



Research review



Interventions to reduce meat consumption by appealing to animal welfare: Meta-analysis and evidence-based recommendations[☆]

Maya B. Mathur^{a,*}, Jacob Peacock^b, David B. Reichling^c, Janice Nadler^{d,e}, Paul A. Bain^f, Christopher D. Gardner^g, Thomas N. Robinson^h

^a Quantitative Sciences Unit, Stanford University, USA

^b The Humane League Labs, USA

^c Department of Oral and Maxillofacial Surgery, University of California at San Francisco (ret.), USA

^d American Bar Foundation, USA

^e Pritzker School of Law, Northwestern University, USA

^f Countway Library of Medicine, Harvard University, USA

^g Stanford Prevention Research Center, Stanford University, USA

^h Stanford Solutions Science Lab, Department of Pediatrics, Stanford University, USA

ARTICLE INFO

Keywords:

Meta-analysis
Nutrition
Behavior interventions
Meat consumption
Planetary health

ABSTRACT

Reducing meat consumption may improve human health, curb environmental damage, and limit the large-scale suffering of animals raised in factory farms. Most attention to reducing consumption has focused on restructuring environments where foods are chosen or on making health or environmental appeals. However, psychological theory suggests that interventions appealing to animal welfare concerns might operate on distinct, potent pathways. We conducted a systematic review and meta-analysis evaluating the effectiveness of these interventions. We searched eight academic databases and extensively searched grey literature. We meta-analyzed 100 studies assessing interventions designed to reduce meat consumption or purchase by mentioning or portraying farm animals, that measured behavioral or self-reported outcomes related to meat consumption, purchase, or related intentions, and that had a control condition. The interventions consistently reduced meat consumption, purchase, or related intentions at least in the short term with meaningfully large effects (meta-analytic mean risk ratio [RR] = 1.22; 95% CI: [1.13, 1.33]). We estimated that a large majority of population effect sizes (71%; 95% CI: [59%, 80%]) were stronger than RR = 1.1 and that few were in the unintended direction. Via meta-regression, we identified some specific characteristics of studies and interventions that were associated with effect size. Risk-of-bias assessments identified both methodological strengths and limitations of this literature; however, results did not differ meaningfully in sensitivity analyses retaining only studies at the lowest risk of bias. Evidence of publication bias was not apparent. In conclusion, animal welfare interventions preliminarily appear effective in these typically short-term studies of primarily self-reported outcomes. Future research should use direct behavioral outcomes that minimize the potential for social desirability bias and are measured over long-term follow-up.

1. Introduction

Excessive consumption of meat and animal products may be deleterious to human health (with meta-analytic evidence regarding cancer (Crippa et al., 2018; Farvid et al., 2018; Gnagnarella et al., 2018; Larsson

& Wolk, 2006), cardiovascular disease (Cui et al., 2019; Guasch-Ferré et al., 2019; Zhang and Zhang, 2018), metabolic disease (Fretts et al., 2015; Kim & Je, 2018; Pan et al., 2011), obesity (Rouhani et al., 2014), stroke (Kim et al., 2017), and all-cause mortality (Larsson & Orsini, 2013; Wang et al., 2016)); promotes the emergence and spread of

[☆] Citation: Mathur MB, Peacock J, Reichling DB, Nadler J, Bain PA, Gardner CD, Robinson TN (in press). Interventions to reduce meat consumption by appealing to animal welfare: Meta-analysis and evidence-based recommendations. *Appetite*.

* Corresponding author. Quantitative Sciences Unit, 1701 Page Mill Road, Palo Alto, CA, 94304, USA.

E-mail addresses: mmathur@stanford.edu (M.B. Mathur), jpeacock@thehumaneleague.org (J. Peacock), nervefiber@gmail.com (D.B. Reichling), jnadler@law.northwestern.edu (J. Nadler), paul_bain@hms.harvard.edu (P.A. Bain), cgardner@stanford.edu (C.D. Gardner), tom.robinson@stanford.edu (T.N. Robinson).

<https://doi.org/10.1016/j.appet.2021.105277>

Received 6 April 2020; Received in revised form 7 August 2020; Accepted 20 April 2021

Available online 11 May 2021

0195-6663/© 2021 The Author(s).

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

pandemics and antibiotic-resistant pathogens (Bartlett et al., 2013; Di Marco et al., 2020; Marshall & Levy, 2011); is a major source of greenhouse gas emissions, environmental degradation, and biodiversity loss (Machovina et al., 2015; Sakadevan & Nguyen, 2017); and contributes to the preventable suffering and slaughter of approximately 500 to 12,000 animals over the lifetime of each human consuming a diet typical of his or her country (Bonnet et al., 2020; Scherer et al., 2019).¹ Therefore, developing simple, effective interventions to reduce meat consumption could carry widespread societal benefits.

“Nudge” interventions that restructure the physical environment, for example by repositioning meat dishes in cafeterias or making vegetarian options the default, may be effective (Bianchi, Garnett, et al., 2018; Garnett et al., 2019; Hansen et al., 2019), as may direct appeals regarding individual health or the environment (Bianchi, Dorsel, et al., 2018; Jalil et al., 2019). Despite sustained academic interest in developing those types of interventions, there has been much less attention to the potential effectiveness of appeals related to animal welfare (Bianchi, Dorsel, et al., 2018). However, the emerging literature on the psychology of meat consumption suggests that appeals to animal welfare might operate on distinct and powerful psychological pathways (Rothgerber, 2020), suggesting that these appeals merit assessment as a potentially effective component of interventions to reduce meat consumption. We first provide a theoretical review of this psychological literature.

1.1. Psychological theory underlying animal welfare interventions

A number of interventions have used psychologically sophisticated approaches to reducing meat consumption by appealing to or portraying the welfare of animals raised for meat (henceforth “animal welfare interventions”). In general, portraying a desired behavior as aligning with injunctive social norms (what others believe one should do) or descriptive social norms (what others actually do) can effectively shift behaviors, including food choices (Higgs, 2015; Schultz et al., 2007). Many animal welfare interventions have invoked **social norms** (Amiot et al., 2018; Hennessy, 2016; Norris, 2014; Norris and Hannan, 2019; Norris and Roberts, 2016; Reese, 2015), for example by stating: “You can’t help feeling that eliminating meat is becoming unavoidably mainstream, with more and more people choosing to become vegetarians by cutting out red meat, poultry, and seafood from their diets” (Macdonald et al., 2016). Animal welfare interventions have also leveraged the “**identifiable victim effect**”, in which people often experience stronger affective reactions when considering a single, named victim rather than multiple victims or a generic group (Jenni & Loewenstein, 1997). In a classic demonstration of this general effect, subjects made larger real donations to the organization Save the Children after reading about a single named child than when reading about multiple children (Västfjäll et al., 2014). Analogously, many animal welfare interventions describe farm animals with reference to specific, named individuals, such as “Leon” the pig (Bertolaso, 2015) or “Lucy” the chicken (Reese, 2015), and this may indeed be more effective at shifting behavior than providing statistical descriptions of the number of animals affected. Last, many interventions provide concrete **implementation suggestions** for reducing meat consumption, for example by listing plant-based dishes or recipes for breakfasts, lunches, and dinners (Norris and Hannan, 2019). These suggestions may help individuals to form concrete intentions for how they would like to respond in the future when faced with food options, which can be an effective means of shifting food choices (Adriaanse et al., 2011).

In addition to leveraging these three well-known standard components of effective behavioral interventions, animal welfare interventions have the potential to harness the unique social, moral, and affective psychology underlying meat consumption (Loughnan et al., 2014;

Rozin, 1996). For example, ethical concern about factory farming conditions is now a majority stance in several developed countries (Cornish et al., 2016), yet meat consumption remains nearly universal (the “meat paradox”; Bastian and Loughnan (2017)). How does meat-eating behavior survive the resulting cognitive dissonance between people’s ethical views and their actual behavior (Rothgerber, 2020)? There are several explanations. First, most individuals in developed countries do not acquire meat by personally raising animals in intensive factory farm conditions, slaughtering, and preparing them, but rather obtain already processed meat that bears little visual resemblance to the animals from which it came. It is therefore rather easy to implicitly view meat as distinct from animals (Benningstad & Kunst, 2020). This situation is captured well in an episode of *The Simpsons* that has been used as an intervention to reduce meat consumption (Byrd-Bredbenner et al., 2010), in which Homer Simpson chastises his newly vegetarian daughter: “Lisa, get a hold of yourself. This is lamb, not a lamb!” Some interventions operate simply by reminding the subject of the connection between meat and animals by, for example, displaying photographs of meat dishes alongside photographs of the animals from which they came; these **meat-animal reminders** seem to consistently reduce meat consumption (Kunst & Hohle, 2016; Kunst & Haugestad, 2018; Earle et al., 2019; Tian et al., 2016; da Silva, 2016; Lackner, 2019).

Second, the public is poorly informed about animal welfare conditions on factory farms, and individuals often deliberately avoid information about farm animal welfare, even admitting to doing so when asked explicitly (Onwezen and van der Weele, 2016). Presumably the public avoids information because they anticipate that the results may be upsetting (Knight & Barnett, 2008). Thus, interventions that circumvent individuals’ cultivated ignorance by graphically describing or depicting conditions on factory farms may provide a “**moral shock**” that could, for some individuals, lead to dietary change, potentially by triggering cognitive dissonance (Jasper & Poulsen, 1995; Rothgerber, 2020; Wrenn, 2013). In principle, animal welfare interventions might be more effective at prompting such dissonance than interventions appealing instead to individual health or the environment, though this point remains speculative (Rothgerber, 2020). However, the use of graphic depictions is controversial, as they might be ineffective or even detrimental in some contexts (Wrenn, 2013).

Third, even when individuals do consider the animal origins of meat and are informed about conditions on factory farms, they may ascribe little or no sentience to animals raised to produce meat or edible animal products (henceforth “farm animals”), limiting the moral relevance of eating meat (Rothgerber, 2020). Indeed, the Cartesian view that animals are automata that do not experience pain, suffering, or emotions, and hence are outside the sphere of moral concern, was once influential in ethical philosophy, though a modern scientific understanding of animal cognition has essentially eliminated this view from scholarly philosophy (Singer, 1995). In fact, individuals may reduce their attributions of mind and sentience post hoc when faced with the dissonance that could otherwise arise from eating meat: subjects randomly assigned to eat beef subsequently reported that cows are less capable of suffering, and they showed less moral concern, than subjects randomly assigned to eat nuts (Loughnan et al., 2010). Interventions that encourage **mind attribution** to farm animals, for example by asking subjects to imagine the cognitive and affective experiences of a cow, may disarm this dissonance-reduction strategy, thus reducing willingness to eat meat (Amiot et al., 2018). Similarly, other interventions leverage the fact that most people already recognize the sentience of companion animals, such as dogs and cats, and therefore incorporate these animals into their spheres of moral concern, even while excluding farm animals with comparable cognitive abilities and capacity for suffering (Rothgerber, 2020). Interventions targeting this form of dissonance highlight the **moral equivalence of farm animals and companion animals**, for example by stating: “If the anti-cruelty laws that protect pets were applied to farmed animals, many of the nation’s most routine farming practices would be illegal in all 50 states.” (Norris and Hannan, 2019).

¹ These are the minimum and maximum taken across 37 countries (Scherer et al., 2019).

Additionally, **physical disgust and moral disgust** are closely intertwined. Experiencing physical disgust can amplify negative moral judgments, even when the two sources of disgust are conceptually unrelated (e.g., viewing physically disgusting video clips versus judging cheating behavior); conversely, experiencing moral disgust can induce physical disgust (Chapman & Anderson, 2013). Given the powerful impact of physical disgust on food choices (Rozin and Fallon, 1980), evoking moral disgust regarding animal welfare may be a particularly potent means of shaping food choices (Feinberg et al., 2019). Indeed, many animal welfare interventions contain graphic verbal or visual depictions of conditions in factory farms that may themselves be physically disgusting. For example, one intervention (Cordts et al., 2014) describes “crowded conditions [and] pens covered in excrement and germs”; a leaflet that has been studied repeatedly describes sows with “deep, infected sores and scrapes from constantly rubbing against the [gestation crate] bars” and “decomposing corpses [found] in cages with live birds” (Vegan Outreach, 2018). We speculate that such interventions might increase moral disgust by triggering physical disgust. Although we are not aware of studies that directly assess this hypothesis using physically disgusting interventions, it is interesting that even interventions that are not obviously physically disgusting, such as the meat-animal reminders described above, seem to operate in part by increasing physical disgust (Earle et al., 2019; Kunst & Hohle, 2016).

Finally, the connection of animal welfare interventions to an existing **social movement** encouraging greater consideration of the welfare of farm animals (Singer, 1995) may further trigger “process motivations” for participation (Robinson, 2010, 2017, chap. 99), such that the process of participating in the social movement (e.g., reducing meat consumption due to ethical concerns) may itself be motivating. Participating in social movements can be intrinsically motivating because they provide opportunities for identity development, social interaction and support, perceived belonging, and activities that boost participants’ perceptions of collective efficacy (Bandura, 2002, chap. 6). That animal welfare interventions are related to a broader social movement may additionally trigger group- or societal-level changes (e.g., increasing public attention, shifting norms regarding meat consumption, or decreasing availability of meat) that may alter the social and physical environments to make it easier to sustain the new behaviors (Robinson, 2010, 2017, chap. 99). Whereas behavior-change appeals emphasizing individual health sometimes suffer from high recidivism (Grattan & Connolly-Schoonen, 2012; Robinson, 2010), interventions that instead link behavior to ethical values, self-identity, and existing social movements may be especially potent and long-lasting (Robinson, 2010, 2017, chap. 99; Walton, 2014). Such interventions have, for example, successfully reduced childhood obesity-related behaviors and risk factors for cardiovascular disease and diabetes by appealing to cultural and ethical values in order to increase physical activity, rather than by appealing directly to obesity reduction or other health motivations (Robinson et al., 2003, 2010; Weintraub et al., 2008). Animal welfare interventions might operate similarly.

In theory, then, animal welfare interventions have the potential to be particularly effective by harnessing:

- (1) multiple general mechanisms of effective behavior interventions (by leveraging social norms and the identifiable victim effect, and by giving implementation suggestions);
- (2) the unique psychology of meat-eating (by invoking meat-animal reminders, moral shocks triggered by graphic depictions of factory farms, mind attribution, the moral equivalence of farm animals and companion animals, and the physical-moral disgust connection); and
- (3) the psychological and practical advantages of connection to a social movement.

