823>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, *45*, 142–152. <http://dx.doi.org/10.1027/1864-9335/a000178>
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234. <http://dx.doi.org/10.1037/h0031564>
- Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1972). *Activation of lexical memory*. Paper presented at the meeting of the Psychonomic Society, St. Louis, MO.
- Mogilner, C. (2010). The pursuit of happiness: Time, money, and social connection. *Psychological Science*, *21*, 1348–1354. <http://dx.doi.org/10.1177/0956797610380696>
- Molinsky, A. L., Grant, A. M., & Margolis, J. D. (2012). The bedside manner of homo economicus: How and why priming an economic schema reduces compassion. *Organizational Behavior and Human Decision Processes*, *119*, 27–37.
- Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS ONE*, *7*, e42510. <http://dx.doi.org/10.1371/journal.pone.0042510>
- Pfeffer, J., & DeVoe, S. E. (2009). Economic evaluation: The effect of money and economics on attitudes about volunteering. *Journal of Economic Psychology*, *30*, 500–508. <http://dx.doi.org/10.1016/j.joep.2008.08.006>
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*, 741–763. <http://dx.doi.org/10.1037/0022-3514.67.4.741>
- Roberts, J. A., & Roberts, C. R. (2012). Money matters: Does the symbolic presence of money affect charitable giving and attitudes among adolescents? *Young Consumers*, *13*, 329–336. <http://dx.doi.org/10.1108/17473611211282572>
- Rubin, Z., & Peplau, L. A. (1975). Who believes in a just world? *Journal of Social Issues*, *31*, 65–89. <http://dx.doi.org/10.1111/j.1540-4560.1975.tb00997.x>
- Schmidt, L. C., & Frieze, I. H. (1997). A mediational model of power, affiliation and achievement motives and product involvement. *Journal of Business and Psychology*, *11*, 425–446. <http://dx.doi.org/10.1007/BF02195890>
- 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. <http://dx.doi.org/10.1177/0956797611417632>
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76–80.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*, 534–547. <http://dx.doi.org/10.1037/a0033242>
- Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, *20*, 580–591. <http://dx.doi.org/10.1177/0146167294205014>
- Stanley, D. J., & Spence, J. R. (2014). Expectations for replications are yours realistic? *Perspectives on Psychological Science*, *9*, 305–318. <http://dx.doi.org/10.1177/1745691614528518>
- Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science*, *314*, 1154–1156. <http://dx.doi.org/10.1126/science.1132491>

(Appendices follow)

Appendix A

Sentence Descramble Task

Subjects in Experiments 2 and 3 completed a sentence descrambling task requiring that they create a sentence using 4 of 5 given words. Each set of words is given below. An asterisk indicates a sentence related to money. This task and these stimuli were first used in studies reported by [Vohs, Mead, and Goode \(2006\)](#).

Money Group

1. you held pencil building the
2. *received a raise blue she
3. *I a cashed pen check
4. to she music listened jump
5. metal I wrote letter the
6. *has the capital line he
7. *received they large city profits
8. we later will mountain swim
9. *revenues our rising book are
10. is green the sweater bottom
11. *hundred bill one bottle dollar
12. you coming are here purple
13. camping ten went girls book
14. *won green the I lottery
15. *he wealthy is cup very
16. is hard he win studying
17. *secure I words financially am
18. sky went gray the is
19. *pockets he deep blue has
20. *we cup afford can it
21. again late worked watch we
22. *finances he manages mouse well
23. paper long going was the
24. is outside cold desk it
25. *liberally money she paperclip spends
26. on printer grass she walked
27. *job well pays the arrow

28. took tight he a glass
29. *salary paying high desk a
30. opens he door his top

Control Group

1. you held pencil building the
2. on printer grass she walked
3. took tight he a glass
4. to she music listened jump
5. metal I wrote letter the
6. ski she to wanted many
7. opens he door his top
8. we later will mountain swim
9. is green the sweater bottom
10. you coming are here purple
11. camping ten went girls book
12. is hard he win studying
13. bill the going sent we
14. sky went gray the is
15. meal she the calendar ate
16. again late worked watch we
17. gift he the helping gave
18. paper long going was the
19. is outside cold desk it
20. dishes we washed song the
21. room dark the city is
22. we coffee for went white
23. walked the keyboard dog she
24. exam was the grass challenging
25. up the stadium pick book
26. was fun outside party the
27. is making sun dinner who
28. read she paper the light
29. deep the water number is
30. volume turn the flower up

Appendix B

System Justification Scale

The measure in Experiment 1 was the system justification scale ([Kay & Jost, 2003](#)). Each item was rated from 1 (*strongly disagree*) to 7 (*strongly agree*). Items 3 and 7 are reverse-scored.

