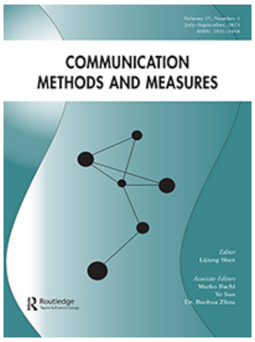

Routledge
Taylor & Francis Group

Communication Methods and Measures

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hcms20

Comparison Conditions in Research on Persuasive Message Effects: Aligning Evidence and Claims About Persuasiveness

Daniel J. O'Keefe

COMMUNICATION METHODS AND MEASURES
2023, VOL. 17, NO. 3, 187–204
https://doi.org/10.1080/19312458.2023.2214949

Comparison Conditions in Research on Persuasive Message Effects: Aligning Evidence and Claims About Persuasiveness

Daniel J. O'Keefe

Communication Studies, Northwestern University, Evanston, Illinois, USA

ABSTRACT
In persuasion message effects research, two kinds of research design are common. One compares the persuasiveness of two different advocacy messages on the same topic. The other compares the persuasiveness of an advocacy message against a no-advocacy-message control condition. Because these two designs contain different comparisons they underwrite different claims, but the designs – and their corresponding claims – are prone to misunderstanding and confusion. And when a study combines the two designs, especially complex issues can arise. This article aims to sort out the relevant issues in the service of better alignment between evidence and claims in persuasive message effects research.

Persuasion message effects research focuses on the effects that messages have on persuasive outcomes such as attitude, intention, and behavior. In this research domain, two kinds of research design are common. One compares the persuasiveness of two different advocacy messages on the same topic, comparing (for instance) a humorous and a non-humorous message (e.g., Nabi, 2016, Study 2), a moral and a pragmatic appeal (Van Zant & Moore, 2015), and so on. The other compares the persuasiveness of an advocacy message against a no-advocacy-message control condition, comparing (for instance) a narrative message and a no-message control condition (e.g., Igartua, 2010, Study 3), a health-appointment reminder and a no-reminder condition (e.g., Ritchie et al., 2000), and so on. These two designs underwrite different kinds of claims, but the designs – and their corresponding claims – are prone to misunderstanding and confusion.

In what follows, each design is described, its attendant misunderstandings sketched, and its contributions articulated. A subsequent section takes up some implications and complexities associated with these research designs – especially those arising when a combination of the two designs is employed. This article aims to sort out the relevant issues in the service of better alignment between claims and evidence in persuasive message effects research.

Two research designs

Two-advocacy-message designs

The design

In persuasion message effects research, one familiar sort of research design compares the persuasiveness of two advocacy messages. The two messages advocate the same position, but differ in a way thought to potentially affect their relative persuasiveness. Participants are randomly assigned to be exposed to one of the two messages.[1]

CONTACT Daniel J. O'Keefe d-okeefe@northwestern.edu Communication Studies, Northwestern University, 999 Michigan Avenue, Apt. 1C, Evanston, IL 60202-1445, USA

[1] Such random assignment is essential to permit appropriate interpretation of experimental results.

As examples: The two messages might differ in whether they discuss opposing arguments: a one-sided message presents only supportive arguments, a two-sided message both presents supportive arguments and discusses opposing arguments. Or the two messages might differ in whether they focus on the advantages of performing the advocated action (a gain-framed message) or the disadvantages of not performing the advocated action (a loss-framed message). (More complex versions of this design are possible, as when a study includes more than one message variation in a factorial design. But the essential elements of the design can be explicated by focusing on the simple two-message case.)

A two-advocacy-message design speaks to claims about the relative persuasiveness of the two messages being compared. The evidence bearing on such claims takes the form of an effect size that represents the difference on the outcome variable between message A and message B (e.g., d, the standardized mean difference). Such data address the question "what difference would it make whether message A, as opposed to message B, was used?" Expressed in practical terms, for a persuasive message designer the question addressed is "should I use message A or message B?"

Misunderstandings of results

Results from such two-advocacy-message designs can be misinterpreted in at least two ways.

One misinterpretation is thinking that the results underwrite claims about broader message categories. For example, results from an experiment that compares one gain-framed message against one loss-framed message might be interpreted as supporting conclusions about differences in the effects of gain-framed and loss-framed messages generally. But such conclusions are plainly not justified by the evidence because the design has only one example of each message category (the classic discussion is Jackson & Jacobs, 1983; see also Highhouse, 2009; Reeves et al., 2016; M. D. Slater et al., 2015; Thorson et al., 2012; Wells & Windschitl, 1999).

A second misinterpretation is thinking that the results underwrite claims about absolute message persuasiveness. For example, an experiment that finds a gain-framed message to be statistically significantly more persuasive than a loss-framed message is sometimes misunderstood as suggesting that the gain-framed message is very persuasive. However, results from this research design do not speak to the absolute persuasiveness of either message. (Confusion on this point has been discussed by O'Keefe, 2017.)

For example, if message A is much more persuasive than message B (i.e., a large effect size), that does not necessarily mean that message A was highly persuasive in absolute terms; both messages might have been relatively ineffective, but with message A being more effective than message B. Similarly, if the two messages are roughly equal in persuasiveness (i.e., a small effect size), that does not necessarily mean that the two messages were ineffective; both messages might have been highly effective, but without much difference between them in persuasiveness. Or if both messages backfire but message A backfires less than message B, message A will appear superior to message B on the outcome measure even though message A was not persuasive from the advocate's point of view.[2]

Briefly, then: In a two-advocacy-message design in which message A was significantly more effective than message B, it's a mistake to say "This shows that messages of kind A are more effective than messages of kind B;" that's a mistake because the design had only one example of each appeal type. And it's also a mistake to say "This shows that message A was effective;" that's a mistake because the result shows that message A was more effective than message B, not that message A was effective in any absolute sense.

2. Imagine a two-advocacy-message design in which message A's mean on the outcome variable was higher than message B's mean (i.e., from the advocate's perspective, message A looked better—that is, more effective). Absent additional information—and, specifically, as discussed below, absent a no-advocacy-message comparison condition—one cannot rule out the possibility that one or both of the messages produced a backfire effect.

The value of the design

Two-advocacy-message designs can be useful for several reasons. Where theoretical predictions about relative persuasiveness are made, such designs can provide relevant evidence. For example, if a theory predicts that loss-framed appeals will be more persuasive than gain-framed appeals (either in general or under specified moderating conditions), then a two-advocacy-message design can contribute useful data.

However, as discussed above, in designs that compare one specific message of one type against another specific message of the other type (e.g., one gain-framed message and one loss-framed message), the results cannot provide a good test of the usual theoretical predictions. Persuasion theories do not commonly predict that *every* message of form A will be more persuasive than *every* message of form B, but rather take the form of expectations about *mean* differences: the prediction is that, on average, messages of form A will be more persuasive than messages of form B (either in general or under specified moderating conditions). A study that compares one form-A message against one form-B message cannot test such predictions.

As data accumulate across many such message pairs, however, meta-analyses can provide better evidence about such theoretical predictions. Better evidence is also obtained when a single study uses multiple examples of each message category (e.g., Andrews et al., 2022), but (ceteris paribus) there is good reason to prefer meta-analytic results across multiple studies – different researchers, different contexts, different media, different sorts of participants, different concrete outcome measures, etc.— over results from any one multiple-message design (this, because of the resulting "heterogeneity of irrelevancies;" see, e.g., Shadish et al., 2002, pp. 361–363).