But do these interventions work in practice? We now turn to the present empirical assessment of this question.

1.2. Objectives of this meta-analysis

We conducted a systematic review and meta-analysis to address the primary research question: “How effective are animal welfare appeals at reducing consumption of meat?”, in which we define “meat” as any edible animal flesh (Section 3.3). We additionally investigated whether interventions’ effectiveness differed systematically based on their content, such as types of dietary recommendations made, use of verbal, visual, and/or graphic content,² or on study characteristics such as length of follow-up and percentage of male subjects (Rozin et al., 2012) (Section 3.5). We evaluated risks of bias in each study (Sections 3.2.4) and conducted numerous sensitivity analyses (Section 3.4). Finally, considering the evidence holistically in light of its methodological strengths and limitations, we discuss what the evidence suggests about the effectiveness of animal welfare interventions (Section 4.1) and give specific recommendations for future research to advance the field (Section 4.3).

2. Methods

See the Reproducibility section for information on the publicly available dataset, code, and materials. All methods and statistical analysis plans were preregistered in detail (<https://osf.io/d3y56/registrations>) and subsequently published as a protocol paper (Mathur et al., 2020). In the Supplement, we describe and justify some deviations from this protocol.

2.1. Systematic search

Our inclusion criteria were as follows. Studies could recruit from any human population, including from online crowdsourcing websites. Studies needed to assess at least one intervention that was intended to reduce meat consumption or purchase, and interventions needed to include any explicit or implicit mention or portrayal of farm animals, their suffering, their slaughter, or their welfare. Composite interventions, defined as those including both an animal welfare appeal and some other form of appeal (e.g., environmental), were included, though we conducted sensitivity analyses excluding such interventions. Studies needed to include a control group, condition, or time period not subjected to any form of intervention intended to change meat consumption. Thus, studies making within-subject comparisons were eligible. Last, studies needed to report an outcome regarding the consumption or purchase of meat or all edible animal products, as assessed by a direct behavioral measure (e.g., the amount of meat that subjects self-served at a buffet), self-reported behavior (e.g., reported meat consumption over the week following exposure to the intervention), or a self-report of intended future behavior (e.g., intended meat consumption over the upcoming week). Although our focus was primarily on meat consumption, we included studies with outcomes related to consumption of all animal products because we anticipated that many interventions designed to reduce meat consumption would in fact make broader recommendations to reduce all animal product consumption and therefore would be assessed using correspondingly broad outcome measures. Further details on inclusion criteria were published previously (Mathur et al., 2020).

In addition to the nascent academic literature, evidence-based nonprofits have conducted numerous studies that have been reported in a separate body of grey literature. We therefore developed sensitive search strategies targeting both the academic literature and the grey literature to identify eligible articles, which could be published or released at any time and written in any language. For non-English articles, we used automated translation to determine eligibility. To search

² We use “graphic content” to refer to explicit, detailed verbal or visual descriptions of conditions on factory farms.

the academic literature, including unpublished dissertations and theses, we collaborated with an academic reference librarian (PAB) to design detailed search strings for each of eight databases (Medline, Embase, Web of Science, PsycInfo, CAB Abstracts, Sociological Abstracts, ProQuest Dissertations & Theses, and PolicyFile). To search the grey literature and help capture potentially missed academic articles, an author (JP) who is the director of an evidence-based animal welfare research organization designed a three-stage search strategy: (1) we screened The Humane League Labs' existing internal compilation of relevant literature, including both academic studies and grey literature; (2) we screened the websites of 24 relevant nonprofits; and (3) we posted a bibliography of literature identified in the first two stages on relevant forums in the animal advocacy research community to solicit additional leads to studies.³ All methods are detailed in the Supplement. We conducted the final searches on January 17, 2020 for the academic literature and November 20, 2019 for the grey literature.

We reviewed articles and managed data using the software applications Covidence (Covidence Development Team, 2019) and Microsoft Excel. Each article retrieved from an academic database first underwent title/abstract screening, conducted independently by at least two reviewers from among DBR, JN, and MBM. In this stage, reviewers excluded only articles that clearly failed the inclusion criteria. Articles receiving an "include" vote from either or both reviewers proceeded to a full-text screening, during which at least two reviewers independently assessed inclusion criteria in detail. We resolved 9 conflicts between reviewers through discussion or adjudication by other authors. Interrater reliability for inclusion decisions was greater than⁴ Cohen's $\kappa = 0.71$. Grey literature articles that JP identified as potentially eligible underwent the same full-text dual review as academic articles. We refer to articles ultimately judged to meet all inclusion criteria as "eligible".

2.2. Data extraction

Basic descriptive characteristics of studies. For each study, we extracted 33 basic descriptive characteristics regarding, for example, the subject population, location, intervention, and outcome. All extracted characteristics are enumerated in the Supplement. One of the extracted variables indicated whether the study was borderline with respect to the inclusion criteria, a classification made through discussion amongst the review team. For example, one "meat-animal reminder" intervention consisting of a photograph of a pork roast with the head still attached (versus a headless roast in the control condition) was classified as borderline because it was difficult to judge whether the intervention was about animal welfare. As described in Section 2.3, we conducted sensitivity analyses excluding borderline studies.

Hypothesized effect modifiers. Also among the extracted variables

³ Because a previous systematic review (Bianchi, Dorsel, et al., 2018) suggested a paucity of academic study on animal welfare appeals to reduce meat consumption, we did not try to identify academic counterparts to these forums. However, participants in the animal advocacy research forums and nonprofits we contacted include numerous academics and researchers with extensive academic training who follow the relevant academic literature.

⁴ Because the three raters rated only partly overlapping subsets of studies, we calculated κ for each of the three pairs of raters and report the mean κ across the three pairs of raters, weighted by the number of articles they both rated. We could only estimate a lower bound on each pairwise κ because a large number of apparent "conflicts" in fact reflected reviewers' different handling of duplicate articles. That is, we initially instructed reviewers to vote "exclude" on duplicate articles, but we then became aware that a large number of articles reporting on duplicated samples provided additional information that would be useful in extracting study characteristics. (This was especially common among the unpublished studies.) We then instructed reviewers to instead vote "include" on all such articles, which created some spurious "conflicts" that were not clearly differentiated by our data extraction software. The 9 conflicts we report above represent only the real conflicts not arising from duplicate articles.

were 13 hypothesized effect modifiers.⁵ We categorize the effect modifiers into 8 "**coarse**" **intervention characteristics**, which we could code for nearly all studies, versus 5 "**fine-grained**" **intervention characteristics**, which we could code for only the $k = 80$ studies for which we had access to all intervention materials or to detailed descriptions of their content. The coarse intervention characteristics were: (1) whether the intervention contained text; (2) whether the intervention contained visuals; (3) whether the intervention contained graphic visual or verbal depictions of factory farm conditions; (4) the nature of the recommendation made in the intervention ("go vegan", "go vegetarian", "reduce consumption", a mixed recommendation [e.g., recommendations to either reduce or eliminate meat consumption], or no recommendation); (5) lasted longer than 5 min; (6) whether the outcome was measured 7 or more days after the intervention (to capture possible decays in intervention effects over time); (7) the percentage of male subjects in the study (because meat-eating may be closely intertwined with masculine identity in Western cultures (Rothgerber, 2020; Rozin et al., 2012; Ruby, 2012); and (8) subjects' mean or median age.⁶

The fine-grained intervention characteristics described whether the intervention: (1) used mind attribution by describing a farm animal's inner states; (2) described social norms in favor of reduced meat consumption; (3) used the identifiable victim effect by giving a proper-noun name to a farm animal (4) described or depicted pets (i.e., companion animals that typically live in people's houses); with or without explicitly comparing them to farm animals; and (5) gave implementation suggestions in the form of describing or depicting a specific plant-based meal, restaurant dish, or recipe.

Risk-of-bias characteristics. We also extracted 9 characteristics related to the study's risks of bias, including design characteristics (e.g., randomization), missing data, analytic reproducibility and preregistration practices, exchangeability of the intervention and control conditions (i.e., avoidance of statistical confounding through randomization and minimization of differential dropout), avoidance of social desirability bias (e.g., by keeping subjects naïve to the purpose of the intervention), and external generalizability (i.e., the extent to which results are likely to apply to subjects in the general population, who may not already be particularly motivated to reduce meat consumption or purchase). Details on the development, contents, and use of our risk-of-bias tool are provided in the Supplement.

Quantitative data extraction. We extracted quantitative data for the meta-analysis as follows. For each eligible article that reported sufficient statistical information, the statistician (MBM) used either custom-written R code or the R package `metafor` (Viechtbauer, 2010) to extract the point estimate(s) and variance estimate(s) most closely approximating the treatment effect of the animal welfare appeal on the risk ratio (RR) scale (see Supplement for details). We synchronized the directions of all studies' estimates such that risk ratios greater than 1 represented reductions in meat consumption or purchase. When relevant statistics were not reported, we hand-calculated them from available statistics, plots, or publicly available datasets as feasible, or made repeated attempts to contact study authors. Articles reporting multiple point estimates on separate subject samples (even those sharing a control group) could contribute all of these point estimates to the analyses; further details on how we handled articles reporting on multiple eligible interventions were published previously (Mathur et al., 2020). For

⁵ Our preregistration and protocol paper (Mathur et al., 2020) gave non-exhaustive examples of candidate effect modifiers we might investigate, but did not formally preregister this list.

⁶ It would have been informative to also investigate possible effect modification by geographic location. However, given the high rate of missing data on this variable and its sparse distribution across countries (Table 1), the meta-regression model described below was not statistically estimable with the inclusion of this covariate, even when location was coarsely coded as "North America" versus "elsewhere".

studies that reported multiple outcomes (e.g., consumption or purchase of all meat, of all edible animal products, and of specific subsets of meats), we used the outcome most closely matching the intended scope of the intervention. For example, we used the outcome “pork consumption” for an intervention specifically targeting pork consumption (Anderson, 2017) and used the outcome “all animal product consumption” for an intervention that recommended going vegan (Bertolaso, 2015).

Although meta-analysts sometimes extract all eligible outcomes from each paper and average them within studies (Sutton et al., 2000), we correctly anticipated that most studies with multiple eligible food outcomes would be those whose primary outcome was a composite measure of total meat or total animal product consumption and whose other eligible food outcomes were subscales of this composite, representing consumption of specific meats and animal products. Averaging these estimates within a given study would therefore yield a point estimate that would be essentially equivalent to the composite itself, but with the additional limitation that its variance would not be estimable without information on the full correlation structure of the various subscales.

When possible, to reduce the possibility of our choosing which estimates to include in a biased manner, we made all decisions about how we would calculate each study’s point estimate, and for which outcome, according to these and other detailed rules (see Supplement) before we actually calculated the estimate. However, doing so was not always feasible (e.g., for studies that directly reported the point estimates we needed to calculate).

2.3. Statistical analysis methods

Main analyses. The statistician (MBM) conducted all statistical analyses in R (R Core Team, 2020b).⁷ To accommodate potential correlation between point estimates contributed by the same article and to obviate distributional assumptions, we fit a robust semiparametric meta-analysis model using the R package `robmeta` (Fisher & Tipton, 2015; Hedges et al., 2010). For the primary analysis, we included all eligible and borderline-eligible studies and estimated the average effect size with a 95% confidence interval and *p*-value, the standard deviation of the true population effects ($\hat{\tau}$), and metrics that characterize evidence strength when effects are heterogeneous. Regarding the latter, we estimated the percentage of true population effects⁸ that were stronger than various effect-size thresholds representing all effects in the beneficial direction (i.e., $RR > 1$) and more stringently representing only effects that

⁷ We used the following R packages: Canty and Ripley (2020); Ooi et al. (2020); Kuhn et al. (2020); Wickham et al. (2020); Wickham (2016); Rodriguez-Sanchez (2018); Wolak et al. (2012); Gamer et al. (2019); Viechtbauer (2010); Mathur et al. (2019); Wickham (2011a); Mathur and VanderWeele (2020a); Neuwirth (2014); Wickham and Bryan (2019); Fisher and Tipton (2015); Yoshida (2020); Wickham (2011b); Wickham et al. (2019); Coburn and Vevea (2019); Dahl et al. (2019); Fox and Weisberg (2019); Dowle and Srinivasan (2019); R Core Team (2020a); Bates et al. (2015); Bulow (2016); Zeileis (2004); Lüdtke (2020); Wickham (2019, 2011b); Wickham and Henry (2020); Wickham et al. (2019); Tierney et al. (2020).

⁸ We use the term “true population effects” to refer to population parameters, rather than to point estimates with statistical error. “True” does not indicate that these effect sizes are free from bias due, for example, to attrition or social desirability bias.

might be considered to be meaningfully large by two different criteria (i.e., $RR > 1.1$ or $RR > 1.2$) (Mathur and VanderWeele, 2019, 2020b).⁹

We also estimated the percentage of true population effects with risk ratios less than 1 and alternatively less than 0.90, representing unintended *detrimental* effects of the interventions (Mathur and VanderWeele, 2019, 2020b). Taken together, these percentage metrics can help identify if: (1) there are few meaningfully large effects despite a “statistically significant” meta-analytic mean; (2) there are some large effects despite an apparently null meta-analytic mean; or (3) strong effects in the direction opposite of the meta-analytic mean also regularly occur (Mathur and VanderWeele, 2019, 2020b). As a hypothesis-generating method to help identify the individual interventions that appear most effective, we estimated the true population effect size in each study using nonparametric calibrated estimates (Wang and Lee, 2019). Intuitively, the calibrated estimates account for differences in precision between studies by shrinking each point estimate toward the meta-analytic mean, such that the least precise studies receive the strongest shrinkage toward the meta-analytic mean.

Publication bias. We assessed publication bias using selection model methods (Vevea & Hedges, 1995), sensitivity analysis methods (Mathur and VanderWeele, 2020c), and the significance funnel plot (Mathur and VanderWeele, 2020c). These methods assume that the publication process favors “statistically significant” (i.e., $p < 0.05$) and positive results over “nonsignificant” or negative results, a typically reasonable assumption that also conforms well to empirical evidence on how publication bias operates in practice (Jin et al., 2015; Mathur and VanderWeele, 2020c). We used the sensitivity analysis methods to estimate the meta-analytic mean under hypothetical worst-case publication bias (i.e., if “statistically significant” positive results were infinitely more likely to be published than “nonsignificant” or negative results). These methods, unlike the selection model, also accommodated the point estimates’ non-independence within articles and did not make any distributional assumptions (Mathur and VanderWeele, 2020c).