1. In general, you find society to be fair.
2. In general, the American political system operates as it should.
3. Americociety needs to be radically restructured.*
4. The United States is the best country in the world to live in.
5. Most policies serve the greater good.
6. Everyone has a fair shot at wealth and happiness.
7. Our society is getting worse every year.*
8. Society is set up so that people usually get what they deserve.

(Appendices continue)

Appendix C

Belief in a Just World Scale

The measure in Experiment 2 was the belief in a just world scale (Rubin & Peplau, 1975). Each item was rated from 0 (*strongly disagree*) to 5 (*strongly agree*). Items 1, 4, 5, 8, 10, 13, 16, 17, and 20 are reverse-scored.

1. I've found that a person rarely deserves the reputation he has.
2. Basically, the world is a just place.
3. People who get "lucky breaks" have usually earned their good fortune.
4. Careful drivers are just as likely to get hurt in traffic accidents as careless ones.
5. It is a common occurrence for a guilty person to get off free in American courts.
6. Students almost always deserve the grade they receive in school.
7. Men who keep in shape have little chance of suffering a heart attack.
8. The political candidate who sticks up for his principles rarely gets elected.
9. It is rare for an innocent man to be wrongly sent to jail.
10. In professional sports, many fouls and infractions never get called by the referee.
11. By and large, people deserve what they get.
12. When parents punish their children, it is almost always for good reason.
13. Good deeds often go unnoticed and unrewarded.
14. Although evil men may hold political power for a while, in the general course of history good wins out.
15. In almost any business or profession, people who do their job well rise to the top.
16. American parents tend to overlook the things most to be admired in their children.
17. It is often impossible for a person to receive a fair trial in the USA.
18. People who meet with misfortune have often brought it on themselves.
19. Crime doesn't pay.
20. Many people suffer through absolutely no fault of their own.

Appendix D

Social Dominance Scale

The measure in Experiment 3 was the social dominance scale (Pratto et al., 1994). Items were rated from 1 (*strongly disagree*) to 7 (*strongly agree*). Items 9 through 16 are reverse-scored.

1. Some groups of people are simply inferior to other groups.
2. In getting what you want, it is sometimes necessary to use force against other groups.
3. It's OK if some groups have more of a chance in life than others.
4. To get ahead in life, it is sometimes necessary to step on other groups.
5. If certain groups stayed in their place, we would have fewer problems.
6. It's probably a good thing that certain groups are at the top and other groups are at the bottom.
7. Inferior groups should stay in their place.
8. Sometimes other groups must be kept in their place.
9. It would be good if groups could be equal.
10. Group equality should be our ideal.
11. All groups should be given an equal chance in life.
12. We should do what we can to equalize conditions for different groups.
13. Increased social equality.
14. We would have fewer problems if we treated people more equally.
15. We should strive to make incomes as equal as possible.
16. No one group should dominate in society.

(Appendices continue)

Appendix E

Fair Market Ideology Scale

The measure in Experiment 4 was the fair market ideology scale (Jost et al., 2003). Items 1 through 15 were rated from -5 (*strongly disagree*) to 5 (*strongly agree*). Items 16 through 25 were rated from -5 (*completely unfair*) to 5 (*completely fair*). Items 3, 4, 6, 8, 13, and 14 are reverse-scored.