Two-advocacy-message designs can also shape practical message design decisions. In the narrowest case, where a design compares the persuasiveness of two specific messages, a campaign planner will be in a position to see whether one message is more effective than the other – and hence make an informed choice between those two specific messages. (In some contexts, designs with such purposes are described as A/B testing; see, e.g., Kohavi & Longbotham, 2017; Kohavi et al., 2020) And, as above: as individual two-advocacy-message design studies accumulate, meta-analyses can reveal whether there is a dependable difference in persuasiveness between the two message kinds (either in general or under specified moderating conditions), thus providing general guidance to message designers.

However, the value of a two-advocacy-message design turns centrally on the nature of the contrast between the two messages. That contrast constrains the sorts of claims to which the study might be relevant. Consider, for example, a design that compares two pro-recycling messages, one gain-framed and one loss-framed (e.g., Anghelcev & Sar, 2014). As mentioned above, such a comparison can be useful to those choosing between gain-framed and loss-framed recycling messages.

But sometimes researchers have used a curious variation on this usual design – a variation in which the two messages advocate *opposing* positions. For example, Thacker et al. (2020) reported that participants exposed to a message describing the advantages of genetically modified foods (GMFs) had significantly more positive attitudes about GMFs than did participants exposed to a message describing the disadvantages of GMFs; Abdel-Raheem and Alkhammash (2022, Experiment 1) reported that participants exposed to a pro-vaccination message about COVID were more willing to be vaccinated than were those exposed to an anti-vaccination message; Gong et al. (2017) found that participants presented with positive information about thrombolytic treatment for acute ischemic stroke preferred thrombolysis more than did participants presented with negative information. (For examples of similar designs, see Majmundar et al., 2020; Tan et al., 2017)

Results from such designs are of limited utility. They do not provide useful message-design guidance; after all, no sensible campaign planner would consider using a message advocating the opposing viewpoint. And it seems unlikely that researchers would be genuinely uncertain about which of the two messages will be more effective in producing a given outcome. However, comparing the effects of a pro-X message and an anti-X message does provide evidence about whether messages can make a difference (e.g., to political opinions); results indicating a significant difference between such messages' means on an outcome variable would be evidence that messages can matter.[3]

In any case, the larger point should be clear: In two-advocacy-message designs, the nature of the contrast between the two messages is crucially important. The claims that a given study can address are limited by the nature of the comparison afforded by the two messages.

And, correspondingly, meta-analyses of two-advocacy-message designs need to be attentive to the message contrast. Meta-analysis is appropriate and informative only when the collected studies all have the same message contrast. For example, imagine a set of studies that all had a one-sided message condition, but one study compared this to a one-sided text-plus-visual-material message, another study compared it to a two-sided message, a third study compared it to a narrative message, and so on.[4] The variety of comparison conditions would make any meta-analytic mean effect size uninterpretable.

As an example: Ma et al. (in press) reported a meta-analysis of vaping prevention messages, but the collection of studies involved a variety of comparisons: an anti-vaping message against a pro-vaping message (e.g., Majmundar et al., 2020), an anti-vaping message against a different anti-vaping message (e.g., Sontag et al., 2019), a pro-vaping message against an ambivalent message (one containing both pro-vaping and anti-vaping information, as in a vaping advertisement accompanied by a health warning; e.g., Katz et al., 2020), and so on. Overall mean effect sizes were reported and interpreted, but these are not meaningful. [For more discussion of this example, see O'Keefe (in press) and Noar et al. (in press). For examples of educational intervention meta-analyses that have encountered this problem, see A. Simpson (2017, pp. 454–456).]

No-advocacy-message control condition

The design

A different sort of research design compares the persuasiveness of an advocacy message against a no-advocacy-message control condition, with participants randomly assigned to condition. The control condition can take a variety of forms: Control-condition participants might not be exposed to any message (a no-message control), might be exposed to a message on a topic unrelated to the advocacy subject (a placebo message control), or might be exposed to a "neutral" message on the advocacy subject (that is, a message that is not an advocacy message) – but the key feature is that control-condition participants are not exposed to any advocacy message on the advocacy subject. One version of a no-message control is a waitlist control, in which participants in the control condition receive the same intervention as those in the treatment condition but only after some time has passed; the experimental comparison is between outcomes for treatment-condition participants and outcomes for waitlist-control participants before the latter have received the treatment.

As examples: Villanti et al. (2019) randomized participants to a treatment condition (nicotine education), a placebo message condition (sun safety education), and a no-message control condition. J. S. Slater et al. (2018) randomized Medicare patients to either an intervention condition (direct mail materials meant to encourage mammography and colonoscopy) or a no-message control condition (a waitlist control).

3. Some designs expose participants to multiple advocacy messages advocating opposing views (i.e., a participant sees both a pro-X message and an anti-X message). In some cases this design is used to explore message order effects ("primacy-recency" effects, e.g., Insko, 1964); in others it is used to simulate a competitive message environment (e.g., Chong and Druckman, 2010; Dillard et al., 2023).

4. Even when a collection of studies all appear to have the same message contrast, there may be hidden complexities. For example, the contrast between one-sided messages (that contain only arguments supporting the advocated view) and two-sided messages (that present supportive arguments but also discuss opposing arguments) is straightforward enough. But Allen(1991) pointed out that one can distinguish two different message contrasts here on the basis of exactly how the two-sided message discusses opposing arguments—by simply acknowledging them or by refuting them.

A no-advocacy-message control-condition design speaks to claims about whether it is more persuasive to expose participants to the advocacy message or to not expose them to the advocacy message. The evidence bearing on such claims takes the form of an effect size (e.g., *d*) describing the difference on the outcome variable between the advocacy-message condition and the no-advocacy-message condition. Such data thus address the question "what if there had been no advocacy message?" Expressed in practical terms, for a persuasive message designer the question addressed is "should I use this advocacy message, or should I say nothing?"

Misunderstandings of results

Results from such a research design can be misinterpreted in at least three ways.

One misinterpretation is thinking that the results underwrite claims about broader message categories. For example, an experiment that compares a narrative message against a no-advocacy-message control condition might be taken as evidence for conclusions about the effects of narrative messages generally. But, as above, when only a single instance of a message category is in hand, conclusions about the broader message category based solely on that one contrast are not justified.

A second misinterpretation is thinking that the active ingredient in the advocacy message can be easily identified. For example, an experiment that found a narrative message to be more persuasive than a no-advocacy-message control condition might point to the narrative format as the factor responsible for the observed effects. But the nature of the comparison does not permit one to identify the aspect of the advocacy-message condition that was responsible for the outcome. It might be, for example, that even a non-narrative advocacy message would have produced similar effects. Identifying the active ingredient requires a comparison condition other than a no-advocacy-message control.[5]

A third misinterpretation is thinking that the results underwrite claims about absolute message persuasiveness. For example, an experiment that finds a narrative message to be more persuasive than the no-advocacy-message control condition might be taken to suggest that the narrative message is very persuasive. However, results from this research design do not speak to the absolute persuasiveness of the advocacy message. These results address the question of the relative persuasiveness of advocating something (the advocacy message condition) or not advocating something on the advocacy subject (the control condition).

To see this clearly, consider these two hypothetical scenarios in which a local food bank randomly either sends or does not send a donation request to people who had donated previously. In scenario #1, the donation rate is 80% in the donation-request condition and 70% in the no-message condition; the effect size (Cohen's *h*, the difference between the arcsine-transformed proportions; Cohen, 1988) is $h = .23$. In scenario #2, the donation rate is 50% in the donation-request condition and 20% in the no-message condition; the effect size is $h = .64$. The proportion of people donating in the donation-request condition is higher in scenario #1 (80%, vs. 50% in scenario #2), but the effect size – the difference in outcome between the donation-request message and the no-message control – is larger in scenario #2; that is, the donation-request message leads to a greater increase in donations compared to the no-message control condition in scenario #2 (from 20% to 50%) than in scenario #1 (from 70% to 80%). The point is: The effect size describing the difference in effectiveness between the advocacy-message condition and a no-advocacy-message control condition (the difference between the two conditions' values on the outcome variable) does not provide information about the effectiveness of the advocacy message (the value of the advocacy condition on the outcome variable).

