Direct and Conceptual Replications of Eskine (2013): Organic Food Exposure Has Little to No Effect on Moral Judgments and Prosocial Behavior

Eileen Moery[1] and Robert J. Calin-Jageman[1]

Abstract

Is there a dark side to organic food? Eskine reported that participants exposed to organic food became much more morally judgmental and much less prosocial relative to participants exposed to neutral or comfort foods. This research sparked tremendous media interest, but was based on one experiment with a small sample size. We report three attempts to replicate Eskine using samples conferring high power, preregistered analysis plans, and original materials. Across two direct replications and an online conceptual replication, we found that organic food exposure has little to no effect on moral judgments ($d = 0.06$, 95% confidence interval [CI] [-0.14, 0.26], $N = 377$) and prosocial behavior ($d = 0.03$, 95% CI [-0.17, 0.23], $N = 377$). Mere exposure to organic food is probably not sufficient to substantially change moral behavior.

The global market for organic food is large and rapidly growing (Sahota, 2008). This growth is driven, in part, by consumers' perceptions that organic food is beneficial for the environment and thus a prosocial consumer choice (Aertsens, Mondelaers, Verbeke, Buysse, & Van Huylenbroeck, 2011; Mondelaers, Verbeke, & Van Huylenbroeck, 2009).

While organic food consumption may be beneficial to the environment, recent research has cautioned that this may be offset by deleterious effects on consumer behavior. Specifically, Eskine (2013) conducted a study in which participants were asked to rate the desirability of different organic, neutral, or comfort foods. In an ostensibly unrelated task, participants then rated the acceptability of a set of moral transgressions and completed a measure of prosocial behavior. Participants who had been exposed to organic food were substantially more harsh in their moral judgments relative to those exposed to control ($d = 0.81$, 95% confidence interval [CI] [0.19, 1.45]) and comfort ($d = 1.17$, 95% CI [0.53, 1.84]) foods. In addition, prosocial behavior was substantially reduced in the organic group relative to the control ($d = -0.64$, 95% CI [-1.28, -0.03]) and comfort groups ($d = -1.42$, 95% CI [-2.12, -0.76]). Eskine (2013) interpreted this finding as a manifestation of moral licensing (Monin & Miller, 2001), though this interpretation has been disputed (Blanken, van de Ven, & Zeelenberg, 2015).

The potential for a dark side to organic food captured tremendous media attention when Eskine (2013) was first published. It was covered extensively in national-level media,

including TV (Fox News, 2012), online (Carbone, 2012), and radio (Limbaugh, 2012).

While the national media was immediately ready to take Eskine (2013) as proof that organic food "turns you into a jerk," the empirical evidence for this claim is actually tenuous. First, Eskine (2013) consists of only a single experiment. Second, the sample size used is fairly small ($n = 21$/group), leading to considerable uncertainty about effect sizes (see CIs above). Third, the wider body of evidence about spillover from pro-environmental acts is mixed, with some evidence of negative spillover, but other findings of no effect or even additional benefits (reviewed by Truelove, Carrico, Weber, Raimi, & Vandenbergh, 2014). Finally, it is difficult to reconcile the findings of Eskine (2013) with other similar findings in social psychology. Moral licensing is typically conceptualized as requiring a virtuous action to actually be performed, not just primed through exposure (Blanken et al., 2015). For example, Mazar and Zhong (2010) contrasted mere exposure to "green" products with actual purchase in a mock online store. They found that purchase of green products did lead to moral

---

[1] Department of Psychology, Dominican University, River Forest, IL, USA

Corresponding Author:
Robert J. Calin-Jageman, Department of Psychology, Dominican University, 7900 West Division, River Forest, IL 60305, USA.
Email: rcalinjageman@dom.edu

licensing, but that mere exposure actually *enhanced* prosocial behavior, the opposite of what Eskine (2013) found with organic product exposure.

Taken together, there seems to be good reason to interpret the results of Eskine (2013) with caution. To help clarify this situation, we undertook a series of direct and conceptual replications of Eskine (2013). Throughout, we adopted new best practices to enhance the rigor and interpretability of our replication attempts. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012). For each experiment, our design, materials, and analysis plan were registered using the replication recipe (Brandt et al., 2014) prior to data collection on the Open Science Framework (https://osf.io/atkn7/); all raw data and analysis files have also been posted there.

Study 1a: Direct Replication

We planned and executed a high-fidelity and high-powered direct replication of Eskine (Eskine, 2013). This was made possible through the gracious cooperation of Kendall Eskine (personal communication, May 27, 2014) who provided the original materials from the study.