Other sensitivity analyses. As further sensitivity analyses, we conducted meta-analyses within 9 separate subsets of studies, namely: (1) excluding studies with borderline eligibility (leaving $k = 91$ analyzed studies); (2) excluding studies with composite interventions ($k = 52$ analyzed); (3) excluding studies that measured intended behavior rather than directly measured or self-reported behavior ($k = 43$ analyzed); (4) including only studies judged to be at the lowest risk of bias as described above ($k = 12$); (5) including only randomized studies ($k = 75$); (6) including only studies that were preregistered and had open data ($k = 21$); (7) including only published studies ($k = 17$); (8) including only unpublished studies ($k = 83$); and (9) excluding one study (FIAPO, 2017) with a very large point estimate and also a very large standard error, visible in Fig. 1 ($k = 99$ analyzed). The first 2 subset analyses were preregistered; the subsequent ones were introduced post hoc. We also conducted a simple post hoc analysis that assessed the sensitivity of our results to potential social desirability bias, conceptualized as bias in which subjects in the intervention group under-report meat and animal product consumption more than subjects in the control group (VanderWeele & Li, 2019). We ultimately did not conduct a preregistered sensitivity analysis in which we would have excluded interventions targeting consumption of only a specific type of meat; this proved

⁹ As previously recommended (Hill et al., 2008), we chose these thresholds with reference to the effect sizes of other health behavior interventions; for example, general nutritional “nudges” (not specific to meat consumption or animal welfare) produce average effect sizes (Arno & Thomas, 2016) of approximately $RR = 1.15$, and graphic warnings on cigarette boxes increase short-term intentions to quit by approximately $RR = 1.14$ upon conversion from the odds ratio scale (Brewer et al., 2016; VanderWeele, 2017, 2019). These effect sizes are close to our two chosen thresholds. Nevertheless, these thresholds are somewhat arbitrary; therefore, Fig. S3 shows the percentage of effects as a continuous function of the threshold chosen.

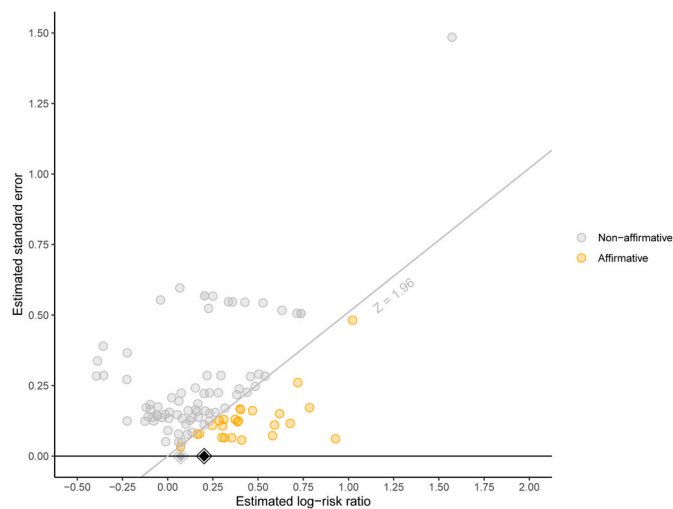


Fig. 1. Significance funnel plot displaying studies' point estimates versus their estimated standard errors. Orange points: affirmative studies ($p < 0.05$ and a positive point estimate). Grey points: nonaffirmative studies ($p \geq 0.05$ or a negative point estimate). Diagonal grey line: the standard threshold of "statistical significance" for positive point estimates; studies lying on the line have exactly $p = 0.05$. Black diamond: main-analysis point estimate within all studies; grey diamond: worst-case point estimate within only the nonaffirmative studies. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

infeasible because of the small number of interventions that unambiguously did so.

Meta-regression on hypothesized effect modifiers. As a secondary analysis, we used robust meta-regression (Hedges et al., 2010) to investigate intervention and study characteristics associated with increased or decreased effectiveness. The meta-regressive covariates were the hypothesized effect modifiers listed in Section 2.2. As described there, we could code the 5 fine-grained intervention characteristics (namely, use of mind attribution, social norms, the identifiable victim effect, depictions of pets, and implementation suggestions) for 80 studies. Additionally, because many of these components were used only in interventions that contained text, these 5 covariates were highly collinear with the covariate indicating that the intervention contained text. We therefore fit two meta-regression models: (1) a "coarse" model that included only the coarse intervention characteristics listed in Section 2.2 ($k = 86$ studies with complete data on the effect modifiers); and (2) a "fine-grained" model containing all fine-grained and the coarse intervention characteristics except the presence of text ($k = 80$ studies). For each model, we included the covariates in the meta-regression simultaneously rather than in separate univariable models because they were fairly highly correlated (Supplementary Table S5).

Metrics of societal impact. We had preregistered an additional type of secondary analysis in which we intended to express intervention effectiveness using metrics that more directly characterize societal impact, such as the estimated reduction in human all-cause mortality events, in the number of animals raised for consumption, and in greenhouse gas emissions. However, limitations in outcome measurement and follow-up duration in the meta-analyzed studies ultimately precluded these analyses; we return to this issue in Section 4.3.

3. Results

3.1. Results of search process

Supplementary Fig. S1 is a PRISMA flowchart depicting the search results. After removing duplicate articles, we screened 4139 articles from the academic database searches and an additional 81 articles

identified from manual searches of the grey literature and the academic literature. We assessed 144 full-text articles for eligibility. In full-text screening, we removed 96 articles that failed inclusion criteria, that reported insufficient information to determine eligibility and for which we could not obtain full texts, or that analyzed the same subject sample as an existing article (e.g., because they were re-analyses of an existing dataset). We thus found that 48 articles met the inclusion criteria. Of these, 7 could not be meta-analyzed because we could not obtain the relevant point estimates from publicly available data, from the paper, or from repeated attempts to contact the authors. (Characteristics of these eligible, but not analyzed, studies are described qualitatively in Supplementary Table S3.)

Of the 41 articles with available statistics, 7 met our preregistered inclusion criteria but were excluded from the main analysis because they reported on a specific, unanticipated study design that is at very high risk of bias. These studies assessed the effectiveness of "challenges" or "pledges" in which, for example, subjects try to follow a vegan diet for one month. These studies shared two key design features: (1) subjects were not specifically recruited, but rather were individuals who had chosen to sign up for the challenge and who therefore may have been already highly motivated to reduce their consumption; and (2) there was no separate control group, but rather the challenge's effectiveness was assessed by comparing within-subject changes in consumption after versus before the challenge. The combination of these two features is highly problematic: subjects with a strong pre-existing motivation to reduce consumption may have done so regardless of whether they took the challenge, and the lack of a separate control group precludes estimation of a valid treatment effect of the challenge even among this subset of highly motivated subjects. These studies also typically provided very limited statistical summaries; in many cases, we would not have had enough information to calculate an effect size that was statistically comparable to the risk ratios we estimated for other studies. For these reasons, we made a post hoc decision during data extraction to exclude studies with both of these features (subject self-selection and lack of a separate control group) from the main analysis. However, in keeping with our preregistered inclusion criteria, we report results that include these "self-selected within-subjects" (SSWS) studies in Section 3.6. Therefore, after excluding the 7 articles reporting on SSWS studies, our main analyses included 34 articles, totaling 100 point estimates.

3.2. Characteristics of included articles and studies

Table 1 summarizes studies' basic characteristics. We meta-analyzed data from 34 articles; 24% were published in peer-reviewed journals, and all others were dissertations, theses, conference proceedings, or reports by nonprofits. The chronologically first study (Byrd-Bredbenner et al., 2010) was published in 2010. The articles contributed a total of 100 point estimates, which were estimated using a total of 24,817 subjects. We will refer to these 100 estimates as "studies" to distinguish them from "articles". We obtained 53% of statistical estimates from publicly available datasets; 28% by manually calculating statistics from information reported in the articles' text, tables, or figures; and the remaining 19% from datasets provided by the study's authors.

Many studies (53%) reported multiple food outcomes that were potentially eligible for inclusion; as described in Section 2.2, we extracted statistics for only the outcome most closely matching the scope of the intervention. For all but one article, the outcome we extracted (e.g., consumption of all meats) was a composite of some or all of the additional food outcomes (e.g., consumption of specific categories of meats), as we had anticipated. For the remaining article (Caldwell, 2016), the primary and additional outcomes were pork consumption and egg consumption respectively, and the latter was measured because the study also contained an ineligible intervention regarding egg production.

Given the large number of studies, we detail their individual characteristics, along with risks of bias ratings, point estimates, variance

Table 1

Basic characteristics of meta-analyzed studies and their interventions. Binary and categorical variables are reported as “count (percentage)”. Continuous variables are reported as “median (first quartile, third quartile)”. ^a: There was a high proportion of missing data regarding studies’ countries because we coded country as missing for online studies (e.g., conducted on Amazon Mechanical Turk) that did not specifically state whether they used geographical restrictions when recruiting subjects.

Characteristic	Number of studies (%) or median (Q1, Q3)
Country	
Canada	1 (1%)
China	2 (2%)
Czech Republic	2 (2%)
Ecuador	1 (1%)
France	3 (3%)
Germany	1 (1%)
India	3 (3%)
Netherlands	1 (1%)
Portugal	2 (2%)
Scotland	2 (2%)
USA	37 (37%)
USA, Canada	2 (2%)
Not reported ^a	43 (43%)
Percent male subjects	43 (36.1, 51.7)
Not reported	4 (4%)
Average age (years)	33.7 (21.9, 35.4)
Not reported	9 (9%)
Student recruitment	
Not undergraduates	77 (77%)
General undergraduates	10 (10%)
Social sciences undergraduates	3 (3%)
Mixed	10 (10%)
Intervention had text	
Yes	83 (83%)
No	14 (14%)
Not reported	3 (3%)
Intervention had visuals	
Yes	62 (62%)
No	36 (36%)
Not reported	2 (2%)
Intervention had graphic content	
Yes	61 (61%)
No	37 (37%)
Not reported	2 (2%)
Intervention used mind attribution	
Yes	65 (65%)
No	28 (28%)
Not reported	7 (7%)
Intervention used social norms	
Yes	29 (29%)
No	62 (62%)
Not reported	9 (9%)
Intervention identified a named victim	
Yes	21 (21%)
No	70 (70%)
Not reported	9 (9%)
Intervention depicted pets	
Yes	32 (32%)
No	60 (60%)
Not reported	8 (8%)
Intervention gave implementation suggestions	
Yes	29 (29%)
No	61 (61%)
Not reported	10 (10%)
Intervention described animal welfare only	
Yes	52 (52%)
No	44 (44%)
Not reported	4 (4%)
Intervention was personally tailored	
Yes	2 (2%)
No	98 (98%)
Not reported	0 (0%)
Intervention’s recommendation	
No recommendation	43 (43%)
"Reduce consumption"	13 (13%)
"Go vegetarian"	9 (9%)
"Go vegan"	7 (7%)

Table 1 (continued)

Characteristic	Number of studies (%) or median (Q1, Q3)
Mixed recommendation	24 (24%)
Not reported	4 (4%)
Intervention’s duration (minutes)	1.5 (1, 5.88)
Not reported	2 (2%)
Outcome category	
Consumption	96 (96%)
Purchase	4 (4%)
Length of follow-up (days)	0 (0, 32.5)
Not reported	1 (1%)
Source of statistics	
Data from author	19 (19%)
Paper	28 (28%)
Public data	53 (53%)

estimates, and additional food outcomes, in a publicly available spreadsheet (<https://osf.io/8zsw7/>; see Supplement for codebook).

3.2.1. Subjects

The median analyzed sample size per article was 522 subjects (25th percentile: 235; 75th percentile: 802). (We report the sample sizes by article rather than by study to avoid double-counting control subjects whose data contributed to multiple point estimates from the same article.) Studies were conducted in at least 11 countries (Canada, China, the Czech Republic, Ecuador, France, Germany, India, the Netherlands, Portugal, Scotland, and the United States) that represented a diverse range of geographical regions as well as cultural characteristics that may be relevant to meat consumption, including affluence, traditional cuisines, and religious and ethical traditions. The median percentage of male subjects was 43%, and the median of studies’ average ages was 33.7 years. 13% of studies recruited only undergraduates, including 2% (2 studies) that specifically recruited social sciences undergraduates. 77% did not specifically recruit undergraduates, and the remaining 10% recruited both undergraduate and non-undergraduate samples.

3.2.2. Interventions

Examples of typical interventions include, not exhaustively, providing informational leaflets about factory farming conditions (Norris and Hannan, 2019), showing photographs of meat dishes accompanied by photographs of the animals they came from (Kunst & Hohle, 2016), and providing mock newspaper articles (Macdonald et al., 2016). Details on the many other interventions represented in our meta-analysis are provided in the aforementioned publicly available spreadsheet of study characteristics. Below, we report on studies’ characteristics using percentages that count missing data as its own category, so percentages may add up to less than 100%. Most studies’ interventions contained text (83%), contained visuals (e.g., photographs, infographics, or videos) (62%), contained graphic verbal or visual depictions of welfare conditions in factory farms (61%), and used mind attribution (65%). Relatively fewer interventions invoked social norms (29%), identified a named victim (21%), depicted pets (32%), and gave implementation suggestions (29%), though these relatively lower percentages also reflect missing data on these variables when we did not have access to all intervention materials.

Most interventions (52%) made appeals only regarding animal welfare, with the remaining interventions also appealing to, for example, individual health or environmental concerns. The interventions were typically quite brief in duration: we estimated that 66% lasted less than 5 min, though there was a wide range, with some interventions lasting only about 30 s and others lasting over 90 min. Interventions also varied considerably in their use of explicit recommendations to the viewer: 13% recommended reducing meat consumption, 9% recommended going vegetarian, 7% recommended going vegan, and 24% made some

Table 2

Study characteristics regarding design, analytic reproducibility, and risks of bias. Binary and categorical variables are reported as “count (percentage)”. Continuous variables are reported as “median (first quartile, third quartile)”. k: Number of studies in subset. RCT: randomized controlled trial. NRCT: non-randomized controlled trial. UCT: uncontrolled trial (i.e., no separate control group) but with subjects serving as own controls.