Briefly, then: In a no-advocacy-message control-condition design in which message A was significantly more persuasive than the no-advocacy-message condition, it's a mistake to say "This shows that messages of kind A are effective;" that's wrong, because the design had only one example of such a message and because one could not be sure what the active ingredient was. It's also a mistake to say

[5] In particular, some appropriate two-advocacy-message design would be useful, a design aimed at isolating potential active ingredients.

"This shows that message A was highly effective;" that's wrong, because the result shows that message A was more effective than saying nothing on the subject, not that it was effective in any absolute sense.

The value of the design

In some ways, no-advocacy-message control-condition designs might seem pointless. For example, suppose one wants to encourage a city's residents to recycle. It seems obvious that a message that tries to encourage recycling will likely be more effective (in persuading residents to recycle) than saying nothing. And the same is true a fortiori of placebo-message designs: a message that tries to encourage recycling will likely be more effective (in persuading residents to recycle) than a message about flu shots. Indeed, one might naturally wonder why researchers would even bother to gather evidence using no-advocacy-message control-condition designs.

But no-advocacy-message control-condition designs can be informative in two ways. First, such designs can speak to the question of whether people are at all persuadable on the advocacy subject. If an advocacy message produces more favorable outcomes than those seen in the no-advocacy-message control condition, then presumably persuasion is possible. (If an advocacy message does not produce more favorable outcomes than those seen in the no-advocacy-message control condition, one cannot conclude that people cannot be persuaded – only that the particular advocacy message tested did not persuade.)

Second, and in some ways more important: No-advocacy-message control-condition designs can provide evidence about potential backfire effects (boomerang effects), that is, message effects that are the opposite of those intended. If the advocacy message produces less favorable outcomes than those seen in the no-advocacy-message control condition, then advocates will want to be alert to the potential dangers of deploying advocacy messages. (If the advocacy message does not produce less favorable outcomes than those seen in the no-advocacy-message control condition, one cannot conclude that backfire effects will not occur – only that the particular advocacy message tested did not backfire.) The adage *primum non nocere*—first do no harm – would recommend attending to the possibility of such boomerang effects. And indeed backfire effects are not uncommon; for some examples, see Calabrese and Zhang (2019), Gosnell (2018), Liang et al. (2018), Lienemann and Siegel (2018), Myers et al. (2012), Nicolla and Lazard (in press), Richter et al. (2018), Ringold (2002), Ryoo and Kim (2023), Sato and Takasaki (2021), Schmid and Betsch (2022), B. Simpson et al. (2018), Skurka (2019), and Winett et al. (2021); for some general discussions, see Byrne and Hart (2009), Kim et al. (2017), and Stibe and Cugelman (2016). (For examples of research with no-advocacy -message conditions motivated by such concerns, see Fishbein et al., 2002; Rode et al., 2023)

So in circumstances in which it's not clear whether people are persuadable, or in which there is reason to fear backfire effects, then designs with a no-advocacy-message control condition can be especially informative.

The key difference between the designs

As is probably apparent, these two research designs address fundamentally different research questions because the designs have crucially different comparisons. A two-advocacy-message design speaks to the question of which of two advocacy messages is more persuasive; the relevant effect size describes the difference in persuasive outcomes between the two messages. A no-advocacy-message control-condition design speaks to the question of whether an advocacy message is more persuasive than not having an advocacy message on the advocacy subject; the relevant effect size describes the difference in persuasive outcomes between the advocacy-message condition and the no-advocacy-message condition.

Another way of expressing this difference between the designs is to say that the two designs address different potential counterfactuals. Specifically, they address different counterfactuals that would arise if only one advocacy message – call it message A – had been deployed. The two-advocacy-message design addresses the question "what if message B had been used rather than message A?" The no-advocacy-message control-condition design addresses the question "what if there had been no message A?"

Combining the designs

An individual primary-research study can combine these two sorts of designs. That is, a research design might (in the simplest case) have two advocacy messages and also a no-advocacy-message control condition. Such designs provide evidence both about the relative persuasiveness of the two advocacy messages studied and about whether either advocacy message is more persuasive than using no advocacy message, and thereby address weakness of each design alone.

For example, a combined design offers the possibility of detecting backfire effects that would go unseen in a two-advocacy-message design. Similarly, where no-advocacy-message designs lack evidence about the active ingredient responsible for any observed difference in effectiveness between the advocacy message and the no-advocacy-message condition, the presence of two advocacy messages can provide evidence about the possible role of the factor that varies between the two advocacy messages. And when the two advocacy messages advocate opposing views, including a no-advocacy-message condition can provide evidence about the question of the relative impact of each message; if, as one might commonly expect, the two messages' means (on an outcome variable such as attitude) are on opposite sides of the mean for the no-advocacy-message condition, then a comparison of the two effect sizes – the effect sizes comparing each advocacy message against the no-advocacy-message control – will indicate the relative effectiveness of each message.

However, results from such combined designs can take a complex variety of forms, and it will be useful to offer a general sketch. The purpose is to display concretely how complicated such results can be. The design to be discussed is the simplest possible one: a design with two advocacy messages (message A and message B) and one no-advocacy-message comparison condition. Even with this simple design, fifteen different patterns of means are possible, generated by the combination of (a) the ordering of the means in the various conditions and (b) the statistical significance of differences between conditions.

Beginning with the simple ordering of the means: In such a design, the means on an outcome variable can be ordered in three abstract ways: (1) message A/message B/no-advocacy; (2) no-advocacy /message A/message B; (3) message A/no-advocacy/message B. Briefly put – and ignoring statistical significance for the moment – the three orderings are: (1) both messages are better than the no-advocacy control, or (2) both messages are worse than the no-advocacy control, or (3) one message is better than, and the other worse than, the no-advocacy control. This is very much a simplification because it treats message A and message B as interchangeable and hence omits an ordering of means such as message B/message A/no-advocacy (because for present purposes that order is functionally equivalent to the message A/message B/no-advocacy order), because it omits cases in which two conditions have the same value on the outcome variable, and so on. Even so, this will be useful as a starting point.

Within each of these three orderings, results might differ depending on the statistical significance of the differences between conditions. As a further simplification, this discussion assumes equal sample sizes among the three conditions. The usefulness of that simplification is that differences in statistical significance – where one comparison is statistically significant and another is not – will reflect differences in effect sizes, not sample sizes; similarly, if a given effect size (say, $d = .2$) is statistically significant in one comparison, an effect size of that same magnitude will also be statistically significant in another comparison. Even with such simplification, each ordering of means can take on five distinct patterns when statistical significance is considered.

If the ordering of means is message A/message B/no-advocacy, the five possible patterns are: (1A) Message A is significantly more effective than message B; message B is significantly more effective than the no-advocacy condition; necessarily, then, given the order of means, message A is significantly more effective than the no-advocacy condition. (1B) Message A is not significantly more effective than message B; message B is not significantly more effective than the no-advocacy condition; message A is not significantly more effective than the no-advocacy condition. (1C) Message A is not significantly more effective than message B; message B is not significantly more effective than the no-advocacy condition; message A is significantly more effective than the no-advocacy condition. (1D) Message A is not significantly more effective than message B; message B is significantly more effective than the no-advocacy condition; necessarily, then, given the order of means, message A is significantly more effective than the no-advocacy condition. (1E) Message A is significantly more effective than message B; message B is not significantly more effective than the no-advocacy condition; necessarily, then, given the order of means, message A is significantly more effective than the no-advocacy condition.