Method

Sampling Plan

For the moral judgment task, Eskine (2013) found $d = 0.83$ for the comparison between the control and organic groups. Based on this we set a minimum sample size of 32/group, which would provide power of 0.90 for this effect size using $\alpha = .05$ (Dupont & Plummer, 1998). We set a stopping rule of ending in the first week of data collection after achieving our minimum sample size. However, we neglected to specify this stopping rule in our preregistered materials. Nevertheless, we did not analyze collected data to determine when to end this study (or any other study reported in this article).

Participants

We recruited participants from introductory psychology and biology courses. Participants received a credit that could be applied to any course with a research participation requirement.

We collected responses from 124 participants. Of these, 1 did not complete any items after the food ratings, and 1 failed the memory manipulation check (see below), leaving 122 responses for analysis (Table 1). Both of these participants were dropped from all analyses.

Participants were randomly assigned to view images of organic food ($n = 40$), comfort food ($n = 41$), or control foods ($n = 41$). Independent of this group assignment, participants were also randomly assigned to either the all-sixes ($n = 58$) or the some-sixes ($n = 64$) condition of the positive control (see below).

Table 1. Overview of Study Characteristics.

Study	n (Organic/Control/ Comfort)	% Female	Mean Age (Years)
Eskine (2013)	62 (21/20/21)	60	Not reported
Study 1a—Direct	122 (40/41/41)	75	21.6
Study 1b—Bigger	112 (55//57/—)	70	21.7
Study 2—MTurk	284 (89/95/100)	40	30.1

Note. % Female is calculated relative to participants who completed this item; nonresponses, though, were less than 1% of all participants. MTurk = Mechanical Turk.

Table 2. Norming Means and Standard Deviations.

Study	Organic	Control	Comfort
Eskine (2013; $n = 28$)	6.6 (1.2)	4.3 (1.4)	1.5 (1.0)
Study 1a—Direct ($n = 25$)	6.2 (0.9)	4.3 (1.1)	1.5 (0.8)
Study 2—MTurk ($n = 60$)	6.3 (1.0)	4.3 (1.3)	1.5 (1.0)

Note. MTurk = Mechanical Turk.

Norming of Food Materials

To prepare the materials for his experiment, Eskine (2013) conducted a preliminary norming study in which participants rated candidate stimuli on a scale from 1 (*typical organic food*) to 7 (*typical organic food*). We repeated this procedure with 25 students from an entry-level psychology course. We presented the same 12 items (1 per page, printed in color) used in the original study (organic: apple, spinach, tomato, carrot; comfort: ice cream, cookie, chocolate, brownie; control: oatmeal, rice, mustard, beans) plus an additional 12 items drawn from the same stock-art collection with the same background and style (additional organic options: banana, celery, grapes, strawberry; additional comfort options: cupcake, doughnut, pudding, and cinnamon roll; additional control items: bread, hard-boiled egg, mayonnaise, pasta). For the norming study, putative organic foods were not marked with a "USDA Organic" symbol.

Based on the norming data we obtained, we replaced the apple image used in Eskine (2013) with a celery image. No other substitutions were made. This yielded a set of food images for each condition with norming data closely matched in central tendency and variation to the original study (Table 2).

Materials

Food stimuli. Food stimuli were presented as in Eskine (2013), with one stimulus per page in color-printed packets labeled "Study 1." For each stimulus, participants rated desirability on a scale from 1 to 7 (1 = *not at all desirable*, 7 = *extremely desirable*). As in the original study, all images of organic food were marked with a "USDA organic" symbol; control and comfort-food images were not.

Moral dilemmas. The six moral dilemmas (originally from Wheatley & Haidt, 2005) were a graduate student stealing from the library, a congressman accepting bribes, a man shoplifting, a lawyer chasing ambulances, second-cousins consensually engaging in sex, and a man eating his already dead dog. The dilemmas were presented one per page in a separate printed packet labeled ''Study 2.'' Participants rated the morality of each situation on a 7-point scale (1 = *not at all morally wrong* to 7 = *very morally wrong*).

Prosocial measures. Eskine (2013) measured prosocial behavior using a verbal cover story after which participants wrote down how many minutes (out of 30) they would be willing to volunteer toward another research study without compensation. We could not obtain, however, a script for the cover story nor materials related to this item.