Characteristic	All studies (k = 100)	Published studies (k = 17)	Unpublished studies (k = 83)
Design			
Between-subjects RCT	72 (72%)	16 (94%)	56 (67%)
Within-subject RCT	1 (1%)	1 (6%)	0 (0%)
Cluster RCT	2 (2%)	0 (0%)	2 (2%)
Between-subjects NRCT	21 (21%)	0 (0%)	21 (25%)
Within-subject UCT	4 (4%)	0 (0%)	4 (5%)
Outcome measurement			
Direct behavioral measure	2 (2%)	1 (6%)	1 (1%)
Self-reported past behavior	41 (41%)	1 (6%)	40 (48%)
Intended future behavior	57 (57%)	15 (88%)	42 (51%)
Percent missing data	7.5 (7.3, 42)	0 (0, 7.95)	13.2 (7.5, 59.75)
Not reported	23 (23%)	2 (12%)	21 (25%)
Exchangeability			
Low risk of bias	52 (52%)	16 (94%)	36 (43%)
Medium risk of bias	9 (9%)	1 (6%)	8 (10%)
High risk of bias	29 (29%)	0 (0%)	29 (35%)
Unclear	10 (10%)	0 (0%)	10 (12%)
Avoidance of social desirability bias			
Low risk of bias	13 (13%)	6 (35%)	7 (8%)
Medium risk of bias	14 (14%)	5 (29%)	9 (11%)
High risk of bias	68 (68%)	6 (35%)	62 (75%)
Unclear	5 (5%)	0 (0%)	5 (6%)
External generalizability			
Low risk of bias	36 (36%)	10 (59%)	26 (31%)
Medium risk of bias	28 (28%)	7 (41%)	21 (25%)
High risk of bias	15 (15%)	0 (0%)	15 (18%)
Unclear	21 (21%)	0 (0%)	21 (25%)
Preregistered			
Yes	25 (25%)	2 (12%)	23 (28%)
No	75 (75%)	15 (88%)	60 (72%)
Public data			
Yes	53 (53%)	5 (29%)	48 (58%)
No	47 (47%)	12 (71%)	35 (42%)
Public code			
Yes	22 (22%)	4 (24%)	18 (22%)
No	78 (78%)	13 (76%)	65 (78%)
Overall lowest risk of bias			
Yes	12 (12%)	9 (53%)	3 (4%)
No	88 (88%)	8 (47%)	80 (96%)

^a: For studies in which missing data was unreported but the outcome was measured in the same session as the intervention, we coded missing data as 0. Details on the risk-of-bias categories are provided in the Supplement.

combination of these recommendations. Another 43% of interventions made no explicit recommendation, instead simply describing or depicting animal welfare in a manner that was intended to reduce meat consumption. The overwhelming majority (98%) of interventions were not personally tailored (i.e., their content was the same for all subjects regardless of personal characteristics); the few tailored interventions were from a single article (FIAPO, 2017) in which the interventions' contents differed for currently vegetarian subjects versus for currently non-vegetarian subjects.

3.2.3. Outcomes

Throughout this section, we describe the outcomes used in our own analysis based on the considerations described in Section 2, which sometimes differed from the main outcome reported in the article, especially when raw data were available. Nearly all studies (96 studies; 96%) assessed consumption of meat or animal products, with the remaining studies instead assessing purchasing. 57% of studies assessed subjects' intended future behavior, another 41% of studies assessed subjects' self-reported behavior, and the remaining 2% (2 studies) used direct objective measures of subjects' behavior (i.e., the amount eaten of an offered sample of ham (Anderson & Barrett, 2016) or the percentage of actually purchased meals that contained meat (Schwitzgebel et al., 2019)). About half of studies (53%) measured outcomes immediately after exposure to the intervention. Among the studies that measured outcomes after some delay, the median and maximum follow-up lengths

were 39 and 120 days, respectively. For 71% of studies, our extracted point estimates were risk ratios of low versus high meat or animal product consumption¹⁰ defined in absolute terms (e.g., being below versus above the baseline median consumption); the remaining estimates were risk ratios of reducing consumption relative to the subject's own previous consumption.

3.2.4. Design quality, reproducibility, and risks of bias

Table 2 summarizes study characteristics related to the methodological strength of study design, analytic reproducibility, potential for selective reporting, and risks of bias. The table presents these characteristics for all studies, for published studies, and for unpublished studies. Overall, 75% of point estimates were from randomized studies¹¹ (with randomization occurring between subjects, within subjects, or by clusters of subjects); the remaining estimates were from nonrandomized designs with a separate control group or in which subjects' meat consumption was assessed before and after the intervention. Regarding analytic reproducibility and the potential for selective reporting, 53% of

¹⁰ As described in Section 2.2, we defined the outcome in terms of animal product consumption rather than meat consumption when the intervention itself targeted animal product consumption.

¹¹ When a study did not explicitly mention randomization, we assumed it was not randomized.

studies had publicly available data, 22% had publicly available analysis code, and 25% had been preregistered. The median percent of missing data at the longest follow-up time point was 7.5%, but 23% of studies did not report on missing data.¹² As described above, nearly all studies (98%) measured outcomes based on subjects' self-reports of behavior or of future intentions, rather than based on direct behavioral measures. We gave "low" or "medium" risk-of-bias ratings (versus "high" or "unclear") to 61% of studies regarding exchangeability of the intervention and control groups, to 27% regarding social desirability bias, and to 64% of studies regarding external generalizability.¹³ Published studies appeared to be at lower risk of bias than unpublished studies on most criteria, except that unpublished studies had notably better analytic reproducibility and preregistration practices (Table 2). Post hoc, we defined studies with the lowest risk of bias overall as those that were randomized, had less than 15% missing data, and had "low" or "medium" risks of bias for exchangeability and social desirability bias.¹⁴ 12 studies, contributed by 6 unique articles (Anderson & Barrett, 2016; Earle et al., 2019; Kunst & Haugestad, 2018; Kunst & Hohle, 2016; Lackner, 2019; Tian et al., 2016), were identified as being at lowest risk of bias.

3.3. Main analyses

The overall pooled risk ratio was 1.22 (95% CI: [1.13, 1.33]; $p < 0.0001$), indicating that on average, interventions increased by 22% a subject's probability of intending, self-reporting, or behaviorally demonstrating low versus high meat consumption, compared to the control condition. Most articles (71%) contributed more than one point estimate to the meta-analysis, and point estimates were moderately correlated within articles (intraclass correlation coefficient = 0.44).

Fig. 2 shows point estimates, 95% confidence intervals [CI], and calibrated estimates for each study. Interventions' effect sizes appeared fairly heterogeneous across studies, with an estimated standard deviation of the log-risk ratios of $\hat{\tau} = 0.12$. Supplementary Fig. S2 shows the estimated distribution of true population effects across all studies. Despite this heterogeneity, the interventions' effects were overwhelmingly in the beneficial direction (i.e., reducing rather than increasing meat consumption, purchase, or relevant intentions), with an estimated 83% (95% CI: [72%, 91%]) of true risk ratios above 1. Upon more stringently considering only risk ratios of at least 1.1, or alternatively 1.2, as being meaningfully large, we estimated that 71% (95% CI: [59%, 80%]) and 53% (95% CI: [38%, 64%]) of effects were stronger than these two thresholds, respectively. Supplementary Fig. S3 displays the estimated percentage of risk ratios above various other thresholds. Considering effects in the detrimental direction (i.e., interventions that "backfired" by increasing rather than decreasing meat consumption

¹² For studies in which outcomes were measured immediately after the intervention and that did not report on missing data, we assumed there was none.

¹³ However, these results regarding exchangeability, social desirability bias, and external generalizability should be interpreted cautiously because interrater reliability for these characteristics was fairly low, reflecting both their somewhat subjective nature and the fact that many studies lacked relevant methodological details (Supplement).

¹⁴ We created this designation in order to compare the *relatively* strongest studies in our meta-analysis to the others as a sensitivity analysis, rather than to claim that the former studies were, in absolute terms, at low risk of bias on all domains. This designation was inherently somewhat arbitrary and considered only a subset of risk-of-bias variables that were most directly related to studies' ability to unbiasedly estimate a causal treatment effect, even if that treatment effect may apply only to certain types of subjects, may apply only to intended behavior, or may apply only to behavior immediately after the intervention. If we had required high ratings on every risk-of-bias variable simultaneously, too few studies would have received the designation for us to conduct the sensitivity analysis.

relative to the control condition), we estimated that 17% (95% CI: [6%, 27%]) of interventions had true population risk ratios below 1. Very few interventions produced risk ratios smaller than 0.90 in the detrimental direction (4%; 95% CI: [0%, 12%]). The 10 studies with the largest calibrated estimates (i.e., the top 10%) used interventions consisting of: brief factual passages that graphically described and/or visually depicted conditions on factory farms (Cordts et al., 2014; Reese, 2015) or fish farms (Rouk, 2017), sometimes combined with health or environmental appeals (Reese, 2015); a 16-page leaflet containing detailed information and graphic portrayals of conditions on factory farms along with health appeals and implementation suggestions ("Even If You Like Meat" leaflet; Norris and Roberts (2016)); a factual mock newspaper article with graphic photographs and descriptions of gestation crates along with discussion of legislation to ban their use (Caldwell, 2016); a 3-min virtual reality video graphically depicting conditions on factory farms (FIPO, 2017); and meat-animal reminders consisting of photographs of meat dishes portrayed alongside photographs of the animals they came from (Earle et al., 2019). Nine of these 10 studies assessed outcomes related to intended behavior, and the remaining study assessed self-reported behavior (Norris and Roberts, 2016).

3.4. Sensitivity analyses

3.4.1. Publication bias

The meta-analytic mean corrected for publication bias (Hedges, 1992) was 1.33 (95% CI: [1.22, 1.45]; $p < 0.0001$). Thus, this estimate was in fact somewhat *larger* than the uncorrected meta-analytic mean of 1.22, though the estimates had substantially overlapping confidence intervals. This result suggests that publication bias was likely mild and did not meaningfully affect results. As a post hoc analysis, we fit the selection model to unpublished and to published studies separately, yielding similar conclusions. The sensitivity analyses for publication bias indicated that even under hypothetical worst-case publication bias (i.e., if "statistically significant" positive results were infinitely more likely to be published than "nonsignificant" or negative results), the meta-analytic mean would decrease to 1.07 (95% CI: [1.02, 1.13]) but would remain positive and with a confidence interval bounded above the null. This worst-case estimate arises from meta-analyzing only the 75 observed "nonsignificant" or negative studies and excluding the 25 observed "significant" or positive studies (Mathur and VanderWeele, 2020c). As a graphical heuristic, the significance funnel plot in Fig. 1 relates studies' point estimates to their standard errors and compares the pooled estimate within all studies (black diamond) to the worst-case estimate (grey diamond). When the diamonds are close to one another or the grey diamond represents a positive, nonnegligible effect size, as seen here, the meta-analysis may be considered fairly robust to publication bias (Mathur and VanderWeele, 2020c). Taken together, the results from the selection model and from the sensitivity analysis suggest that publication bias appeared negligible and not likely to substantially attenuate the results.

3.4.2. Social desirability bias

We conducted a simple sensitivity analysis (VanderWeele and Li (2019); Mathur and VanderWeele (2020b); see Supplement for methodological details) that characterizes how severe social desirability bias would have to have been in the meta-analyzed studies in order for: (1) the pooled point estimate of $RR = 1.22$ to be entirely attributable to social desirability bias (i.e., such that the interventions in fact had no effect on *actual* consumption on average); or (2) the percentage of true population effects stronger than $RR = 1.1$ to be reduced from our estimated 71% to only 10%. Specifically, the sensitivity analyses characterize the hypothetical severity of social desirability bias as how strongly the interventions affected subjects' *reported* consumption independently of their effects on subjects' *actual* consumption (VanderWeele & Li, 2019). For example, this form of social desirability bias could arise in a study in which the intervention did not affect subjects' actual