If the ordering is no-advocacy/message A/message B, the five possible patterns are: (2A) The no-advocacy condition is significantly more effective than message A; message A is significantly more effective than message B; necessarily, then, given the order of means, the no-advocacy condition is significantly more effective than message B. (2B) The no-advocacy condition is not significantly more effective than message A; message A is not significantly more effective than message B; the no-advocacy condition is not significantly more effective than message B. (2C) The no-advocacy condition is not significantly more effective than message A; message A is not significantly more effective than message B; the no-advocacy condition is significantly more effective than message B. (2D) The no-advocacy condition is not significantly more effective than message A; message A is significantly more effective than message B; necessarily, then, given the order of means, the no-advocacy condition is significantly more effective than message B. (2E) The no-advocacy condition is significantly more effective than message A; message A is not significantly more effective than message B; necessarily, then, given the order of means, the no-advocacy condition is significantly more effective than message B.

If the ordering is message A/no-advocacy/message B, the five possible patterns are: (3A) Message A is significantly more effective than the no-advocacy condition; the no-advocacy condition is significantly more effective than message B; necessarily, then, given the order of means, message A is significantly more effective than message B. (3B) Message A is not significantly more effective than the no-advocacy condition; the no-advocacy condition is not significantly more effective than message B; message A is not significantly more effective than message B. (3C) Message A is not significantly more effective than the no-advocacy condition; the no-advocacy condition is not significantly more effective than message B; message A is significantly more effective than message B. (3D) Message A is not significantly more effective than the no-advocacy condition; the no-advocacy condition is significantly more effective than message B; necessarily, then, given the order of means, message A is significantly more effective than message B. (3E) Message A is significantly more effective than the no-advocacy condition; the no-advocacy condition is not significantly more effective than message B; necessarily, then, given the order of means, message A is significantly more effective than message B.

Laying out these various scenarios permits one to see more clearly the differences between them – and to avoid mistaken assumptions about how results might look. For example, one might be tempted to suppose that if message A is significantly more effective than message B, then message A must also be significantly more effective than a no-advocacy-message control condition.[6] But, as scenarios 2A, 2D, 3C, and 3D illustrate, that reasoning is mistaken. Or one might think that if message B is significantly more effective than the no-advocacy-message control, and message A and message B do not differ significantly in effectiveness, than message A must also be significantly more effective than the no-advocacy-message control. But, as scenario 1D illustrates, that reasoning is mistaken.

The research literature does contain examples of various patterns of results. For example, Powell-Jackson et al. (2018) found that gain-framed and loss-framed childhood vaccination messages increased vaccination relative to a no-message control condition, but the two

6. Looking at you, reviewer 2.

treatment messages did not differ significantly, an example of scenario 1D (see similarly Busso et al., 2019; Divdar et al., 2021; Nelson et al., 2021, study 1). Seiter et al. (in press) found that participants exposed to positive online comments about an online anti-vaping ad had significantly more positive attitudes toward the ad than participants exposed to negative comments, but neither condition differed significantly from a no-comment control condition. Ariel (2012) found that taxpayer compliance did not differ across a deterrence-message condition, a moral-persuasion-message condition, and a no-message control condition, an example of scenario 1B or 2B or 3B depending on the ordering of the means (see similarly Legate et al., 2022, p. 3; or, more complicatedly, Fesenfeld et al., 2021).

But even these 15 scenarios represent a very simplified picture. The possibilities are more complex if the three conditions have unequal sample sizes, or if more than two advocacy messages are studied, or if the design includes more than one advocacy-message variable. As examples: Severson and Coleman (2015) compared six messages invoking different arguments for climate-change mitigation policies (arguments based on economic efficiency, religious moral grounds, scientific evidence, etc.) with a no-message control condition, finding that some but not all of the advocacy messages were more persuasive than the no-message control. Dijkstra and Rotelli (2022) compared messages advocating lower meat consumption using arguments invoking environmental, health, or animal welfare considerations with a no-message control condition; as just one illustration of the complex results, for participants with relatively high meat consumption, heath-based arguments were more effective than, but environment-based arguments seemed less effective than, the no-message control condition.

These various possible patterns of results do underscore the difference between the two underlying types of research design. When an experiment has two advocacy messages and also a no-advocacy-message control condition, it can speak to two distinct research questions—one concerning the relative persuasiveness of the two advocacy messages studied and the other concerning whether either advocacy message is more persuasive than providing no advocacy on the subject. Unsurprisingly, then, research results can be complicated.

Implications, complexities, nuances

Persuasiveness is always relative to some comparison condition

One key idea underlying the analysis above is worth explicit discussion. Briefly: It does not make sense to describe a given message as persuasive simpliciter. The question must be: persuasive as compared to what?

In the two research designs described above, different comparison conditions are used and hence different claims about (relative) persuasiveness are assessed. The two-advocacy-message design asks how the persuasiveness of message A compares to the persuasiveness of message B (that is, whether message A is more or less persuasive than message B). The no-advocacy-message-control design asks whether message A is more or less persuasive than having no advocacy message on the subject. But neither design underwrites claims such as "message A is persuasive," that is, persuasive in absolute terms.

Compliance percentages

There is a natural temptation to suppose sometimes the absolute persuasiveness of a message can be assessed – namely, in cases where the outcome of interest is represented by a compliance percentage. A message that produces 20% compliance will be thought unpersuasive, and one that produces 80% compliance will be thought persuasive. But on closer consideration, the appropriateness of such descriptions can be seen to turn on just what comparisons are being made.

Imagine, for example, two applications of some compliance strategy (different contexts, different behaviors of interest, different target audiences, etc.). In application #1, the strategy produces 20%

compliance, but in a no-advocacy-message control condition the compliance rate was 10%. In application #2, the strategy produces 80% compliance, but in a no-advocacy-message control condition the compliance rate was 90%. Plainly, the strategy was successful in application #1 (doubling the compliance rate from 10% to 20%), but it backfired in application #2 (reducing the compliance rate from 90% to 80%).

Thus it does not seem to be meaningful to talk of message persuasiveness simpliciter, even where compliance percentages are concerned. If no comparison condition is specified, a claim of the form "that message was persuasive" is insufficiently articulated. There is no scale of absolute persuasiveness and so it does not make sense to speak of a message being "persuasive" or "unpersuasive" in absolute terms. And that's true even if (e.g.) "90% of people exposed to this message subsequently performed the advocated action," because one can still ask "what if there had been no message?" If 90% would have performed the action anyway, it would be misleading to describe the message as having been persuasive; if 95% would have performed the action anyway, it would be wrong to describe the message as having been persuasive. The point, thus, is: The comparison condition always matters, even for compliance percentages.

Pretest values as a comparison condition

When outcome variables are measured both before and after message exposure, post-exposure assessments can be compared to pre-exposure assessments (a pre/post design). Indeed, traditional communication campaign evaluations often take the form of comparing message recipients' pre-campaign and post-campaign values on variables of interest. (Some examples: Bleakley et al., 2018; Fagan et al., 2020; Kite et al., 2018; Weiler et al., 2017.)

It can be tempting to think of pretest values as the answer to the question "what would have happened if there had been no message?" and so to think of pre/post designs as approximating designs with a no-message control condition. The pre-campaign assessment is taken to represent what would have been the case if there had been no campaign (the equivalent of the no-message control condition); the post-campaign assessment represents the effect of the campaign (the equivalent of the advocacy message condition).

However, pre-campaign assessments cannot be assured to represent what would have happened if the campaign had not been conducted. That is, pre-campaign assessments are not necessarily the equivalent of a no-message control condition, because confounding factors might be present, that is, influences on the outcome of interest other than the campaign.