We developed a similar item that would avoid the need for verbal intervention. Specifically, we added the following prompt at the end of a page of demographic questions: ''Basic science research provides many benefits to society. Would you be interested in volunteering for additional research projects?'' Participants then marked their interest on a scale from 1 to 7 (1 = *absolutely not,* 4 = *maybe*, and 7 = *definitely*). Then, on the next page, we added a second measure of prosocial behavior with the following open-ended prompt: ''If you indicated an interest in volunteering, leave your student number so that we can contact you when additional studies are available.'' Participants who responded with their student number, phone number, or e-mail were scored as prosocial; all other responses and blank responses were scored as nonprosocial.

Memory manipulation check. We added a memory manipulation check to the end of the study. Specifically, participants were asked, ''What foods did you rate in Study 1?'' with three multiple-choice options listing the names of the foods from each condition. As specified in our preregistered analysis plan, we excluded participants who failed this manipulation check from all analyses (one participant).

Positive control. To help indicate the overall quality of our replication attempt, we included as a positive control an additional experiment with a well-defined effect size. Specifically, we used the retrospective gambler's task (Oppenheimer & Monin, 2009). In this task, participants were asked to imagine entering a casino where they observe a gambler role three dice and obtain either (a) three sixes (the all-sixes condition) or (b) two sixes and a three (the some-sixes condition). After imagining the scenario, participants were asked to estimate how many times the gambler had already rolled the dice (open-ended response). The expected effect is for those who read the all-sixes scenario to estimate more prior rolls than those who read the some-sixes scenario.

We obtained the materials for this task from the Many Labs project (Klein et al., 2014). This positive control was selected because (a) the Many Labs project has recently shown that this effect is highly robust, (b) the expected effect size ($d = 0.61$) is similar to that observed for moral judgments in the target study, and (c) the effect depends critically on participant's reading carefully enough to respond differently to a subtle difference between the two scenarios.

Group assignment to the positive control was made randomly and independent of food condition. As in the Many Labs project, we applied a square root transformation to estimates from this task, but we report raw scores for ease of interpretation.

Procedure

The experiment was administered by one of the three female lab assistants. A script was followed to ensure regularity in the administration of the experiment. Participants completed the study in a classroom typesetting.

As in Eskine's original experiment, participants were told that they would be participating in two different studies administered together for the sake of efficiency. First, packets for ''Study 1'' were passed out, containing the food images and desirability ratings. After students completed the food ratings, packets for ''Study 2'' were passed out, containing the remaining items. Packets for Study 2 contained either the all-sixes scenario or the some-sixes scenario for the positive control.

Analysis

We used the same analysis strategy as Eskine (2013). We also calculated standardized effects sizes, CIs, and integrated effect size estimates. The standardized effect sizes we report are corrected for bias (Hedges, 1981). Effect size CIs and integrated effect size estimates were calculated using Exploratory Software for Confidence Intervals (ESCI) (Cumming, 2011) and Meta-Essentials (Van Rhee, Suurmond, & Hak, 2015) using random effects meta-analysis. Statistics reported in this article were checked for typos using the statcheck package for R (version 1.0.1; Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2015).

Results

In contrast to Eskine (2013), we did not observe a large effect of food exposure on moral judgments, $F(2, 119) = 0.43$, $p = .65$, $\eta^2 = 0.01$. The organic food group did express the most severe moral judgments, but the effect size was small (Table 3). While ratings of food desirability did vary significantly across condition, $F(2, 119) = 21.5$, $p < .001$, $\eta^2 = 0.27$, these ratings were not related to moral judgments and using desirability ratings as a covariate did not reveal a strong impact of food on moral judgments, $F(2, 118) = 0.73$, $p = .48$, $\eta^2 = 0.01$.

We also did not observe a large effect of food condition on prosocial behavior, $F(2, 119) = 0.19$, $p = .83$, $\eta^2 = 0.003$. The organic group was less willing to volunteer for a different study, but the effect size was small (Table 4). Similarly, we found that the proportion of participants who left contact information to volunteer for a study was similar between groups ($P_{organic} = 0.50$, $P_{comfort} = 0.37$, $P_{control} = 0.56$; $\chi^2(2) = 3.27$,

Table 3. Morality Ratings by Study Conditions.