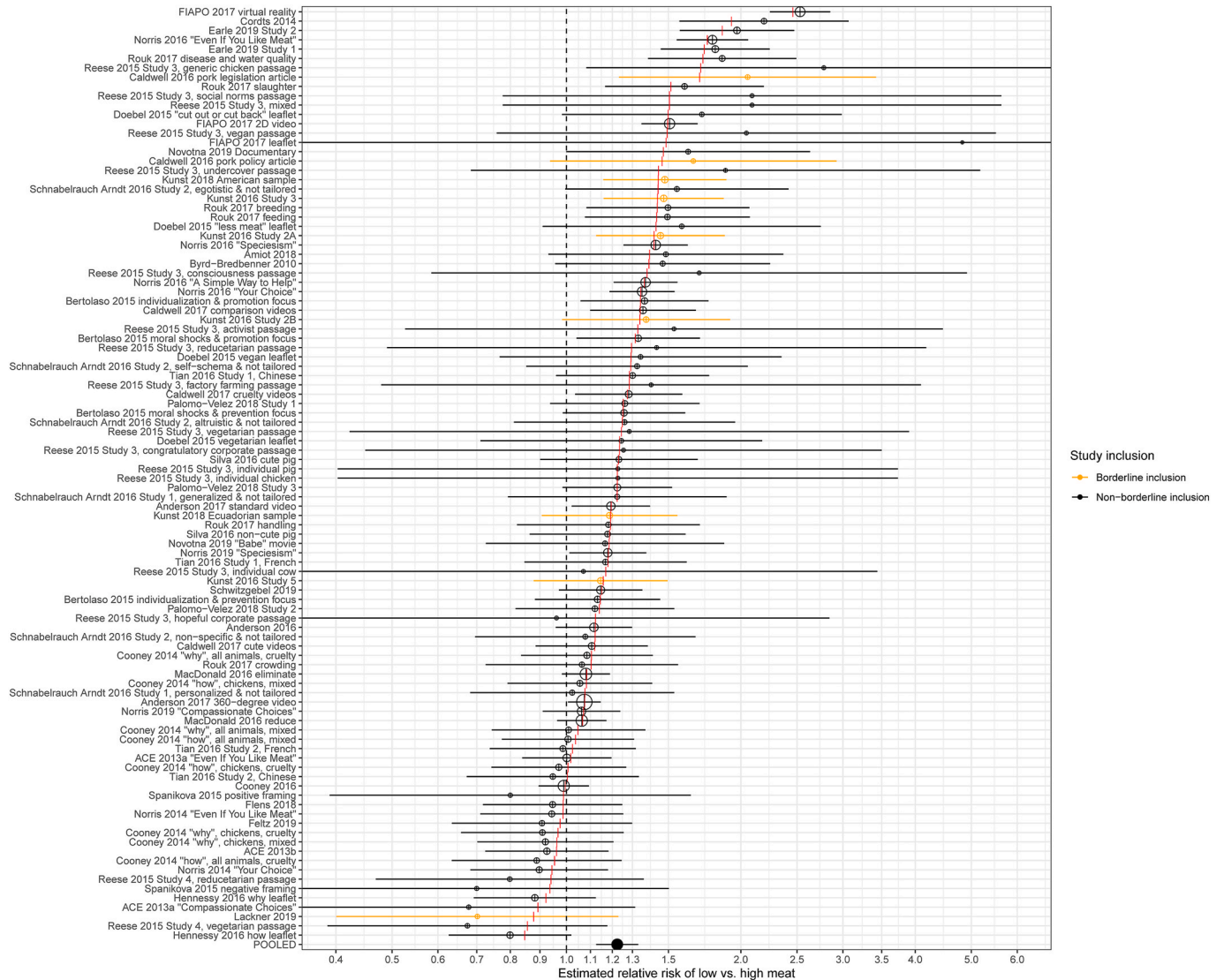


Fig. 2. Point estimates in each study (open circles), ordered by the study’s calibrated estimate (vertical red tick marks), and the overall meta-analytic mean (solid circle). Areas of open circles are proportional to the estimate’s relative weight in the meta-analysis. Orange estimates were borderline with respect to inclusion criteria and were excluded in sensitivity analysis. The x-axis is presented on the log scale. Error bars represent 95% confidence intervals. The vertical, black dashed line represents the null (no intervention effect). (For interpretation of the references to colour bars in this figure legend, the reader is referred to the Web version of this article.)

Table 3

Sensitivity analyses conducted on different groups of studies, with the overall estimate from the main analysis reported for comparison. Mean risk ratio: meta-analytic mean with 95% confidence interval. p-value: for mean risk ratio versus null of 1. k: Number of studies in subset. $\hat{\tau}$: estimated standard deviation of true population effects on log-risk ratio scale. Final three columns: estimated percentage of true population effects stronger than various thresholds on risk ratio scale. Bracketed values are 95% confidence intervals for the percentage of effects stronger than a threshold, which were sometimes not estimable (“NA”) when $\hat{\tau} = 0$ exactly or when the estimation proportion was very high.

Studies analyzed	k	Mean risk ratio	p-value	$\hat{\tau}$	% above 1	% above 1.1	% above 1.2
Main analysis	100	1.22 [1.13, 1.33]	<0.0001	0.12	83 [72, 91]	71 [59, 80]	53 [38, 64]
Excluding borderline-eligible studies	91	1.21 [1.11, 1.33]	0.0003	0.12	82 [69, 91]	68 [55, 77]	52 [35, 63]
Excluding composite interventions	52	1.29 [1.18, 1.40]	<0.0001	0.11	98 [NA, NA]	87 [71, 94]	71 [38, 83]
Studies at lowest risk of bias	12	1.3 [0.98, 1.72]	0.06	0.21	92 [NA, NA]	75 [0, 92]	58 [8, 83]
Excluding studies measuring intended behavior	43	1.11 [0.98, 1.26]	0.08	0.14	70 [47, 88]	47 [26, 60]	30 [9, 49]
Randomized studies	75	1.24 [1.14, 1.34]	<0.0001	0.15	93 [79, 99]	80 [66, 88]	61 [36, 77]
Preregistered studies with open data	21	1.09 [0.99, 1.19]	0.06	0	100 [NA, NA]	0 [NA, NA]	0 [NA, NA]
Published studies	17	1.35 [1.09, 1.67]	0.02	0.16	100 [NA, NA]	88 [NA, NA]	71 [0, 94]
Unpublished studies	83	1.19 [1.08, 1.32]	0.001	0.13	80 [65, 89]	66 [52, 77]	51 [34, 64]
Excluding one extreme estimate	99	1.22 [1.13, 1.33]	<0.0001	0.1	83 [71, 91]	71 [59, 80]	53 [36, 64]
Including SSWS studies	108	1.31 [1.19, 1.44]	<0.0001	0	81 [71, 87]	69 [56, 76]	56 [44, 65]

consumption at all, but did induce them to under-report meat consumption more than they otherwise would, thus yielding a spurious intervention effect estimate. If there were no social desirability bias of this form, the interventions would affect reported consumption exclusively via their effects on actual consumption.

Another plausible form of social desirability bias could arise if subjects were to systematically under-report meat consumption, but by a similar degree in the intervention and control groups (e.g., Hebert et al. (1995); Neff et al. (2018); Rothgerber (2020)). Critically, this form of under-reporting would in general leave estimates of intervention effects either unbiased or would bias them toward, rather than away from, the null (Rothman et al., 2008). Hence, this form of under-reporting is of less concern than the type of differential social desirability bias that we considered in the sensitivity analyses, which could bias the estimates away from the null (Rothman et al., 2008).

The sensitivity analyses indicated that, for the observed $RR = 1.22$ to be entirely attributable to social desirability bias, the interventions would have needed to increase subjects’ probability of reporting low meat consumption, entirely independently of their potential effects on actual consumption, by at least 22% (95% CI: [17%, 29%]) on average. If the severity of social desirability bias were approximately the same for all studies, then to reduce the percentage of true population effects stronger than $RR = 1.1$ from our estimated 71% to only 10% would require that each intervention had increased subjects’ probability of reporting low meat consumption, independent of its effects on actual consumption, by at least 37% (95% CI: [31%, 55%]).

Table 4

Meta-regressive estimates of effect modification by various study design and intervention characteristics (**coarse model**). Intercept: estimated mean risk ratio when all listed covariates are set to 0. For binary covariates, estimates represent risk ratios for the increase in intervention effectiveness associated with a study’s having, versus not having, the covariate. For the percentage of male subjects, the estimate represents the risk ratio for the increase in intervention effectiveness associated with a 10-percentage-point increase in males. For the average age, the estimate represents the risk ratio for the increase in intervention effectiveness associated with a 5-year increase in average subject age. CI: confidence interval. p-values are for the effect modification coefficients themselves, not for the subset of studies having the listed characteristic.

Coefficient	Effect modification RR [95% CI]	p-value
Intercept	1.11 [0.66, 1.86]	0.66
Intervention had text	0.85 [0.68, 1.06]	0.12
Intervention had visuals	0.96 [0.81, 1.13]	0.57
Intervention had graphic content	1.10 [0.97, 1.24]	0.13
Intervention’s recommendation		
No recommendation	Ref.	Ref.
“Reduce consumption”	1.00 [0.77, 1.31]	0.98
“Go vegetarian”	1.03 [0.78, 1.36]	0.81
“Go vegan”	1.31 [1.06, 1.62]	0.03
Mixed recommendation	0.99 [0.83, 1.19]	0.94
Intervention duration >5 min	1.03 [0.86, 1.24]	0.70
Follow-up length at least 7 days	0.81 [0.68, 0.97]	0.03
Percentage male subjects (10-pt increase)	1.00 [0.95, 1.06]	0.88
Average subject age (5-year increase)	1.04 [0.98, 1.10]	0.15

Table 5

Meta-regressive estimates of effect modification by various study design and intervention characteristics (**fine-grained model**), presented as in Table 4.

Coefficient	Effect modification RR [95% CI]	p-value
Intercept	0.87 [0.51, 1.46]	0.56
Intervention used mind attribution	1.00 [0.74, 1.33]	0.97
Intervention used social norms	1.03 [0.64, 1.65]	0.88
Intervention identified victim	0.94 [0.65, 1.35]	0.68
Intervention depicted pets	1.11 [0.70, 1.78]	0.47
Intervention used implementation suggestions	1.24 [0.85, 1.80]	0.22
Intervention had visuals	1.04 [0.87, 1.25]	0.61
Intervention had graphic content	1.07 [0.88, 1.29]	0.45
Intervention's recommendation		
No recommendation	Ref.	Ref.
"Reduce consumption"	0.89 [0.68, 1.17]	0.33
"Go vegetarian"	0.91 [0.63, 1.31]	0.55
"Go vegan"	1.25 [0.78, 1.99]	0.29
Mixed recommendation	0.78 [0.54, 1.12]	0.14
Intervention duration >5 min	0.96 [0.66, 1.40]	0.81
Follow-up length at least 7 days	0.76 [0.50, 1.14]	0.15
Percentage male subjects (10-pt increase)	0.98 [0.91, 1.06]	0.57
Average subject age (5-year increase)	1.08 [1.00, 1.17]	0.06

to have larger effects than studies of interventions making no recommendation (effect modification $RR = 1.31$; 95% CI: [1.06, 1.62]; $p = 0.03$), and point estimates in studies whose interventions recommended “going vegetarian” (1.03) or “reducing consumption” (1.00) heuristically suggested some degree of dose-response such that studies with broader-scoped recommendations (e.g., “go vegan” versus “reduce meat consumption”) typically had larger effects. Regarding other study characteristics, studies measuring outcomes after a relatively long delay (≥ 7 days) typically had smaller effect sizes (effect modification $RR = 0.81$; 95% CI: [0.68, 0.97]; $p = 0.03$), which could reflect decays in intervention effectiveness over time. See the Discussion for important guidance on the interpretation of these meta-regression results. [Supplementary Figs. S4–S6](#) plot individual studies’ calibrated estimates versus the interventions’ durations, studies’ lengths of follow-up, and average ages.

In the fine-grained meta-regression model ($k = 80$ studies; Table 5), the point estimates for the fine-grained intervention components were again usually close to the null, with the possible exception of using implementation suggestions (effect modification $RR = 1.24$; 95% CI: [0.85, 1.80]; $p = 0.22$). In most cases, the confidence intervals were fairly wide. Estimates for the other intervention components (Table 5) were usually similar to those in the coarse model.

3.6. Self-selected within-subjects studies

As described in Section 3.1, self-selected within-subjects (SSWS) studies as defined in Section 3.1 were excluded post hoc from main analyses, though they did meet our preregistered inclusion criteria. When we instead included the 8 SSWS studies from which we could obtain point estimates, the meta-analytic point estimate was somewhat larger than in main analyses (Table 3), but overall conclusions were not affected.

4. Discussion

4.1. Summary of findings

Despite authoritative calls for academic research and public policy regarding reducing meat consumption (Gardner et al., 2019; Godfray et al., 2018; Scherer et al., 2019; Tilman and Clark, 2014), the study of interventions to reduce meat consumption by appealing to animal welfare has largely remained outside the purview of academic research. Few, if any, attempts have been made to synthesize the evidence. In our systematic meta-analysis of this literature, we found that the body of evidence on this topic is well-developed both in terms of its size (100

studies reported in 34 articles) and in terms of the psychological sophistication of existing interventions (Section 1.1). The interventions appeared consistently beneficial at least in the short term and with outcomes primarily based on self-reported behavior or intended behavior, on average increasing by 22% an individual’s probability of intending, self-reporting, or behaviorally demonstrating low versus high meat consumption.

Although it seemed plausible, *a priori*, that in some settings, animal welfare appeals could be in danger of backfiring (e.g., by being perceived as intrusively self-righteous and moralizing), our results suggested that this danger was rarely realized, as we estimated that the large majority of interventions (83%) had true population effects in the beneficial rather than detrimental direction. This finding also suggests that the interventions were consistently effective across the culturally diverse samples we meta-analyzed, spanning at least 11 countries and four continents. However, it remains theoretically possible that even interventions with beneficial average effects could backfire for specific individuals (Rothgerber, 2020), a possibility that could be addressed in large studies by measuring individual-level effect modifiers or by assessing individually tailored interventions (Section 4.3).

These generally favorable results regarding animal welfare interventions do not imply, however, that appealing to animal welfare is necessarily more effective than, for example, appealing to concerns about health or the environment. To our knowledge, no quantitative meta-analysis of the latter types of interventions has been performed, precluding direct comparisons of evidence strength. Additionally, the literature on individuals’ reported motivations for reducing meat consumption is somewhat mixed: a nationally representative survey in the United States (McCarthy and DeKoster, 2020) suggested that individuals who reported having reduced their meat consumption most often cited health as a major or minor motivation (90% of subjects), though animal welfare was another important motivation (65% of subjects). Another survey suggested that primary motivations were cost (51% of subjects) and health (50%), with concerns about animal welfare (12%) and the environment (12%) cited considerably less often (Neff et al., 2018).

We would speculate three potential explanations for these findings regarding the motivating role of animal welfare concerns relative to other concerns. First, these differences in reported motivations may simply reflect the prevalence of receiving information and interventions with different appeals, rather than differences in effectiveness between appeals. As described in the Introduction, the public is largely uninformed and even misinformed about farm animal welfare, reflecting successful efforts by the animal agriculture industry as well as individuals’ deliberate avoidance of distressing information (Cornish et al., 2016; Rothgerber, 2020), whereas the public may be better

informed about potential health consequences of excessive meat consumption. Second, if animal welfare interventions are indeed effective, this may be in part because they provide information that dissonates with many individuals' stated ethical values (Cornish et al., 2016; Rothgerber, 2020), perhaps leading these individuals to reduce meat consumption even if they would not do so out of concern for their own health or the environment. Third, food choices in general are shaped more by unconscious adherence to habit and situational cues than by conscious motivations (van't Riet et al., 2011); we would therefore speculate that non-educational components of animal welfare interventions (such as leveraging the physical-moral disgust connection, invoking social norms, and giving implementation suggestions) might largely bypass conscious motivations in their influence on behavior.