Consider, for example, a study of mailing postcards to encourage flu shots among Medicare recipients in Utah and Nevada who were not vaccinated in the prior year (described by Maglione et al., 2002, p. 44). Among participants who received a postcard providing information about susceptibility and severity, the vaccination rate was 18.9%; this was plainly a noticeable increase over the previous year's figure (zero). However, among participants who were not vaccinated in the prior year and did *not* receive a postcard, the vaccination rate was 18.3%.

As this illustrates, pre-exposure assessments are not always a good substitute for a straightforward no-message control condition.[7] If one wants to know what would have happened if the campaign had not occurred, data from a contemporaneous no-advocacy-message condition is likely preferable to pre-exposure data. For example, Reger et al. (1999) assessed the effects of a campaign to encourage switching to low-fat milk by (inter alia) comparing milk sales in the intervention city and in a matched no-intervention city (see, similarly, Morley et al., 2018; for a review of such studies of health campaigns, see Anker et al., 2016).

[7] Researchers will also want to bear in mind larger potential problems with pretest assessments, such as pretest sensitization effects. The classic discussion is Campbell and Stanley, (1963, e.g., p. 18); see also Shadish et al., (2002, e.g., p. 260).

Challenges for research synthesis

The presence of both kinds of research design in the persuasion-effects literature can invite some misunderstandings, especially where research synthesis is of interest. As emphasized above, these two designs address fundamentally different research questions and hence results from the two designs cannot meaningfully be combined. In meta-analytic research, uninterpretable results are created if the meta-analysis combines effect sizes from studies with two advocacy messages and effect sizes from studies comparing an advocacy message against a no-advocacy-message condition.

Appropriate meta-analyses

Persuasion meta-analyses have often been conducted in ways that respect the difference between these two designs. For example, many meta-analyses have focused exclusively on studies using two-advocacy-message designs, examining such message contrasts as narrative vs. non-narrative (Shen et al., 2015), metaphorical vs. literal (Sopory & Dillard, 2002; Van Stee, 2018), gain-framed vs. loss-framed (Gallagher & Updegraff, 2012; O'Keefe & Jensen, 2006), humorous vs. non-humorous (Walter et al., 2018), strong vs. weak fear appeals (Witte & Allen, 2000), text-only warnings vs. pictorial warnings (Noar et al., 2016, 2020), and so on.

Similarly, a number of meta-analyses have focused exclusively on studies using no-advocacy-message control-condition designs. For example, Carey et al. (2013) reported a meta-analysis of driver-behavior studies that compared threat appeals against a no-advocacy-message control condition. In Rode et al. (2021) meta-analysis of studies of climate-change messages, the included studies compared a climate-change message condition against a no-advocacy-message comparison condition. Braddock and Dillard (2016) meta-analyzed experiments that compared a narrative message against a no-advocacy message control condition. (See, similarly, Nisa et al., 2019; Rohde et al., 2021; Steinmetz et al., 2016)

And some meta-analytic reviews have included both kinds of designs, but have – appropriately— analyzed data from the two kinds of design separately. For example, Abrahamse and Steg's (2013) meta-analysis of the effects of social influence interventions on resource conservation appropriately distinguished (and analyzed separately) two sorts of effect size, one describing the effect of social influence approaches as compared to a no-intervention control condition, the other describing the relative effectiveness of social influence approaches as compared to other kinds of intervention. Similarly, Wang and Miller's (2020) meta-analysis of just-in-time tailored health interventions distinguished comparisons of such treatments to waitlist controls and comparisons to alternative kinds of treatments. (For examples of systematic reviews – not meta-analyses – that similarly distinguished studies on the basis of the comparison condition, see Frascella et al., 2020; Okuhara et al., 2023; Perrier & Martin Ginis, 2017)

Inappropriate meta-analyses

However, other meta-analytic reviews have included effect sizes from both kinds of designs without recognizing that the resulting collection of effect sizes cannot be interpreted straightforwardly. For example, Tannenbaum et al.'s (2015) meta-analysis compared fear appeals against a combination of a number of different control conditions, including lower-fear-appeal messages and no-message control conditions: "The comparison group could have been a group that was not exposed to any message, a group that was exposed to a message that was not designed to induce fear, or a message that was designed to induce less fear than the treatment group's message" (Tannenbaum et al., 2015; see, similarly, Bigsby & Albarracín, 2022, p. 247).

So some of those effect sizes represented an answer to the question of whether a fear-appeal message was more persuasive than having no advocacy message on the advocacy topic, and other effect sizes represented an answer to the question of whether a high-fear-appeal message was more persuasive than a low-fear-appeal message. This mixture of comparison conditions thus made the overall mean effect sizes uninterpretable. For example, a positive mean effect size (favoring fear-appeal messages over the control conditions) computed across such a combination of designs would be consistent with there being no difference between the high-fear and low-fear message conditions but there being a substantial difference between the fear-message and no-message conditions; with such a pattern of effects, one would want to use an advocacy message but it wouldn't matter whether it was high-fear or low-fear. Alternatively, such a mean effect size might reflect high-fear messages being more persuasive than low-fear messages; if those were the results, then advocates would want to be sure to deploy high-fear messages rather than low-fear messages. The larger point is: The overall mean effect sizes cannot be meaningfully interpreted because the mixture of comparison conditions means that the effect sizes do not all address the same research question. (For a report of a separate subset of those cases concerning specifically the high-fear-vs.-low-fear message contrast, see White & Albarracín, 2018.)

A more subtle version of this mistake appears in meta-analyses that treat the variation in comparison condition as a "moderator" variable. As examples: Yang's (2017) review of social-networking-site-based health behavior interventions included both studies comparing such interventions to a no-intervention control condition and studies comparing such interventions to an alternative intervention; control condition was treated as a moderator (see p. 229). Li et al.'s (2019) meta-analysis of studies of visualized nutrition education interventions included both studies comparing such an intervention to a no-intervention condition and studies comparing such an intervention to other kinds of intervention; control condition was treated as a moderator (see Li et al. 2019, p. 1981). In a review of online health behavior change interventions, Cugelman et al. (2011) included both studies with a no-intervention control condition (waitlist and placebo controls) and studies that compared online interventions with either low-tech (website) or print interventions; comparison condition was treated as a moderator (see Cugelman et al. 2011, p. 10). (For other examples of this practice, see Head et al., 2013; Huang & Shen, 2016; Lustria et al., 2013; Noar et al., 2007, 2009, 2010; Rhodes et al., 2020; Vereen et al., 2023; Yang et al., 2020)

But the difference in designs – the difference in comparison conditions – is not appropriately treated as a moderator because the two designs generate fundamentally different effect sizes. The customary role for a meta-analytic moderator variable arises when a set of effect sizes all address the same research question, but some property of those effect sizes (the possible moderating factor) varies across studies and hence the collection of effect sizes is subdivided on the basis of differing values of the moderating factor. Analysis of the different subsets indicates how the same research question might receive different answers depending on the value of the moderator.

In the circumstance under discussion, however, the collected effect sizes do not all address the same research question. Rather, the two kinds of effect size describe different relationships, arising from two very different comparisons-one between an advocacy message and a no-advocacy-message control condition, the other between two advocacy messages. Because the two sets of effect sizes address different questions ("is it more persuasive to use an advocacy message or not?" and "Is it more persuasive to use message A or message B?"), the underlying design difference is not appropriately described or analyzed as a moderator – and the overall mean effect sizes based on combining the two kinds of effect size are not meaningfully interpretable.

When comparison condition is treated as a moderator, one might think that if the moderator effect is not significant (i.e., the two mean effect sizes do not differ significantly), then combining the effect sizes is appropriate (see, e.g., Noar et al., in press). But this reasoning is mistaken. A finding that the size of the mean difference between message A and message B was statistically indistinguishable from the size of the mean difference between message A and a no-advocacy message control condition would not imply that the two effect sizes mean the same thing; it would not show that the two sorts of effect sizes are appropriately combined.