Study	Organic	Control	Comfort	Cohen's d (Organic to Control)	Power (if d = 0.81)
Eskine (2013)	5.58 (0.59)	5.08 (0.62)	4.89 (0.57)	0.81 [0.19, 1.45]	0.71
Study 1a—Direct	5.75 (0.73)	5.65 (0.90)	5.58 (0.86)	0.12 [−0.32, 0.55]	0.95
Study 1b—Bigger	5.81 (0.76)	5.66 (1.02)	—	0.17 [−0.21, 0.54]	0.99
Study 2—MTurk	5.22 (0.83)	5.25 (0.91)	5.29 (0.87)	−0.03 [−0.32, 0.26]	0.99
Overall	—	—	—	0.18 [−0.10, 0.47]	

Note. Cohen's *d* is reported with correction for bias (Hedges, 1981). Overall shows integrated effect size over all studies using a random effect meta-analysis conducted using ESCI (Cumming, 2011). A test for heterogeneity of effect size was not significant: $Q(3) = 5.94$, $p = .11$. MTurk = Mechanical Turk.

Table 4. Prosocial Behavior by Study Condition.

Study	Scale	Organic Food	Control Food	Comfort Food	Cohen's d (Organic to Control)	Power (if d = 0.64)
Eskine (2013)	0–30 min	13.40 (9.38)	19.88 (10.33)	24.55 (5.49)	−0.64 [−1.28, −0.03]	0.51
Study 1a—Direct	1–7 willingness	5.17 (1.41)	5.36 (1.56)	5.31 (1.37)	−0.13 [−0.56, 0.31]	0.81
Study 1b—Bigger	0–30 min	9.53 (12.05)	9.04 (11.04)	—	0.04 [−0.33, 0.41]	0.92
Study 2—MTurk	1–5 willingness	4.36 (0.79)	4.28 (.93)	4.25 (0.96)	0.09 [−0.20, 0.38]	0.99
Overall	—	—	—	—	−0.07 [−0.33, 0.18]	

Note. Cohen's *d* is reported with correction for bias (Hedges, 1981). Measurement scales varied over these studies. Eskine (2013) measured the number of minutes (0–30) participants were willing to commit to an additional study without compensation. Study 1a used a 7-point scale to indicate willingness to volunteer for another study (1 = *absolutely not*, and 7 = *definitely*). Study 1b used the same scale as in Eskine (2013). Study 2 used a 5-point scale to indicate willingness to volunteer for another study (1 = *absolutely not*, and 5 = *definitely*). MTurk = Mechanical Turk.

Table 5. Positive Control: Estimated Number of Roles in Retrospective-Gambler's Fallacy Scenarios.

Study	All-sixes scenario	Some-sixes scenario	Cohen's d
Study 1a—Direct	8.61 (9.31)	4.92 (5.70)	0.49 [0.10, 0.88]
Study 1b—Bigger	34.24 (61.9)	9.96 (24.6)	0.52 [0.14, 0.91]
Study 2—MTurk	28.41 (68.9)	8.91 (18.7)	0.41 [0.17, 0.64]
Overall			0.45 [0.30, 0.60]

Note. Overall row shows integrated effect size using a random effect meta-analysis conducted with Meta-Essentials (Van Rhee et al., 2015). A test for heterogeneity of variance was not significant: $Q(2) = 0.30$, $p = .86$. For Study 1a, results are using a 10% trimmed mean approach due to the presence of a single extreme outlier: $n = 51$ and 55 for the all-six and some-six conditions, respectively. For Study 1b, $n = 58$, 53 and for Study 2, $n = 148$, 136 for the all-sixes and some-sixes conditions, respectively. Cohen's *d* and inferential statistics were calculated using square root transformed responses but the means and standard deviations reported here are raw score responses for ease of interpretation. MTurk = Mechanical Turk.

$p = .20$, Cramer's $V = 0.12$). As expected, responses on these two prosocial measures were positively correlated ($r = .53$, 95% CI [0.39, 0.65], $N = 123$, $p < .001$), indicating reasonable convergent validity.

To help determine the overall quality of the data collected, we examined responses on our positive control, the retrospective gambler's task (Table 5). Four participants did not respond to the scenario (three in the some-sixes group, one in the all-sixes group). In addition, one participant gave an extreme outlier response ($z = 10.8$, estimated 3,000,000 rolls). This extreme outlier disrupted our plan to use a standard filter for outlier responses ($|z| > 3$, as used in the Many Labs project).

Instead, we used a trimmed mean approach and analyzed group differences after dropping the highest 10% of all valid responses (12 participants total; 6 from each condition). Within the remaining 106 responses we found the expected effect, with participants given the all-sixes scenario estimating about 75% more rolls than those given the some-sixes scenario, $t(104) = 2.51$, $p = .01$. This suggests that our study was of sufficient quality to detect a subtle effect of moderate size.

We also tried reanalyzing the main dependent variables within only the 106 participants with valid/nontrimmed responses to the positive control. This did not, however, reveal a strong effect of food exposure on moral judgments, $F(2, 103) = 0.85$, $p = .43$, $\eta^2 = 0.02$, nor on prosocial behavior, $F(2, 103) = 0.74$, $p = .48$, $\eta^2 = 0.01$.