1. The free market system is a fair system.
2. Common or “normal” business practices must be fair, or they would not survive.
3. In many markets, there is no such thing as a true “fair” market price.*
4. Ethical businesses are *not* as profitable as unethical businesses.*
5. The most fair economic system is a market system in which everyone is allowed to independently pursue their own economic interests.
6. Acting in response to market forces is *not* always a fair way to conduct business.*
7. The free market system is an efficient system.
8. The free market system has nothing to do with fairness.*
9. Acting in response to market forces is an ethical way to conduct business.
10. In free market systems, people tend to get the outcomes that they deserve.
11. The fairest outcomes result from transactions in which the buyers pay the “fair” market price.
12. Profitable businesses tend to be more morally responsible than unprofitable businesses.
13. Regulated trade is fair trade.*
14. Economic markets do *not* fairly reward people.*
15. Whatever price a buyer and seller agree to trade at is a fair price.
16. When a company raises the prices that it charges its customers for its goods, because management has obtained market research which suggests that its customers are willing to pay more, it is . . .
17. When a professional athlete receives a raise because a raise has been received by another league player of comparable ability, but none of the other team members receive comparable raises, it is . . .
18. The fact that scarce goods tend to cost more in a free market system is . . .
19. When a company downsizes in order to reduce its costs to be more competitive with rival companies, it is . . .
20. When concessions at airports and concerts charge higher prices for beverages because they know that their customers have no alternatives, it is . . .
21. The fact that wealthier people live in bigger homes and better neighborhoods than poorer people who cannot afford to pay the same prices is . . .
22. When a company lays off higher-cost employees in the United States and replaces them with lower wage workers in a foreign country in order to make higher profits, it is . . .
23. The fact that housing prices in Palo Alto, California are 4 to 6 times those for comparable houses in Chicago is . . .
24. The fact that more educated employees tend to earn higher wages than less-educated employees is . . .
25. The fact that some working families can afford to hire more household help than others is . . .

Appendix F

First Two Versions of Experiment 2

We conducted three versions of Experiment 2 in Caruso et al. (2013). The third attempt is reported in this paper as Experiment 2, and our first two attempts are described below.

First Version of Experiment 2

In our first attempt, the Likert scale for the dependent measure (belief in a just world scale, Appendix C) differed from that used

in the original study by Caruso et al. (2013) in two ways. First, the endpoints in our study were inadvertently labeled *unjust* and *just* (rather than *strongly disagree* and *strongly agree*). Second, we used a 6-point scale instead of a 7-point scale because Caruso et al. mistakenly reported that they used a 7-point scale. (After we completed this study and sent the results to Caruso, he informed us that they had in fact used a 6-point scale; personal communication, October 31, 2012.)

(Appendices continue)

We tested 212 adults (mean age = 34.34; 48% women). Subjects were drawn from an online subject pool of the UCSD Learning, Attention, and Perception lab. Subjects in this pool are paid about \$10 per hour. Otherwise, the method was identical to that reported in Experiment 2.

We found no effect, $t(210) = 0.0$, $d = 0$. The money prime group scored 4.35 ($SD = 0.81$), and the control group scored 4.35 ($SD = 0.79$). The mislabeled Likert scale had no discernible effect on the measure's reliability (Cronbach's alpha = 0.85, with 9 of 20 items reverse-scored). We suspect that subjects are so familiar with Likert scales that they did not even read the labels, and instead interpreted the scale as measuring the degree of agreement.

Second Version of Experiment 2

In our next attempt, we used a Likert scale that was identical to the one used in Experiment 2 of Caruso et al. (2013). We tested 191 adults (mean age = 20.6; 70% women). Each was randomly

assigned to either the money condition ($n = 91$) or the control condition ($n = 100$). The sample was drawn from the UCSD psychology subject pool, and each subject first passed an online English proficiency test. Subjects received partial course credit. Otherwise, the method was identical to that used in Experiment 2.

The money prime effect was not statistically significant, $t(189) = 1.70$, $p = .09$, $d = 0.23$. The money prime group scored 2.53 ($SD = 0.43$), and the control group scored 2.43 ($SD = 0.43$). Reliability was good (Cronbach's alpha = 0.90).

We reported this study in an earlier version of this paper, and one reviewer wrote that the sample size ($n = 191$) was too small and should be 2.5 times as large as the sample used in the original study ($n = 168$). We therefore conducted a third version that is reported as Experiment 2 in the main text.

Received August 6, 2013

Revision received January 1, 2015

Accepted January 9, 2015 ■