By way of illustration: Imagine a meta-analyst combined effect sizes from (a) studies comparing the persuasive effects of humorous and non-humorous messages and (b) studies comparing the persuasive effects of messages varying in the depicted severity (high vs. low) of a threat, reporting an overall mean effect size across the two kinds of study. When faced with the objection "but those different effect sizes can't appropriately be combined because they describe different relationships," imagine the reply being "yes, combining those effect sizes *is* legitimate – because the mean effect size in the first set of studies (mean $r = .12$, 95% CI [.06, .18]; Walter et al., 2018) is not statistically significantly different from the mean effect size in the second set of studies (mean $r = .13$, 95% CI [.07, .18]; de Hoog et al., 2007). Therefore it is entirely appropriate to combine them."[8]

Obviously that's not a good argument for combining those effect sizes. Put more generally, the fact that mean effect size X and mean effect size Y are not statistically significantly different says nothing about the meaning of the effect sizes being analyzed, and is not a justification for combining effect sizes. The only suitable justification for combining effect sizes is that the effect sizes all speak to the same research question – and whether they all speak to the same research question cannot be decided by seeing whether the two mean effect sizes are statistically significantly different.

In the applications of interest here, effect sizes from studies using two-advocacy-message designs and effect sizes from studies using no-advocacy-message-control designs are not appropriately combined because the two kinds of effect size describe different relationships. Even if, in a given meta-analysis, the two mean effect sizes are not statistically significantly different, the two mean effect sizes nevertheless describe different relationships.

Conclusion

In persuasion message effects research, two kinds of research design are common, with different comparisons being made. One compares the persuasiveness of two different advocacy messages on the same topic. The other compares the persuasiveness of an advocacy message against a no-advocacy-message control condition. These two designs underwrite different kinds of claims, but each design – and its corresponding claims – is prone to misunderstanding and confusion, especially in the context of meta-analytic research synthesis. Designs that combine these two designs offer significant advantages, but these are accompanied by related complexities in results. This article has tried to sort out the relevant issues in the service of better alignment between claims and evidence in persuasive message effects research.

Disclosure statement

I have no known conflict of interest to disclose. Thanks to the journal's reviewers and to Hans Hoeken for useful commentary.

Notes on contributor

Daniel J. O'Keefe is the Owen L. Coon Professor Emeritus in the Department of Communication Studies at Northwestern University

ORCID

Daniel J. O'Keefe http://orcid.org/0000-0003-1594-8892

References

Abdel-Raheem, A., & Alkhammash, R. (2022). "To get or not to get vaccinated against COVID-19": Saudi women, vaccine hesitancy, and framing effects. *Discourse & Communication*, *16*(1), 21–36. https://doi.org/10.1177/17504813211043724

[8] The results of de Hoog et al. (2007) were originally reported using d as the effect size index; the results have been converted to r for easy comparison with the results of Walter et al. (2018).

Abrahamse, W., & Steg, L. (2013). Social influence approaches to encourage resource conservation: A meta-analysis. *Global Environmental Change*, *23*(6), 1773–1785. https://doi.org/10.1016/j.gloenvcha.2013.07.029

Allen, M. (1991). Meta-analysis comparing the persuasiveness of one-sided and two-sided messages. *Western Journal of Speech Communication*, *55*(4), 390–404. https://doi.org/10.1080/10570319109374395

Andrews, M. E., Mattan, B. D., Richards, K., Moore-Berg, S. L., & Falk, E. B. (2022). Using first-person narratives about healthcare workers and people who are incarcerated to motivate helping behaviors during the COVID-19 pandemic. *Social Science & Medicine*, *299*, 114870. https://doi.org/10.1016/j.socscimed.2022.114870

Anghelcev, G., & Sar, S. (2014). In the mood for [the right kind of] social marketing communication: How congruity between consumer mood and message framing influences intentions to recycle. *Journal of Social Marketing*, *4*(1), 38–57. https://doi.org/10.1108/JSOCM-04-2013-0025

Anker, A. E., Feeley, T. H., McCracken, B., & Lagoe, C. A. (2016). Measuring the effectiveness of mass-mediated health campaigns through meta-analysis. *Journal of Health Communication*, *21*(4), 439–456. https://doi.org/10.1080/10810730.2015.1095820

Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial. *Criminology*, *50*(1), 27–69. https://doi.org/10.1111/j.1745-9125.2011.00256.x

Bigsby, E., & Albarracín, D. (2022). Self- and response efficacy information in fear appeals: A meta-analysis. *Journal of Communication*, *72*(2), 241–263. https://doi.org/10.1093/joc/jqab048

Bleakley, A., Jordan, A., Mallya, G., Hennessy, M., & Piotrowski, J. T. (2018). Do you know what your kids are drinking? Evaluation of a media campaign to reduce consumption of sugar-sweetened beverages. *American Journal of Health Promotion*, *32*(6), 1409–1416. https://doi.org/10.1177/0890117117721320

Braddock, K., & Dillard, J. P. (2016). Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Communication Monographs*, *83*(4), 446–467. https://doi.org/10.1080/03637751.2015.1128555

Busso, D. S., Volmert, A., & Kendall-Taylor, N. (2019). Reframing aging: Effect of a short-term framing intervention on implicit measures of age bias. *Journals of Gerontology, Series B: Psychological Sciences & Social Sciences*, *74*(4), 559–564. https://doi.org/10.1093/geronb/gby080

Byrne, S., & Hart, P. S. (2009). The boomerang effect: A synthesis of findings and a preliminary theoretical framework. *Annals of the International Communication Association*, *33*(1), 3–37. https://doi.org/10.1080/23808985.2009.11679083

Calabrese, C., & Zhang, J. (2019). Inferring norms from numbers: Boomerang effects of online virality metrics on normative perceptions and behavioral intention. *Telematics and Informatics*, *45*, 101279. article no. 101279. https://doi.org/10.1016/j.tele.2019.101279

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Carey, R. N., McDermott, D. T., Sarma, K. M., & Rustichini, A. (2013). The impact of threat appeals on fear arousal and driver behavior: A meta-analysis of experimental research 1990–2011. *PLos One*, *8*(5), e62821. https://doi.org/10.1371/journal.pone.0062821

Chong, D., & Druckman, J. (2010). Dynamic public opinion: Communication effects over time. *The American Political Science Review*, *104*(4), 663–680. https://doi.org/10.1017/S0003055410000493

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cugelman, B., Thelwall, M., & Dawes, P. (2011). Online interventions for social marketing health behavior change campaigns: A meta-analysis of psychological architectures and adherence factors. *Journal of Medical Internet Research*, *13*(1), e17. https://doi.org/10.2196/jmir.1367

de Hoog, N., Stroebe, W., & de Wit, J. B. F. (2007). The impact of vulnerability to and severity of a health risk on processing and acceptance of fear-arousing communications: A meta-analysis. *Review of General Psychology*, *11*(3), 258–285. https://doi.org/10.1037/1089-2680.11.3.258

Dijkstra, A., & Rotelli, V. (2022). Lowering red meat and processed meat consumption with environmental, animal welfare, and health arguments in Italy: An online experiment. *Frontiers in Psychology*, *13*, 877911. https://doi.org/10.3389/fpsyg.2022.877911

Dillard, J. P., Cruz, S. M., & Shen, L. (2023). Spillover effects of anti-sugar-sweetened beverage messages: From consumption decisions to policy preferences. *Social Science & Medicine*, *320*, 115693. https://doi.org/10.1016/j.socscimed.2023.115693