Discussion

Although we replicated Eskine (2013) with high fidelity and power, we observed negligible effects of food exposure on moral judgments and prosocial behavior. Still, the effect size estimates we obtained are fairly broad. In addition, our prosocial item was not perfectly matched to the original study, as it differed in response scale (1–7 rather than 0–30 min), wording, and social context (request made by female undergraduate rather than a male faculty member).

Study 1b: A Bigger and Improved Direct Replication

At the request of reviewers, we conducted a second direct replication with a larger sample size and a prosocial item even more

closely matched to the original study (same scale, wording more similar to original, and request made by male faculty member). To achieve the larger sample size, we dropped the comfort-food condition, as this contrast to control was non-significant in the original study.

Method

Sampling Plan

We set a sample-size goal of at least 52 participants per group to achieve the standard of at least $2.5\times$ original sample size suggested by Simonsohn (2015). We set a stopping rule of ending data collection the first week in which this minimum sample size was obtained.

Participants

Participants were recruited as in Study 1a. We collected responses from 113 students. Of these, one failed the memory manipulation check and was not analyzed further. Of the 112 remaining responses, 55 viewed organic food and 57 viewed control food. Independent of this group assignment, 53 responded to the some-sixes scenario for the positive control, and 58 responded to the all-sixes scenario.

Materials

We even more closely matched our prosocial measure to the one used by Eskine (2013). Specifically, participants were given the following prompt adapted from the original manuscript:

> Another professor from another department is also conducting research and really needs volunteers. If you volunteer you will not receive course credit or compensation for your help. How many minutes would you be willing to volunteer? **Specify 0 to 30 minutes:**

Participants then gave an open-ended response to this prompt. Participants were still asked to leave their contact information as an additional measure of prosocial behavior.

Procedure

To match the social context of the original study, a male faculty member (second author) administered the study. In all other respects, the administration of this study was the same as for Study 1a.

Results

We again failed to observe a large effect of food exposure on moral judgments, $t(110) = 0.89$, $p = .39$. The organic food group did express the most severe moral judgments, but the effect size was again small (Table 3). Adjusting for ratings of food desirability did not alter this conclusion, $F(1, 111) = 1.19$, $p = 0.27$, $\eta^2 = 0.01$.

Prosocial behavior also did not vary strongly by group, $t(110) = 0.26$, $p = .82$, and in fact trended in the opposite direction than in the original study (Table 4). The proportion of participants who left contact information to volunteer was also similar between groups, $P_{organic} = 0.38$, $P_{control} = 0.46$; $\chi^2(1) = 0.64$, $p = .43$, Cramer's $V = 0.08$. Responses on the two prosocial measures were again positively correlated ($r = .74$, 95% CI [0.64, 0.81], $n = 112$, $p < .001$).

We again found the expected effect on the positive control (Table 5), with participants responding to the all-six scenario estimating more roles than those responding to the some-sixes scenario, $t(109) = 2.75$, $p = .007$.

Discussion

Although we replicated Eskine (2013) with even higher fidelity and power, we observed negligible effects of food exposure on moral judgments and prosocial behavior. However, our participant pool could be distinctive in a way that moderated the expected effect.

Study 2: Online Conceptual Replication

To sample a more heterogeneous participant pool, we conducted an online conceptual replication of Eskine (2013) with participants recruited from Amazon's Mechanical Turk (AMT).

Method

Sampling Plan

We set a goal of 89 participants/group (267 overall). This was based on Simonsohn's suggestion (2015) of providing 80% power for the smallest effect size that could have been detected with 33% power in the original study. However, we calculated our sample size for a two-sided test rather than the one-sided used by Simonsohn. We did this to adopt an even more conservative approach to allow for possible attenuation of effect in an online context.

To ensure a high quality of response, we preregistered a number of quality controls (see below). From previous work (Cusack, Vezenkova, Gottschalk, & Calin-Jageman, 2015), we estimated that $\sim 30\%$ of AMT participants would fail these quality controls. We thus set a quota of 350 responses and set a stopping rule of ending the study once this target was reached.

Participants

Participants were recruited via AMT and paid US$0.40 for completing the study. Recruitment was restricted to U.S.-based participants with a lifetime human intelligence task (HIT) approval rate > 90%.

Due to a glitch in setting our recruitment quota, 356 complete responses were collected. Of these, 284 (80%) passed all quality controls—only their data are reported here.