We investigated the associations of specific study and intervention characteristics with effectiveness. Critically, these meta-regression estimates of effect modification based on covariates that vary across studies should not be interpreted as causal interactions due to potential confounding, as we further discuss below (Thompson & Higgins, 2002; VanderWeele & Knol, 2011). That is, these estimates represent the correlation between effect size and various characteristics of studies and interventions, not causation. We found that, on average, studies of interventions that recommended going vegan had effects that were 31% larger than interventions making no recommendation, and effects in studies whose interventions made intermediate recommendations to "go vegetarian" or "reduce consumption" heuristically suggested dose-response such that studies of interventions making more forceful recommendations (e.g., "go vegan" versus "reduce meat consumption") may have had larger effects on average. These findings preliminarily suggest that interventions that make more forceful recommendations do not seem to backfire and may in fact be more effective than subtler interventions. Studies measuring outcomes after a relatively long delay (≥ 7 days) typically had smaller effect sizes by an estimated 19%, though there was considerable statistical uncertainty due to the small number of studies with long follow-up times. Interventions' durations, their inclusion of text or visuals, and studies' percentages of male subjects did not appear meaningfully associated with effect size. Again, these estimates may not represent causal effects of different intervention designs; they could be confounded if, for example, more forceful interventions were typically used in studies of individuals who were already particularly motivated to reduce their meat consumption. Future research should conduct more head-to-head comparisons of interventions varying on these apparent effect modifiers, using randomization for those that can be directly manipulated (as in Cooney (2014); Macdonald et al. (2016)).

4.2. Strengths and limitations

The studies we meta-analyzed had some notable methodological strengths, including predominantly randomized designs, large sample sizes, and, especially among unpublished studies, a relatively high rate of data-sharing and preregistration. The studies also had limitations, including the potential for social desirability bias, the use of short lengths of follow-up (and typically high attrition when longer lengths of follow-up were used), and the use of self-reported rather than direct behavioral outcome measures; these issues were most apparent in the unpublished studies. Study designs appeared to be typically stronger with respect to exchangeability and external generalizability. However, a number of studies did appear susceptible to confounding due to a lack of randomization or differential attrition, and others may have had limited generalizability due to recruitment strategies that targeted highly motivated subjects or demographically restricted subjects (with samples tending to over-represent young North Americans, especially college undergraduates). Nevertheless, a number of sensitivity analyses, including those that included only studies at the lowest risk of bias by different criteria, had little effect on the results. Additionally, given the typically short lengths of follow-up, it is not clear whether interventions'

effects were sustained long enough to meaningfully improve individual health, a point we discuss further in Section 4.3. On the other hand, inexpensive interventions such as leaflets or online videos could be deployed for a large number of individuals, potentially still producing meaningfully large environmental and animal welfare impacts in aggregate even if effects might have been short-lived for any given individual.

Regarding potential bias due to the literature's predominant reliance on self-reported outcome measures, such measures can introduce random noise if subjects estimate their consumption imperfectly. However, if this misreporting is, on average, no more or less severe for the intervention group than the control group, then typically intervention effect estimates would be either unbiased or biased toward, rather than away from, the null (Rothman et al., 2008). A more pernicious problem with self-report measures is their potential to also induce misreporting that is *differential* between the intervention and control group, which could potentially inflate estimates away from the null. Sensitivity analyses indicated that for the pooled point estimate to have been entirely attributable to such social desirability biases, interventions would have had to increase subjects' probability of reporting low meat consumption, independently of any effects the interventions may have had on *actual* consumption, by at least 22%; and for social desirability bias of constant severity across studies to reduce to only 10% the percentage of true population effects stronger than $RR = 1.1$, each intervention would have had to increase subjects' probability of reporting low meat consumption, independently of actual consumption, by at least 37%. While it is clear that individuals in some cultures systematically under-report meat consumption in the absence of any meat-reducing intervention (Neff et al., 2018; Rothgerber, 2020), less is known about whether interventions reliably produce *differential* under-reporting of 22%–37% in under-reporting above and beyond this baseline under-reporting.

Our meta-analysis contained mostly unpublished studies, including a large number conducted by animal advocacy nonprofits. These studies might in principle be more subject to influence by research funders than peer-reviewed academic studies. On the other hand, peer-reviewed studies might in principle be subject to stronger traditional publication bias favoring positive results. However, we observed no evidence to support either possibility. Instead, it appeared that the literature on animal welfare interventions overall did not show evidence of publication bias, and this held even when we considered only unpublished, or only published, studies (Section 3.4.1). Furthermore, the meta-analytic means were similar in unpublished and published studies, and the former estimate was in fact somewhat smaller (Section 3.5). Additionally, unpublished studies showed a higher rate of preregistration, presumably limiting selective reporting. Selective reporting could also arise during our own processes of data extraction, analysis, or reporting, though we took considerable precautions to minimize this possibility, including publishing and preregistering the protocol in detail (Mathur et al., 2020), defining rules for data extraction that were as detailed and objective as possible, and documenting all post hoc analyses as such.

4.3. Recommendations for future research

To further advance the field, one approach is to continue improving the interventions themselves, for example based on the potential effect modifiers identified above. Additionally, content could be tailored to individual subject characteristics, such as personality traits (Kaptein et al., 2012), individual receptivity to different types of recommendations or to graphic depictions of factory farming conditions, or disgust sensitivity (Schnall et al., 2008). Studies could also assess whether exposing subjects to the interventions repeatedly rather than only once might increase effects. However, given the substantial existing body of research and the psychological sophistication and apparent effectiveness of existing interventions, we believe the field would be best served if future research were to first prioritize increasingly rigorous assessment

of existing interventions.

To this end, we make three key methodological recommendations. Given financial constraints on study designs, we give concrete recommendations for study designs that may facilitate adopting these recommendations with relatively modest increases in cost, along with examples of existing studies that have demonstrated the feasibility of these designs. First, studies should **use direct behavioral outcome measures** of meat consumption rather than self-reports of behavior or of future intentions (Peacock, 2018), and the measures should be designed to minimize social desirability bias (Hebert et al., 1995). Self-reported dietary data can be subject to substantial measurement error even when collected via measures such as 24-h recalls or food frequency questionnaires (Freedman et al., 2014; Peacock, 2018). Additionally, self-reported measures may exacerbate bias due to social desirability concerns: it is plausible that exposure to an animal welfare intervention could increase subjects' perceptions of social pressure to report lower meat consumption, and if outcomes are self-reported, subjects exposed to the intervention may report lower consumption than control subjects purely as an artifact of differential social desirability concerns. Some existing work has proposed or used behavioral measures based on purchase data from college dining halls or similar cafeteria settings (Garnett et al., 2019; Haile et al., 2020; Jalil et al., 2019; Schwitzgebel et al., 2019), meal choices at conferences (Hansen et al., 2019) or cafés (Anderson, 2020; Sparkman & Walton, 2017), or commercial marketing data providers (Peacock, 2018). Lab-based measurement of inexpensive clinical biomarkers of meat consumption, such as those obtained from hair samples, may be another cost-effective future avenue (Cuparencu et al., 2019; Peacock, 2018). Using measures such as purchase data and clinical biomarkers would also facilitate keeping subjects naïve to the outcome of interest and thus reduce social desirability bias.

Second, to determine how long interventions' effects last, studies should **measure outcomes at longer follow-up times after the intervention** (e.g., 1 month or more). When used to supplement pilot studies using shorter-term designs, such longitudinal follow-up would help identify whether some interventions' effects might be initially large but fade quickly, while other interventions' effects might be initially modest but sustained better. Additionally, longitudinal studies would enable principled assessment of psychological mediators, such as disgust or empathy, that underlie interventions' immediate and sustained effects (VanderWeele, 2015), which might help further refine interventions' content. Nevertheless, longer-term randomized studies may be prohibitively expensive and can introduce other methodological challenges, such as attrition. Introducing quasi-experimental designs, such as controlled interrupted time series or difference-in-difference methods (Bernal et al., 2017), to this field may facilitate following these recommendations. For example, quasi-experiments in which the intervention is deployed for a large group of subjects at the same time, rather than randomly assigned to individual subjects, might facilitate longitudinal collection of direct behavioral measures because these measures could be collected at the aggregate rather than individual level, and additionally without making subjects aware of that purpose of the study. Converging results from high-quality quasi-experiments and from randomized trials would strengthen the overall evidence base.

Third, outcome measures should be designed, when possible, to **assess the numerical volume of meat eaten or purchased** rather than only frequencies of consumption or purchase. Many existing studies measure meat consumption in terms of, for example, Likert-type items that categorize the number of weekly meals containing meat (e.g., "none", "1–5 meals", etc.) or in terms of reductions from one's previous consumption. When possible, using finer-grained absolute measures, such as the number of servings of poultry, beef, pork, lamb, fish, etc., would enable effect sizes to be translated into direct measures of societal impact, as described in Section 2.3. The typically coarse outcome measures in the current literature were not well-suited to conducting such analyses.

4.4. Conclusion

Our review and meta-analysis suggests that animal welfare interventions preliminarily appear effective, with meaningfully large effects, in these typically short-term studies of primarily self-reported outcomes. However, the existing literature does have some important methodological limitations. Although animal welfare interventions have received little academic attention to date, with the majority of studies reported in unpublished grey literature, our findings point to the importance of subjecting these promising interventions to increasingly rigorous and detailed empirical assessment. We recommend that interventions that appear effective in initial short-term studies with simple self-reported outcome measures – and our review has identified numerous existing strong candidates – should then be studied more rigorously using fine-grained, behavioral outcomes that minimize social desirability bias, and at longer time points following the intervention. Further developing the field of animal welfare interventions is a promising avenue for developing simple, effective interventions to reduce meat consumption, with potentially broad-reaching societal benefits.

List of abbreviations

CI = confidence interval; RR = risk ratio; SSWS = self-selected within-subjects.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data, materials, and analysis code required to reproduce all results are publicly available (<https://osf.io/8zsw7/>), with the exception of individual-subject data that we obtained by contacting authors and that cannot be made public at the authors' request. We have reported all deviations from the preregistration in the main text or Supplement.

Funding

This research was funded in part by The Pershing Square Fund for Research on the Foundations of Human Behavior. The funders had no role in the design, conduct, or reporting of this research.

Authors' contributions

MBM conceptualized the project. MBM, JP, DBR, and PAB designed the methodology with contributions from the other authors. DBR, JN, MBM, and JP reviewed articles and extracted qualitative data. MBM and JP extracted quantitative data. MBM conducted statistical analyses and wrote the manuscript, with all authors contributing critical revisions. All authors read and approved the final draft.

Declaration of competing interest

JP is employed by The Humane League Labs, which is a program of, and currently fully funded through, The Humane League, a 501(c) (3) nonprofit organization that "exists to end the abuse of animals raised for food". The Humane League Labs is editorially independent from The Humane League, and any other potential funders, in reporting research results. The Humane League had no role in the design, conduct, or reporting of this research. All other authors have no conflicts of interest to report.

Acknowledgments

Jo Anderson (Research Director, Faunalytics) provided helpful suggestions on a draft of the protocol. Sophie Attwood (Senior Behavioral Scientist, Better Buying Lab) provided leads to grey literature. We are grateful to the many authors and organizations who made their data publicly available, who shared raw data with us upon request (including Jack Norris [Vegan Outreach] and Koushik Raghavan [Federation of Indian Animal Protection Organizations]), and who answered questions about their studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.appet.2021.105277>.