Divdar, M., Araban, M., Heydarabadi, A. B., Cheraghian, B., & Stein, L. (2021). Effectiveness of message-framing to improve oral health behaviors and dental plaque among pregnant women. *Archives of Public Health*, *79*(1), 117. https://doi.org/10.1186/s13690-021-00640-1

Fagan, J., Frye, V., Calixte, R., Jain, S., Molla, L., Lawal, A., Mosley, M. P., Greene, E., Mayer, K. H., & Zingman, B. S. (2020). "It's like Plan B but for HIV!" Design and evaluation of a media campaign to drive demand for PEP. *AIDS and Behavior*, *24*(12), 3337–3345. https://doi.org/10.1007/s10461-020-02906-1

Fesenfeld, L. P., Sun, Y., Wicki, M., & Bernauer, T. (2021). The role and limits of strategic framing for promoting sustainable consumption and policy. *Global Environmental Change*, *68*, 102266. https://doi.org/10.1016/j.gloenvcha.2021.102266

Fishbein, M., Hall-Jamieson, K., Zimmer, E., von Haeften, I., & Nabi, R. (2002). Avoiding the boomerang: Testing the relative effectiveness of antidrug public service announcements before a national campaign. *American Journal of Public Health*, *92*(2), 238–245. https://doi.org/10.2105/AJPH.92.2.238

Frascella, B., Oradini-Alacreu, A., Balzarini, F., Signorelli, C., Lopalco, P. L., & Odone, A. (2020). Effectiveness of email-based reminders to increase vaccine uptake: A systematic review. *Vaccine*, *38*(3), 433–443. https://doi.org/10.1016/j.vaccine.2019.10.089

Gallagher, K. M., & Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: A meta-analytic review. *Annals of Behavioral Medicine*, *43*(1), 101–116. https://doi.org/10.1007/s12160-011-9308-7

Gong, J., Zhang, Y., Gao, H., Wei, W., Lv, J., Liu, H., & Huang, Y. (2017). Choices regarding thrombolysis are modified by the way to transfer the messages. *Frontiers in Neurology*, *8*, 589. https://doi.org/10.3389/fneur.2017.00589

Gosnell, G. K. (2018). Communicating resourcefully: A natural field experiment on environmental framing and cognitive dissonance in going paperless. *Ecological Economics*, *154*, 128–144. https://doi.org/10.1016/j.ecolecon.2018.07.020

Head, K. J., Noar, S. M., Iannarino, N. T., & Harrington, N. G. (2013). Efficacy of text messaging-based interventions for health promotion: A meta-analysis. *Social Science & Medicine*, *97*, 41–48. https://doi.org/10.1016/j.socscimed.2013.08.003

Highhouse, S. (2009). Designing experiments that generalize. *Organizational Research Methods*, *12*(3), 554–566. https://doi.org/10.1177/1094428107300396

Huang, Y., & Shen, F. (2016). Effects of cultural tailoring on persuasion in cancer communication: A meta-analysis. *Journal of Communication*, *66*(4), 694–715. https://doi.org/10.1111/jcom.12243

Igartua, J. -J. (2010). Identification with characters and narrative persuasion through fictional feature films. *Communications*, *35*(4), 347–373. https://doi.org/10.1515/COMM.2010.019

Insko, C. A. (1964). Primacy versus recency in persuasion as a function of the timing of arguments and measures. *Journal of Abnormal & Social Psychology*, *69*(4), 381391. https://doi.org/10.1037/h0042765

Jackson, S., & Jacobs, S. (1983). Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research*, *9*(2), 169181. https://doi.org/10.1111/j.1468-2958.1983.tb00691.x

Katz, S. J., Shi, W., Erkkinen, M., Lindgren, B., & Hatsukami, D. (2020). High school youth and e-cigarettes: The influence of modified risk statements and flavors on e-cigarette packaging. *American Journal of Health Behavior*, *44*(2), 130–145. https://doi.org/10.5993/AJHB.44.2.2

Kim, S., Levine, T. R., & Allen, M. (2017). The intertwined model of reactance for resistance and persuasive boomerang. *Communication Research*, *44*(7), 931–951. https://doi.org/10.1177/0093650214548575

Kite, J., Gale, J., Grunseit, A., Bellew, W., Li, V., Lloyd, B., Maxwell, M., Vineburg, J., & Bauman, A. (2018). Impact of the make healthy normal mass media campaign (Phase 1) on knowledge, attitudes and behaviours: A cohort study. *Australian and New Zealand Journal of Public Health*, *42*(3), 269–276. https://doi.org/10.1111/1753-6405.12779

Kohavi, R., & Longbotham, R. (2017). Online controlled experiments and A/B testing. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning and data mining*. Springer. https://doi.org/10.1007/978-1-4899-7687-1_891

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press.

Legate, N., Nguyen, T. -V., Weinstein, N., Moller, A., Legault, L., Vally, Z., Tajchman, Z., Zsido, A. N., Zrimsek, M., Chen, Z., Ziano, I., Gialitaki, Z., Ceary, C. D., Jang, Y., Lin, Y., Kunisato, Y., Yamada, Y., Xiao, Q., Jiang, X. . . . Psychological Science Accelerator Self-Determination Theory Collaboration. (2022). A global experiment on motivating social distancing during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *119*(22), e2111091119. https://doi.org/10.1073/pnas.2111091119

Liang, Y., Henderson, L. K., & Kee, K. F. (2018). Running out of water! Developing a message typology and evaluating message effects on attitude toward water conservation. *Environmental Communication*, *12*(4), 541–557. https://doi.org/10.1080/17524032.2017.1288648

Lienemann, B. A., & Siegel, J. T. (2018). Increasing help-seeking outcomes among people with elevated depressive symptomatology with public service announcements: An examination of functional matching and message sidedness. *Journal of Health Communication*, *23*(1), 28–39. https://doi.org/10.1080/10810730.2017.1396630

Li, X., Huang, Y., Yin, R., Pan, C., Cai, Y., & Wang, Z. (2019). Visualized nutrition education and dietary behavioral change: A systematic review and meta-analysis. *Critical Reviews in Food Science and Nutrition*, *59*(12), 1976–1985. https://doi.org/10.1080/10408398.2018.1469466

Lustria, M. L. A., Noar, S. M., Cortese, J., Van Stee, S. K., Glueckauf, R. L., & Lee, J. (2013). A meta-analysis of web-delivered tailored health behavior change interventions. *Journal of Health Communication*, *18*(9), 1039–1069. https://doi.org/10.1080/10810730.2013.768727

Maglione, M. A., Stone, E. G., & Shekelle, P. G. (2002). Mass mailings have little effect on utilization of influenza vaccine among Medicare beneficiaries. *American Journal of Preventive Medicine*, *23*(1), 43–46. https://doi.org/10.1016/s0749-3797(02)00443-9

Majmundar, A., Cornelis, E., & Moran, M. B. (2020). Examining the vulnerability of ambivalent young adults to e-cigarette messages. *Health Marketing Quarterly*, *37*(1), 73–88. https://doi.org/10.1080/07359683.2019.1680119

Ma, H., Kieu, T.K. -T., Ribisl, K. M., & Noar, S. M. (in press). Do vaping prevention messages impact adolescents and young adults? A meta-analysis of experimental studies. *Health Communication*. https://doi.org/10.1080/10410236.2023.2185578

Morley, B. C., Niven, P. H., Dixon, H. G., Swanson, M. G., McAleese, A. B., & Wakefield, M. A. (2018). Controlled cohort evaluation of the *LiveLighter* mass media campaign's impact on adults' reported consumption of sugar-sweetened beverages. *British Medical Journal Open*, *8*(4), e019574. https://doi.org/10.1136/bmjopen-2017-019574