Norming Study

We again conducted a preliminary norming study to ensure that AMT participants perceived the food stimuli as typically organic, comfort, or neither. The stimuli were the same as in the first norming study, but presented via an online survey. We collected 60 valid responses. Based on the ratings made, the control-group image of beans was replaced by an image of a hard-boiled egg. No other substitutions were made. This resulted in a set of food stimuli with norming data very similar in central tendency and variation to the original study (Table 2).

Materials and Procedure

The main survey consisted of an online adaptation of Eskine (2013). First, participants entered a screening survey that filtered out participants from the norming study and provided an instructional manipulation check (Oppenheimer, Meyvis, & Davidenko, 2009; see Supplemental Material). Next, participants were randomly assigned to rate the desirability of organic, comfort, or control foods. One food item and rating scale were presented per screen. Next, under the guise of a second study, the moral dilemmas were presented (1/screen).

After the moral dilemmas, prosocial behavior was measured by asking participants if they would be willing to volunteer for additional studies like this one. Interest was rated on a scale from 1 to 5 (1 = *absolutely not*, and 5 = *definitely*). There was no mechanism for participants to follow-through on the volunteer request (e.g., leaving contact information, etc.), and the lack of payment was not stressed as in the original study. We acknowledge this is a poor approximation of the prosocial measure used by Eskine and did not consider it a key aspect of the replication attempt.

The positive control was presented next (retrospective gambler's task). After this, participants reported basic demographics (gender, age, and ethnicity). Finally, with the assurance that payment was guaranteed, participants were asked to honestly report (a) how familiar they were with the moral dilemmas on a scale from 1 to 4 (1 = *not familiar*, 4 = *very familiar*), (b) if English is the first language, (c) if they were currently living in the United States, and (d) their guess for the hypothesis of the study.

Quality Controls

To help ensure that our AMT sample would represent a high quality of response, we specified a number of quality-control filters in our preregistered analysis plan. These filtered out 72 responses in total, excluding participants who used Internet protocol address outside the United States (4), who failed the instructional manipulation check more than 2 times (7), who classified themselves as nonnative English speakers (4), who were already familiar with the moral dilemmas (31), who took an unusually long or short time to complete the survey (22), who gave an outlier response ($|z| > 4$) on the positive control

(4), or who guessed the Hypothesis (0). See Supplemental Material for more details.

For the manipulation of food exposure, this left 95 participants in the control condition, 100 in the comfort-food condition, and 89 in the organic food condition. For the independently assigned positive control, sample sizes were 148 in the all-sixes condition and 136 in the two-sixes condition. Essentially identical results were obtained analyzing the entire response set.

Results and Discussion

We again found that food exposure has little to no effect on moral judgments, $F(2, 281) = 0.14$, $p = .87$, $\eta^2 = 0.001$. In this case, the organic food group was very slightly *less* judgmental than controls, the opposite of the trend expected (Table 3). Ratings of food desirability did vary by condition, $F(2, 281) = 30.5$, $p < .001$, $\eta^2 = 0.18$, but adjusting for desirability did not reveal a strong effect of food exposure on moral judgments, $F(2, 280) = 0.35$, $p = .71$, $\eta^2 = 0.002$.

We also found that food exposure had little to no impact on our single-item measure of prosocial behavior, $F(2, 280) = 0.39$, $p = .68$, $\eta^2 = 0.003$ (Table 4). Nearly all participants marked either the highest or second-highest level of willingness to participate in additional studies. As the expected effect was a *decrease* in prosocial behavior, this ceiling effect is not extremely problematic. However, given the online context of the study and the payment model for our participants, this measure has dubious validity as a measure of prosocial behavior.

The lack of efficacy for the food exposure variable is probably not due to frank problems of quality or engagement, as we were able to detect the expected effect on our positive control $t(214.08) = 3.53$, $p = .001$ (Table 5).

Discussion

We again failed to detect a strong effect of food exposure on moral judgments. It is clear that AMT workers can repeatedly encounter popular psychological stimuli and that this repeated exposure can greatly attenuate effects found with naive participants (Rand et al., 2014). For our study, however, we consider the risk of this problem as small, as we specifically measured familiarity with the moral dilemmas in a way meant to encourage honest responding. Not only was the level of prior familiarity relatively low, but we also excluded participants who did report nonnaiveté.

General Discussion

Across the studies reported here, organic food had little to no effect on moral judgments ($d = 0.06$, 95% CI [−0.14, 0.26]). We found a similar result comparing the moral judgments of participants who were arriving at or departing from an organic food market ($d = 0.15$, 95% CI [−0.28, 0.59]). The review team for this article required, however, that this field study not

be reported or interpreted in this article because it is quite different from Eskine's original study.