References

- Adriaanse, M. A., Vinkers, C. D., De Ridder, D. T., Hox, J. J., & De Wit, J. B. (2011). Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. *Appetite*, *56*(1), 183–193. <https://doi.org/10.1016/j.appet.2010.10.012>
- Amiot, C. E., Boutros, G. E. H., Sukhanova, K., & Karelis, A. D. (2018). Testing a novel multicomponent intervention to reduce meat consumption in young men. *PLoS One*, *13*(10), Article e0204590. <https://doi.org/10.1371/journal.pone.0204590>
- Anderson, J. (2017). *An experimental investigation of the impact of video media on pork consumption*. Technical report, Faunalytics. <https://drive.google.com/file/d/1ovZ0xvJPKLGayAX-uCyGuZJKCFcY0TP/view>.
- Anderson, J. (2020). "Reduce" or "go veg?" effects on meal choice. Technical report, Faunalytics. <https://faunalytics.org/reduce-or-go-veg/>.
- Anderson, E. C., & Barrett, L. F. (2016). Affective beliefs influence the experience of eating meat. *PLoS One*, *11*(8), Article e0160424. <https://doi.org/10.1371/journal.pone.0160424>
- Arno, A., & Thomas, S. (2016). The efficacy of nudge theory strategies in influencing adult dietary behaviour: A systematic review and meta-analysis. *BMC Public Health*, *16*(1), 676. <https://doi.org/10.1186/s12889-016-3272-x>
- Bandura, A. (2002). Social foundations of thought and action: A social cognitive theory. In D. F. Marks (Ed.), *The health psychology reader* (pp. 94–106). <https://doi.org/10.1017/S0813483900008238> (chapter 6).
- Bartlett, J. G., Gilbert, D. N., & Spellberg, B. (2013). Seven ways to preserve the miracle of antibiotics. *Clinical Infectious Diseases*, *56*(10), 1445–1450.
- Bastian, B., & Loughnan, S. (2017). Resolving the meat-paradox: A motivational account of morally troublesome behavior and its maintenance. *Personality and Social Psychology Review*, *21*(3), 278–299. <https://doi.org/10.1177/1088868316647562>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Benningstad, N. C., & Kunst, J. R. (2020). Dissociating meat from its animal origins: A systematic literature review. *Appetite*, *147*. <https://doi.org/10.1016/j.appet.2019.104554>
- Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, *46*(1), 348–355. <https://doi.org/10.1093/ije/dyw098>
- Bertolaso, C. (2015). *Investigating the effectiveness of message framing and regulatory fit in increasing positive animal attitude and reducing animal products consumption: A study for animal equality*. Master's thesis. Universiteit van Amsterdam <https://scriptie.uba.uva.nl/scriptie/570273>.
- Bianchi, F., Dorsel, C., Garnett, E., Aveyard, P., & Jebb, S. A. (2018). Interventions targeting conscious determinants of human behaviour to reduce the demand for meat: A systematic review with qualitative comparative analysis. *International Journal of Behavioral Nutrition and Physical Activity*, *15*(1), 102. <https://doi.org/10.1186/s12966-018-0729-6>
- Bianchi, F., Garnett, E., Dorsel, C., Aveyard, P., & Jebb, S. A. (2018). Restructuring physical micro-environments to reduce the demand for meat: A systematic review and qualitative comparative analysis. *The Lancet Planetary Health*, *2*(9), e384–e397. [https://doi.org/10.1016/S2542-5196\(18\)30188-8](https://doi.org/10.1016/S2542-5196(18)30188-8)
- Bonnet, C., Bouamra-Mechemache, Z., Réquillart, V., & Treich, N. (2020). *Regulating meat consumption to improve health, the environment and animal welfare*. Food Policy. <https://doi.org/10.1016/j.foodpol.2020.101847>, 101847.
- Brewer, N. T., Hall, M. G., Noar, S. M., Parada, H., Stein-Seroussi, A., Bach, L. E., Hanley, S., & Ribisl, K. M. (2016). Effect of pictorial cigarette pack warnings on changes in smoking behavior: a randomized clinical trial. *JAMA Internal Medicine*, *176*(7), 905–912. <https://doi.org/10.1001/jamainternmed.2016.2621>
- Bulow, E. (2016). *rccmisc: Miscellaneous R Functions for Swedish Regional Cancer Centers*. R package version 0.3.7. <https://CRAN.R-project.org/package=rccmisc>.
- Byrd-Bredbenner, C., Grenici, A., & Quick, V. (2010). Effect of a television programme on nutrition cognitions and intended behaviours. *Nutrition and Dietetics*, *67*(3), 143–149. <https://doi.org/10.1111/j.1747-0080.2010.01446.x>
- Caldwell, K. (2016). *Welfare reforms and meat consumption*. Technical report, Mercy for Animals. <https://mercyforanimals.org/welfare-reforms-survey>.
- Canty, A., & Ripley, B. D. (2020). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-25.
- Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin*, *139*(2), 300. <https://doi.org/10.1037/a0030964>
- Coburn, K. M., & Vevea, J. L. (2019). *weightr: Estimating weight-function models for publication bias*. R package version 2.0.2. <https://CRAN.R-project.org/package=weightr>.
- Cooney, N. (2014). *What elements make a vegetarian leaflet more effective? Technical report, Humane League Labs*. <http://www.humaneleague.org/blog/2014-05-20-what-elements-make-vegetarian-leaflet-more-effective/>.
- Cordts, A., Nitzko, S., & Spiller, A. (2014). Consumer response to negative information on meat consumption in Germany. *The International Food and Agribusiness Management Review*, *17*(Special Issue A), 83–106. <https://doi.org/10.22004/ag.econ.164599>
- Cornish, A., Raubenheimer, D., & McGreevy, P. (2016). What we know about the public's level of concern for farm animal welfare in food production in developed countries. *Animals*, *6*(11), 74. <https://doi.org/10.3390/ani6110074>
- Covidence Development Team. (2019). *Covidence systematic review software*. Melbourne, Australia: Veritas Health Innovation.
- Crippa, A., Larsson, S. C., Discacciati, A., Wolk, A., & Orsini, N. (2018). Red and processed meat consumption and risk of bladder cancer: A dose–response meta-analysis of epidemiological studies. *European Journal of Nutrition*, *57*(2), 689–701. <https://doi.org/10.1007/s00394-016-1356-0>
- Cui, K., Liu, Y., Zhu, L., Mei, X., Jin, P., & Luo, Y. (2019). Association between intake of red and processed meat and the risk of heart failure: A meta-analysis. *BMC Public Health*, *19*(1), 354. <https://doi.org/10.1186/s12889-019-6653-0>
- Cuparencu, C., Praticó, G., Hemeryck, L. Y., Harsha, P. S. S., Noerman, S., Rombouts, C., Xi, M., Vanhaecke, L., Hanhineva, K., Brennan, L., et al. (2019). Biomarkers of meat and seafood intake: An extensive literature review. *Genes & Nutrition*, *14*(1), 1–30. <https://doi.org/10.1186/s12263-019-0656-4>
- Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4. <https://CRAN.R-project.org/package=xtable>.
- Di Marco, M., Baker, M. L., Daszak, P., De Barro, P., Eskew, E. A., Godde, C. M., Harwood, T. D., Herrero, M., Hoskins, A. J., Johnson, E., et al. (2020). Opinion: Sustainable development must account for pandemic risk. *Proceedings of the National Academy of Sciences*, *117*(8), 3888–3892. <https://doi.org/10.1073/pnas.2001655117>
- Dowle, M., & Srinivasan, A. (2019). *data.table: Extension of 'data.frame'*. R package version 1.12.8 <https://CRAN.R-project.org/package=data.table>.
- Earle, M., Hodson, G., Dhont, K., & MacInnis, C. (2019). Eating with our eyes (closed): Effects of visually associating animals with meat on antivegan/vegetarian attitudes and meat consumption willingness. *Group Processes & Intergroup Relations*, *22*(6), 1–18. <https://doi.org/10.1177/1368430219861848>
- Farvid, M. S., Stern, M. C., Norat, T., Sasazuki, S., Vineis, P., Weijenberg, M. P., Wolk, A., Wu, K., Stewart, B. W., & Cho, E. (2018). Consumption of red and processed meat and breast cancer incidence: A systematic review and meta-analysis of prospective studies. *International Journal of Cancer*, *143*(11), 2787–2799. <https://doi.org/10.1002/ijc.31848>
- Feinberg, M., Kovacheff, C., Teper, R., & Inbar, Y. (2019). Understanding the process of moralization: How eating meat becomes a moral issue. *Journal of Personality and Social Psychology*, *117*, 50–72. <https://doi.org/10.1037/pspa0000149>
- FIAPO. (2017). *Effectiveness assessment of outreach methods used by federation of Indian animal protection organisations (FIAPO)*. New Delhi, India: Technical report.
- Fisher, Z., & Tipton, E. (2015). *Robumeta: An R-package for robust variance estimation in meta-analysis*. <https://github.com/zackfisher/robumeta>. arXiv preprint arXiv: 1503.02220.
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks CA: Sage <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Freedman, L. S., Commins, J. M., Moler, J. E., Arab, L., Baer, D. J., Kipnis, V., ... Prentice, R. L., et al. (2014). Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology*, *180*(2), 172–188. <https://doi.org/10.1093/aje/kwu116>
- Fretts, A. M., Follis, J. L., Nettleton, J. A., Lemaitre, R. N., Ngwa, J. S., Wojczynski, M. K., ... Houston, D. K., et al. (2015). Consumption of meat is associated with higher fasting glucose and insulin concentrations regardless of glucose and insulin genetic risk scores: A meta-analysis of 50,345 caucasians. *American Journal of Clinical Nutrition*, *102*(5), 1266–1278. <https://doi.org/10.3945/ajcn.114.101238>
- Gamer, M., & Lemon, J. (2019). *irr: Various coefficients of interrater reliability and agreement*. <puspendra.pusp22@gmail.com>, I. F. P. S, R package version 0.84.1 <https://CRAN.R-project.org/package=irr>.
- Gardner, C. D., Hartle, J. C., Garrett, R. D., Offringa, L. C., & Wasserman, A. S. (2019). Maximizing the intersection of human health and the health of the environment with regard to the amount and type of protein produced and consumed in the United States. *Nutrition Reviews*, *77*(4), 197–215. <https://doi.org/10.1093/nutrit/nuy073>
- Garnett, E. E., Balmford, A., Sandbrook, C., Pilling, M. A., & Marteau, T. M. (2019). Impact of increasing vegetarian availability on meal selection and sales in cafeterias. *Proceedings of the National Academy of Sciences*, *116*(42), 20923–20929.
- Gnagnarella, P., Caini, S., Maisonneuve, P., & Gandini, S. (2018). Carcinogenicity of high consumption of meat and lung cancer risk among non-smokers: A comprehensive meta-analysis. *Nutrition and Cancer*, *70*(1), 1–13. <https://doi.org/10.1080/01635581.2017.1374420>
- Godfray, H. C. J., Aveyard, P., Garnett, T., Hall, J. W., Key, T. J., Lorimer, J., Pierrehumbert, R. T., Scarborough, P., Springmann, M., & Jebb, S. A. (2018). Meat

- consumption, health, and the environment. *Science*, 361(6399). <https://doi.org/10.1126/science.aam5324>. eam5324.
- Grattan, B. J., & Connolly-Schoonen, J. (2012). Addressing weight loss recidivism: A clinical focus on metabolic rate and the psychological aspects of obesity. *ISRN Obesity*. <https://doi.org/10.5402/2012>, 2012.
- Guasch-Ferré, M., Satiya, A., Blondin, S. A., Janiszewski, M., Emlen, E., O'Connor, L. E., Campbell, W. W., Hu, F. B., Willett, W. C., & Stampfer, M. J. (2019). Meta-analysis of randomized controlled trials of red meat consumption in comparison with various comparison diets on cardiovascular risk factors. *Circulation*, 139(15), 1828–1845. <https://doi.org/10.1161/CIRCULATIONAHA.118.035225>
- Haille, M., Jalil, A., & Tasoff, J. (2020). *Changing hearts and plates: The effect of animal-advocacy pamphlets on meat consumption*. Preprint retrieved from <https://papers.ssrn.com/sol3/papers.cfm?abstractid=3753878>.
- Hansen, P. G., Schilling, M., & Malthesen, M. S. (2019). Nudging healthy and sustainable food choices: Three randomized controlled field experiments using a vegetarian lunch-default as a normative signal. *Journal of Public Health*.
- Hebert, J. R., Clemow, L., Pbert, L., Ockene, I. S., & Ockene, J. K. (1995). Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *International Journal of Epidemiology*, 24(2), 389–398. <https://doi.org/10.1093/ije/24.2.389>
- Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*, 7(2), 246–255. <https://doi.org/10.1214/ss/1177011364>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/rsm.5>
- Hennessy, S. R. (2016). *The impact of information on animal product consumption*. Master's thesis. University of Illinois at Urbana-Champaign <https://www.ideals.illinois.edu/bitstream/handle/2142/92865/HENNESSY-THESIS-2016.pdf?sequence=1&isAllowed=y>.
- Higgs, S. (2015). Social norms and their influence on eating behaviours. *Appetite*, 86, 38–44. <https://doi.org/10.1016/j.appet.2014.10.021>
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177. <https://doi.org/10.1111/j.1750-8606.2008.00061.x>
- Jalil, A., Tasoff, J., & Bustamante, A. (2019). Eating to save the planet: Evidence from a randomized controlled trial using individual-level food purchase data. *Social Science Research Network (SSRN) Electronic Journal*. <https://doi.org/10.2139/ssrn.3444642>
- Jasper, J. M., & Poulsen, J. D. (1995). Recruiting strangers and friends: Moral shocks and social networks in animal rights and anti-nuclear protests. *Social Problems*, 42(4), 493–512. <https://doi.org/10.2307/3097043>
- Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, 14(3), 235–257. <https://doi.org/10.1023/A:1007740225484>
- Jin, Z.-C., Zhou, X.-H., & He, J. (2015). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine*, 34(2), 343–360. <https://doi.org/10.1002/sim.6342>
- Kaptein, M., De Ruyter, B., Markopoulos, P., & Aarts, E. (2012). Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TIS)*, 2(2), 10. <https://doi.org/10.1145/2209310.2209313>
- Kim, K., Hyeon, J., Lee, S. A., Kwon, S. O., Lee, H., Keum, N., Lee, J.-K., & Park, S. M. (2017). Role of total, red, processed, and white meat consumption in stroke incidence and mortality: A systematic review and meta-analysis of prospective cohort studies. *Journal of the American Heart Association*, 6(9), Article e005983. <https://doi.org/10.1161/JAHA.117.005983>
- Kim, Y., & Je, Y. (2018). Meat consumption and risk of metabolic syndrome: Results from the Korean population and a meta-analysis of observational studies. *Nutrients*, 10(4), 390. <https://doi.org/10.3390/nu10040390>
- Knight, S., & Barnett, L. (2008). Justifying attitudes toward animal use: A qualitative study of people's views and beliefs. *Anthrozoös*, 21(1), 31–42. <https://doi.org/10.2752/089279308X274047>
- Kuhn, M., Jackson, S., & Cimentada, J. (2020). *corr: Correlations in R*. R package version 0.4.2. <https://CRAN.R-project.org/package=corr>.
- Kunst, J. R., & Haugstad, C. A. P. (2018). The effects of dissociation on willingness to eat meat are moderated by exposure to unprocessed meat: A cross-cultural demonstration. *Appetite*, 120, 356–366. <https://doi.org/10.1016/j.appet.2017.09.016>
- Kunst, J. R., & Hohle, S. M. (2016). Meat eaters by dissociation: How we present, prepare and talk about meat increases willingness to eat meat by reducing empathy and disgust. *Appetite*, 105, 758–774. <https://psycnet.apa.org/doi/10.1016/j.appet.2017.09.016>
- Lackner, Z. (2019). *The effect of dissociation between meat and animal on the experience of ambivalence towards meat*. Université de Grenoble Alpes.
- Larsson, S. C., & Orsini, N. (2013). Red meat and processed meat consumption and all-cause mortality: A meta-analysis. *American Journal of Epidemiology*, 179(3), 282–289. <https://doi.org/10.1093/aje/kwt261>
- Larsson, S. C., & Wolk, A. (2006). Meat consumption and risk of colorectal cancer: A meta-analysis of prospective studies. *International Journal of Cancer*, 119(11), 2657–2664. <https://doi.org/10.1002/ijca.293.2.172>
- Loughnan, S., Bastian, B., & Haslam, N. (2014). The psychology of eating animals. *Current Directions in Psychological Science*, 23(2), 104–108. <https://doi.org/10.1177/0963721414525781>
- Loughnan, S., Haslam, N., & Bastian, B. (2010). The role of meat consumption in the denial of moral status and mind to meat animals. *Appetite*, 55(1), 156–159. <https://doi.org/10.1016/j.appet.2010.05.043>
- Lüdtke, D. (2020). *sjstats: Statistical functions for regression models*. Version 0.18.0 <https://CRAN.R-project.org/package=sjstats>.
- Macdonald, B., Caldwell, K., & Boese, G. (2016). *The effects of 'reduce' and 'eliminate' appeals on individual meat consumption*. Technical report. Reductarian Foundation. <https://osf.io/nxrx3/>.
- Machovina, B., Feeley, K. J., & Ripple, W. J. (2015). Biodiversity conservation: The key is reducing meat consumption. *The Science of the Total Environment*, 536, 419–431. <https://doi.org/10.1016/j.scitotenv.2015.07.022>
- Marshall, B., & Levy, S. (2011). Food animals and antimicrobials: Impacts on human health. *Clinical Microbiology Reviews*, 24(4), 718–733. <https://doi.org/10.1128/cmr.00002-11>
- Mathur, M. B., Robinson, T. N., Reichling, D. B., Gardner, C. D., Nadler, J., Bain, P. A., & Peacock, J. (2020). Reducing meat consumption by appealing to animal welfare: Protocol for a meta-analysis and theoretical review. *Systematic Reviews*, 9(1), 1–8. <https://doi.org/10.1186/s13643-019-1264-5>
- Mathur, M. B., & VanderWeele, T. J. (2019). New metrics for meta-analyses of heterogeneous effects. *Statistics in Medicine*, 1336–1342. <https://doi.org/10.1002/sim.8057>
- Mathur, M. B., & VanderWeele, T. J. (2020a). *PublicationBias: Sensitivity analysis for publication bias in meta-analyses*. R package version 2.1.0 <https://CRAN.R-project.org/package=PublicationBias>.
- Mathur, M. B., & VanderWeele, T. J. (2020b). Robust metrics and sensitivity analyses for meta-analyses of heterogeneous effects. *Epidemiology*, 9(1), 1–8. <https://doi.org/10.1097/ede.0000000000001180>
- Mathur, M. B., & VanderWeele, T. J. (2020c). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series A C*, 5(69), 1091–1119.
- Mathur, M. B., Wang, R., & VanderWeele, T. J. (2019). *MetaUtility: Utility Functions for Conducting and interpreting meta-analyses*. R package version 2.1.0. <https://CRAN.R-project.org/package=MetaUtility>.
- McCarthy, J., & DeKoster, S. (2020). *Nearly one in four in U.S. have cut back on eating meat*. <https://news.gallup.com/poll/282779/nearly-one-four-cut-back-eating-meat.aspx>.
- Neff, R. A., Edwards, D., Palmer, A., Ramsing, R., Righter, A., & Wolfson, J. (2018). Reducing meat consumption in the USA: A nationally representative survey of attitudes and behaviours. *Public Health Nutrition*, 21(10), 1835–1844. <https://doi.org/10.1017/S1368890017004190>
- Neuwirth, E. (2014). *RColorBrewer: ColorBrewer palettes*. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>.
- Norris, J. (2014). *Booklet comparison study: Pay per read #1*. Technical report. Vegan Outreach. <https://veganoutreach.org/pay-per-read-study-results/>.
- Norris, J., & Hannan, R. (2019). *Leafletting and booklet effectiveness*. Technical report. Vegan Outreach. <https://veganoutreach.org/leafletting-and-booklet-effectiveness/>.
- Norris, J., & Roberts, E. (2016). *Booklet comparison study - pay per read #2*. Technical report. Vegan Outreach. <https://veganoutreach.org/ppr-2016/>.
- Onwezen, M. C., & van der Weele, C. N. (2016). When indifference is ambivalence: Strategic ignorance about meat consumption. *Food Quality and Preference*, 52, 96–105. <https://doi.org/10.1016/j.foodqual.2016.04.001>
- Ooi, H., de Vries, A., & Microsoft. (2020). *checkpoint: Install packages from snapshots on the checkpoint server for reproducibility*. R package version 0.4.9. <https://CRAN.R-project.org/package=checkpoint>.
- Outreach, V. (2018). *Compassionate choices*. <https://veganoutreach.org/wp-content/uploads/2020/01/CC.pdf>.
- Pan, A., Sun, Q., Bernstein, A. M., Schulze, M. B., Manson, J. E., Willett, W. C., & Hu, F. B. (2011). Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *American Journal of Clinical Nutrition*, 94(4), 1088–1096. <https://doi.org/10.3945/ajcn.111.018978>
- Peacock, J. (2018). *Measuring change in diet for animal advocacy*. Technical report, the Humane League Labs. <https://thehumaneleague.org/article/E009R01-measuring-diet-animal-advocacy>.
- R Core Team. (2020a). *foreign: Read data stored by 'minitab', 'S', 'SAS', 'SPSS', 'stata', 'sysstat', 'weka', 'dBase', ...* R package version 0.8-80. <https://CRAN.R-project.org/package=foreign>.
- R Core Team. (2020b). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reese, J. (2015). *Testing the effectiveness of animal advocacy messages with Amazon Mechanical Turk*. Technical report. The Sentience Institute. https://jacyreese.com/assets/downloads/Testing_the_Effectiveness.pdf.
- van't Riet, J., Sijtsma, S. J., Dagevos, H., & De Bruijn, G.-J. (2011). The importance of habits in eating behaviour: an overview and recommendations for future research. *Appetite*, 57(3), 585–596. <https://doi.org/10.1016/j.appet.2011.07.010>
- Robinson, T. N. (2010). Save the world, prevent obesity: Piggybacking on existing social and ideological movements. *Obesity*, 18(S1), S17–S22. <https://doi.org/10.1038/oby.2009.427>
- Robinson, T. N. (2017). Stealth interventions for obesity: Strategies for behavioral, social, and policy changes. In K. D. Brownell, & B. T. Walsh (Eds.), *Eating disorders and obesity: A comprehensive handbook* (pp. 609–613). Guilford Publications (chapter 99).
- Robinson, T. N., Killen, J. D., Kraemer, H. C., Wilson, D. M., Matheson, D. M., Haskell, W. L., Pruitt, L. A., Powell, T. M., Owens, A., & Thompson, N. (2003). Dance and reducing television viewing to prevent weight gain in african-american girls: The stanford GEMS pilot study. *Ethnicity & Disease*, 13(Suppl 1), S1–S65. <https://www.ethndis.org/priorsuparchives/ethn-13-s1-65.pdf>.
- Robinson, T. N., Matheson, D. M., Kraemer, H. C., Wilson, D. M., Obarzanek, E., Thompson, N. S., Alhassan, S., Spencer, T. R., Haydel, K. F., Fujimoto, M., et al. (2010). A randomized controlled trial of culturally tailored dance and reducing screen time to prevent weight gain in low-income African American girls: Stanford GEMS. *Archives of Pediatrics and Adolescent Medicine*, 164(11), 995–1004. <https://doi.org/10.1001/archpediatrics.2010.197>