We also observed little to no effect of organic food exposure on prosocial behavior ($d = 0.03$, 95% CI [−0.17, 0.23]). We acknowledge that the prosocial measure in our online study is of dubious validity, but excluding this study does not greatly alter the effect size estimates ($d = -0.03$, 95% CI [−0.31, 0.25]).

Overall, the effect sizes we obtained could not have been reliably detected with the sample size of the original study reasonable power (Simonsohn, 2015). In interpreting these estimates, it is important to keep in mind that we followed Eskine in presenting the organic food stimuli with labels but the other food stimuli without labels. It is ambiguous, then, if these effect size estimates represent changes due to organic food exposure or to label exposure.

Combining our results with Eskine (2013) still suggests little to no effect of organic food on moral reasoning and prosocial behavior (see bottom rows of Tables 3 and 4). The integrated CIs still leave some uncertainty, though, and cannot rule out moderate effects in the predicted direction. On the other hand, these CIs are also consistent with no effect or even weak effects in the opposite direction. We did not detect heterogeneity of effect size in these integrated analyses; this is primarily because the data from Eskine (2013) are consistent with a very wide range of possible effect sizes.

Why did our results diverge so sharply from those of Eskine (2013)? One possibility is that our negative results are wrong and that we have underestimated the true impact of organic food exposure. This could have occurred due to insufficient manipulation of the independent variable. This does not seem likely in our studies, however. We obtained original materials from Eskine (2013). In addition, we used the same type of norming studies to ensure that our study populations perceived these materials as equally prototypical of organic, comfort, and neutral foods. Finally, the memory manipulation check in the direct replications shows participants did attend to the food stimuli. It seems more likely that organic food exposure was manipulated as strongly as in the original study.

Negative results are frequently blamed on researcher error (e.g., Mitchell, 2015). To allay such concerns, we used a positive control, the retrospective gambler's fallacy (Oppenheimer & Monin, 2009). In all cases, we observed the expected pattern of results. Furthermore, while any one replication could fail due to a procedural problem, we consistently observed similar results across a range of experimental contexts. Thus, there seems to be reasonable assurance against substantive researcher error.

Another possibility to consider is that both ours and Eskine's results are valid, but differ due to a strong moderator. For example, our participant pools could have differed in a way that alters how well organic food makes moral identity salient. This also seems unlikely, though. Our two direct replications used a similar student participant pool as Eskine (2013). In addition, we examined a diverse online participant pool, and it seems unlikely that these participants also differed systematically in the same direction along an unforeseen moderator. Although the possibility of a moderator cannot be ruled out, there does not seem to be a strong case for this interpretation.

The final possibility to consider is that our results diverge sharply because Eskine (2013) substantially overestimates the effect of organic food exposure on moral judgments. This, to us, seems the most likely explanation. The small sample size in the original experiment is associated with considerable risk of measurement error. Moreover, this would fit the now well-documented pattern of the "winner's curse" (Young, Ioannidis, & Al-Ubaydli, 2008): the tendency for extreme estimates of effect size to be prominently published, only to be undermined by subsequent research that more accurately estimates the true effect size.

If Eskine (2013) overestimates the effect of organic food exposure on moral behavior, then by how much? A recent meta-analysis of moral licensing suggests an average effect size of 0.31 (Blanken et al., 2015), which is well within the upper bound of the CI we obtained for moral reasoning when integrating our results with Eskine (2013). This comparison must be made with caution, though, as this meta-analysis was based solely on studies in which participants actually completed a prosocial behavior, and excluded studies like Eskine (2013) where mere exposure was manipulated. Furthermore, this meta-analysis indicated substantial upward bias in published studies. It seems, then, that our finding of little to no effect of organic food on moral behavior is currently the most plausible.

Overall, our conclusion is that organic food exposure has much less impact on moral reasoning than found by Eskine, potentially down to no effect at all. It is not clear why the original research reached such a different outcome. There does seem, however, to be a clear lesson here related to research dissemination. The credulous public response that followed the publication of Eskine (2013) indicates the better need for all parties in the dissemination process to more clearly communicate the level of uncertainty associated with a scientific study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material

The online supplement is available at http://spp.sagepub.com/supplemental.

References

Aertsens, J., Mondelaers, K., Verbeke, W., Buysse, J., & Van Huylen-broeck, G. (2011). The influence of subjective and objective knowledge on attitude, motivations and consumption of organic food. *British Food Journal*, *113*, 1353–1378. doi:10.1108/00070701111179988

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*. doi:10.1177/0146167215572134

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., . . . van 't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224. doi:10.1016/j.jesp.2013.10.005

Carbone, N. (2012). *Study: Buying organic food may "Harshen moral judgments."* Retrieved August 6, 2015, from http://newsfeed.time.com/2012/05/21/does-organic-food-turn-you-into-a-jerk/

Cumming, G. (2011). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Cusack, M., Vezenkova, N., Gottschalk, C., & Calin-Jageman, R. J. (2015). Direct and Conceptual Replications of Burgmer & Englich (2012): Power may have little to no effect on motor performance. *PLOS ONE*, *10*, e0140806. doi:10.1371/journal.pone.0140806

Dupont, W. D., & Plummer, W. D. (1998). Power and sample size calculations for studies involving linear regression. *Controlled Clinical Trials*, *19*, 589–601. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9875838

Eskine, K. J. (2013). Wholesome foods and wholesome morals?: Organic foods reduce prosocial behavior and harshen moral judgments. *Social Psychological and Personality Science*, *4*, 251–254. doi:10.1177/1948550612447114

Fox News. (2012). *Organic foods can make some people act more judgmental, study finds*. Retrieved August 6, 2015, from http://www.foxnews.com/health/2012/05/22/organic-foods-can-make-some-people-act-more-judgmental-study-finds/

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*, 107–128. doi:10.3102/10769986006002107

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A "Many Labs" replication project. *Social Psychology*, *45*, 142–152. doi:10.1027/1864-9335/a000178

Limbaugh, R. (2012). Organic food and liberal ideology — The Rush Limbaugh show. Retrieved August 6, 2015, from http://www.rushlimbaugh.com/daily/2012/05/21/organic_food_and_liberal_ideology

Mazar, N., & Zhong, C.-B. (2010). Do green products make us better people? *Psychological Science?: A Journal of the American Psychological Society / APS*, *21*, 494–498. doi:10.1177/0956797610363538

Mitchell, J. (2015). *On the emptiness of failed replications*. Retrieved August 6, 2015, from https://web.archive.org/web/20150604192510/http://wjh.harvard.edu/∼jmitchel/writing/failed_science.htm

Mondelaers, K., Verbeke, W., & Van Huylenbroeck, G. (2009). Importance of health and environment as quality traits in the buying decision of organic products. *British Food Journal*, *111*, 1120–1139. doi:10.1108/00070700910992952

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, *81*, 33–43. doi:10.1037/0022-3514.81.1.33

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2015). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*. doi:10.3758/s13428-015-0664-2

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872. doi:10.1016/j.jesp.2009.03.009

Oppenheimer, D. M., & Monin, B. (2009). The retrospective gambler's fallacy: Unlikely events, constructing the past, and multiple universes. *Judgment and Decision Making*, *4*, 326–334. Retrieved from http://www.decisionsciencenews.com/sjdm/journal.sjdm.org/9609/jdm9609.pdf

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. a., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677. doi:10.1038/ncomms4677

Sahota, A. (2008). The global market for organic food and drink. In H. Willer, M. Yussefi-Menzler, & N. Sorensen (Eds.), *The World of organic agriculture: Statistics and emerging trends 2008* (pp. 53–58). New York, NY: Routledge.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word solution. *SSRN Electronic Journal*, 1–4. doi:10.2139/ssrn.2160588

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*, 559–569. doi:10.1177/0956797614567341

Truelove, H. B., Carrico, a. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, *29*, 127–138. doi:10.1016/j.gloenvcha.2014.09.004

Van Rhee, H. J., Suurmond, R., & Hak, T. (2015). *User manual for meta-essentials: Workbooks for meta-analyses*. Rotterdam, The Netherlands: Erasmus Research Institute of Management. Retrieved from www.erim.eur.nl/research-support/meta-essentials

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*, 780–784. doi:10.1111/j.1467-9280.2005.01614.x

Young, N. S., Ioannidis, J. P. a., & Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Medicine*, *5*, e201. doi:10.1371/journal.pmed.0050201

Author Biographies

Eileen Moery is a recent graduate (with honors) of Dominican University with a double-major in neurobiology and clinical psychology. She is planning on pursuing a PhD in clinical psychology.

Robert J. Calin-Jageman is a professor of psychology, electrophysiologist, and scientific gadfly. His research interests are in the transcriptional mechanisms of long-term memory, computational modeling of neural networks, statistics, and meta-science.