- Rodriguez-Sanchez, F. (2018). *grateful: Facilitate Citation of R Packages*. R package version 0.0.2. <https://github.com/Pakillo/grateful>.
- Rothgerber, H. (2020). Meat-related cognitive dissonance: A conceptual framework for understanding how meat eaters reduce negative arousal from eating animals. *Appetite*, 146, 104511. <https://doi.org/10.1016/j.appet.2019.104511>
- Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology*. Lippincott Williams & Wilkins.
- Rouhani, M., Salehi-Abargouei, A., Surkan, P., & Azadbakht, L. (2014). Is there a relationship between red or processed meat intake and obesity? A systematic review and meta-analysis of observational studies. *Obesity Reviews*, 15(9), 740–748. <https://www.researchgate.net/deref/http>.
- Rouk, S. (2017). Which farmed fish welfare issues does the public care most about? (Neutral messaging). Technical report, Mercy for Animals <https://mercyforanimals.org/which-fish-welfare-issue-generates-the-most>.
- Rozin, P. (1996). Towards a psychology of food and eating: From motivation to module to marker, morality, meaning, and metaphor. *Current Directions in Psychological Science*, 5(1), 18–24. <https://doi.org/10.1111/1467-8721.ep10772690>
- Rozin, P., & Fallon, A. (1980). The psychological categorization of foods and non-foods: A preliminary taxonomy of food rejections. *Appetite*, 1(3), 193–201. [https://doi.org/10.1016/S0195-6663\(80\)80027-4](https://doi.org/10.1016/S0195-6663(80)80027-4)
- Rozin, P., Hormes, J. M., Faith, M. S., & Wansink, B. (2012). Is meat male? A quantitative multimethod framework to establish metaphoric relationships. *Journal of Consumer Research*, 39(3), 629–643. <https://doi.org/10.1086/664970>
- Ruby, M. B. (2012). Vegetarianism. A blossoming field of study. *Appetite*, 58(1), 141–150. <https://doi.org/10.1016/j.appet.2011.09.019>
- Sakadevan, K., & Nguyen, M.-L. (2017). Livestock production and its impact on nutrient pollution and greenhouse gas emissions. *Advances in Agronomy*, 141, 147–184. <https://doi.org/10.1016/bs.agron.2016.10.002>
- Scherer, L., Behrens, P., & Tukker, A. (2019). Opportunity for a dietary win-win-win in nutrition, environment, and animal welfare. *One Earth*, 1(3), 349–360. <https://doi.org/10.1016/j.oneear.2019.10.020>
- Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34(8), 1096–1109. <https://doi.org/10.1177/0146167208317771>
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5), 429–434. <https://doi.org/10.1111/j.1467-9280.2007.01917.x>
- Schwitzgebel, E., Cokelet, B., & Singer, P. (2019). Ethics classes can influence student behavior: Students purchase less meat after discussing arguments for vegetarianism. *The Splintered Mind (Blog)*. <http://schwitsplinters.blogspot.com/2019/07/ethics-classes-can-influence-student.html>.
- da Silva, C. R. (2016). *Am I too cute to eat? The effect of cuteness appeal towards the promotion of a more plant-based diet*. Master's thesis. University Institute of Lisbon <http://hdl.handle.net/10071/12583>.
- Singer, P. (1995). *Animal liberation*. Random House.
- Sparkman, G., & Walton, G. M. (2017). Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological Science*, 28(11), 1663–1674. <https://doi.org/10.1177/0956797617719950>
- Sutton, A. J., Abrams, K. R., Jones, D. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester: Wiley.
- Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21(11), 1559–1573. <https://doi.org/10.1002/sim.1187>
- Tian, Q., Hilton, D., & Becker, M. (2016). Confronting the meat paradox in different cultural contexts: Reactions among Chinese and French participants. *Appetite*, 96, 187–194. <https://doi.org/10.1016/j.appet.2015.09.009>
- Tierney, N., Cook, D., McBain, M., & Fay, C. (2020). *naniar: Data structures, summaries, and visualisations for missing data*. R package version 0.5.2. <https://CRAN.R-project.org/package=naniar>.
- Tilman, D., & Clark, M. (2014). Global diets link environmental sustainability and human health. *Nature*, 515(7528), 518–522. <https://doi.org/10.1038/nature13959>
- VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press.
- VanderWeele, T. J. (2017). On a square-root transformation of the odds ratio for a common outcome. *Epidemiology*, 28(6), e58–e60. <https://doi.org/10.7326/0003-4819-154-10-201105170-00008>
- VanderWeele, T. J. (2019). *Optimal approximate conversions of odds ratios and hazard ratios to risk ratios*. Biometrics.
- VanderWeele, T. J., & Knol, M. J. (2011). Interpretation of subgroup analyses in randomized trials: Heterogeneity versus secondary interventions. *Annals of Internal Medicine*, 154(10), 680–683. <https://doi.org/10.7326/0003-4819-154-10-201105170-00008>
- VanderWeele, T. J., & Li, Y. (2019). Simple sensitivity analysis for differential measurement error. *American Journal of Epidemiology*, 188(10), 1823–1829. <https://doi.org/10.1093/aje/kwz133>
- Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: Affect and charity are greatest for a single child in need. *PLoS One*, 9(6), Article e100115. <https://doi.org/10.1371/journal.pone.0100115>
- Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419–435. <https://doi.org/10.1007/BF02294384>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://www.jstatsoft.org/v36/i03/>.
- Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, 23(1), 73–82. <https://doi.org/10.1177/0963721413512856>
- Wang, C.-C., & Lee, W.-C. (2019). A simple method to estimate prediction intervals and predictive distributions: Summarizing meta-analyses beyond means and confidence intervals. *Research Synthesis Methods*, 10(2), 255–266. <https://doi.org/10.1002/jrsm.1345>
- Wang, X., Lin, X., Ouyang, Y. Y., Liu, J., Zhao, G., Pan, A., & Hu, F. B. (2016). Red and processed meat consumption and mortality: Dose-response meta-analysis of prospective cohort studies. *Public Health Nutrition*, 19(5), 893–905. <https://doi.org/10.1017/s1368980015002062>
- Weintraub, D. L., Tirumalai, E. C., Haydel, K. F., Fujimoto, M., Fulton, J. E., & Robinson, T. N. (2008). Team sports for overweight children: The stanford sports to prevent obesity randomized trial (SPORT). *Archives of Pediatrics and Adolescent Medicine*, 162(3), 232–237. <https://doi.org/10.1001/archpediatrics.2007.43>
- Wickham, H. (2011a). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1), 1–29. <http://www.jstatsoft.org/v40/i01/>.
- Wickham, H. (2011b). testthat: Get started with testing. *The R Journal*, 3, 5–10. <https://doi.org/10.32614/RJ-2011-002>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. <https://ggplot2.tidyverse.org>.
- Wickham, H. (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0. <https://CRAN.R-project.org/package=stringr>.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., & Bryan, J. (2019). *readxl: Read Excel Files*. R package version 1.3.1. <https://CRAN.R-project.org/package=readxl>.
- Wickham, H., François, R., Henry, L., & Müller, K. (2020). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.0. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, H., & Henry, L. (2020). *tidyr: Tidy Messy Data*. R package version 1.1.0. <https://CRAN.R-project.org/package=tidyr>.
- Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, 3(1), 129–137. <https://doi.org/10.1111/j.2041-210X.2011.00125.x>
- Wrenn, C. L. (2013). Resonance of moral shocks in abolitionist animal rights advocacy: Overcoming contextual constraints. *Society and Animals*, 21(4), 379–394. <https://doi.org/10.1163/15685306-12341271>
- Yoshida, K. (2020). *tableone: Create 'table 1' to describe baseline characteristics*. R package version 0.11.1. <https://CRAN.R-project.org/package=tableone>.
- Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10), 1–17. <https://doi.org/10.18637/jss.v011.i10>
- Zhang, Y., & Zhang, D. Z. (2018). Red meat, poultry, and egg consumption with the risk of hypertension: A meta-analysis of prospective cohort studies. *Journal of Human Hypertension*, 32(7), 507. <https://doi.org/10.1038/s41371-018-0068-8>