Myers, T. A., Nisbet, M. C., Maibach, E. W., & Leiserowitz, A. A. (2012). A public health frame arouses hopeful emotions about climate change. *Climatic Change*, *113*(3–4), 1105–1112. https://doi.org/10.1007/s10584-012-0513-6

Nabi, R. L. (2016). Laughing in the face of fear (of disease detection): Using humor to promote cancer self-examination behavior. *Health Communication*, *31*(7), 873–883. https://doi.org/10.1080/10410236.2014.1000479

Nelson, K. M., Bauer, K. M., & Partelow, S. (2021). Informational nudges to encourage pro-environmental behavior: Examining differences in message framing and human interaction. *Frontiers in Communication*, *5*, 610186. https://doi.org/10.3389/fcomm.2020.610186

Nicolla, S., & Lazard, A. J. (in press). Social media communication about sexual violence may backfire: Online experiment with young men. *Journal of Health Communication*. https://doi.org/10.1080/10810730.2023.2174214

Nisa, C. F., Bélanger, J. J., Schumpe, B. M., & Faller, D. G. (2019). Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. *Nature Communications*, *10*(1), 4545. https://doi.org/10.1038/s41467-019-12457-2

Noar, S. M., Benac, C. N., & Harris, M. S. (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin*, *133*(4), 673–693. https://doi.org/10.1037/0033-2909.133.4.673

Noar, S. M., Black, H. G., & Pierce, L. B. (2009). Efficacy of computer technology-based HIV prevention interventions: A meta-analysis. *AIDS*, *23*(1), 107–115. https://doi.org/10.1097/QAD.0b013e32831c5500

Noar, S. M., Hall, M. G., Francis, D. B., Ribisl, K. M., Pepper, J. K., & Brewer, N. T. (2016). Pictorial cigarette pack warnings: A meta-analysis of experimental studies. *Tobacco Control*, *25*(3), 341–354. https://doi.org/10.1136/tobaccocontrol-2014-051978

Noar, S. M., Ma, H., Kieu, T.K. -T., & Ribisl, K. M. (in press). Can extant vaping prevention message experiments tell us something about what works? A response to O'Keefe. *Health Communication*. https://doi.org/10.1080/10410236.2023.2212195

Noar, S. M., Pierce, L. B., & Black, H. G. (2010). Can computer-mediated interventions change theoretical mediators of safer sex? A meta-analysis. *Human Communication Research*, *36*(3), 261–297. https://doi.org/10.1111/j.1468-2958.2010.01376.x

Noar, S. M., Rohde, J. A., Barker, J. O., Hall, M. G., & Brewer, N. T. (2020). Pictorial cigarette pack warnings increase some risk appraisals but not risk beliefs: A meta-analysis. *Human Communication Research*, *46*(2–3), 250–272. https://doi.org/10.1093/hcr/hqz016

O'Keefe, D. J. (2017). Misunderstandings of effect sizes in message effects research. *Communication Methods and Measures*, *11*(3), 210–219. https://doi.org/10.1080/19312458.2017.1343812

O'Keefe, D. J. (in press). Commentary on "Do vaping prevention messages impact adolescents and young adults? A meta-analysis of experimental studies. *Health Communication*. https://doi.org/10.1080/10410236.2023.2212467

O'Keefe, D. J., & Jensen, J. D. (2006). The advantages of compliance or the disadvantages of noncompliance? A meta-analytic review of the relative persuasive effectiveness of gain-framed and loss-framed messages. *Annals of the International Communication Association*, *30*(1), 1–43. https://doi.org/10.1080/23808985.2006.11679054

Okuhara, T., Kagawa, Y., Okada, H., Tsunezumi, A., & Kiuchi, T. (2023). Intervention studies to encourage HPV vaccination using narrative: A scoping review. *Patient Education & Counseling*, *111*, 107689. https://doi.org/10.1016/j.pec.2023.107689

Perrier, M. -J., & Martin Ginis, K. A. (2017). Narrative interventions for health screening behaviours: A systematic review. *Journal of Health Psychology*, *22*(3), 375–393. https://doi.org/10.1177/1359105315603463

Powell-Jackson, T., Fabbri, C., Dutt, V., Tougher, S., Singh, K., & Tumwine, J. K. (2018). Effect and cost-effectiveness of educating mothers about childhood DPT vaccination on immunisation uptake, knowledge, and perceptions in Uttar Pradesh, India: A randomised controlled trial. *PLoS Medicine*, *15*(3), e1002519. https://doi.org/10.1371/journal.pmed.1002519

Reeves, B., Yeykelis, L., & Cummings, J. J. (2016). The use of media in media psychology. *Media Psychology*, *19*(1), 49–71. https://doi.org/10.1080/15213269.2015.1030083

Reger, B., Wootan, M. G., & Booth-Butterfield, S. (1999). Using mass media to promote healthy eating: A community-based demonstration project. *Preventive Medicine*, *29*(5), 414–421. https://doi.org/10.1006/pmed.1998.0570

Rhodes, N., Shulman, H. C., & McClaren, N. (2020). Changing norms: A meta-analytic integration of research on social norms appeals. *Human Communication Research*, *46*(2–3), 161–191. https://doi.org/10.1093/hcr/hqz023

Richter, I., Thøgersen, J., & Klöckner, C. A. A social norms intervention going wrong: Boomerang effects from descriptive norms information. (2018). *Sustainability*, *10*(8), 2848. article no. 2848. https://doi.org/10.3390/su10082848

Ringold, D. J. (2002). Boomerang effect in response to public health interventions: Some unintended consequences in the alcoholic beverage market. *Journal of Consumer Policy*, *25*(1), 27–63. https://doi.org/10.1023/A:1014588126336

Ritchie, P. D., Jenkins, M., & Cameron, P. A. (2000). A telephone call reminder to improve outpatient attendance in patients referred from the emergency department: A randomised controlled trial. *Australian and New Zealand Journal of Medicine*, *30*(5), 585–592. https://doi.org/10.1111/j.1445-5994.2000.tb00860.x

Rode, J. B., Dent, A. L., Benedict, C. N., Brosnahan, D. B., Martinez, R. L., & Ditto, P. H. (2021). Influencing climate change attitudes in the United States: A systematic review and meta-analysis. *Journal of Environmental Psychology*, *76*, 101623. https://doi.org/10.1016/j.jenvp.2021.101623

Rode, J. B., Dent, A. L., & Ditto, P. H. (2023). Climate change consensus messages may cause reactance in conservatives, but there is no meta-analytic evidence that they backfire. *Environmental Communication*, *17*(1), 60–66. https://doi.org/10.1080/17524032.2022.2101501

Rohde, J. A., Barker, J. O., & Noar, S. M. (2021). Impact of eHealth technologies on patient outcomes: A meta-analysis of chronic gastrointestinal illness interventions. *Translational Behavioral Medicine*, *11*(1), 1–10. https://doi.org/10.1093/tbm/ibz166

Ryoo, Y., & Kim, W. (2023). Using descriptive and injunctive norms to encourage COVID-19 social distancing and vaccinations. *Health Communication*, *38*(4), 732–741. https://doi.org/10.1080/10410236.2021.1973702

Sato, R., & Takasaki, Y. (2021). Backfire effect of salient information on vaccine take-up experimental evidence from scared-straight intervention in rural northern Nigeria. *Human Vaccines & Immunotherapeutics*, *17*(6), 1703–1713. https://doi.org/10.1080/21645515.2020.1836917

Schmid, P., & Betsch, C. (2022). Benefits and pitfalls of debunking interventions to counter mRNA vaccination misinformation during the COVID-19 pandemic. *Science Communication*, *44*(5), 531–558. https://doi.org/10.1177/10755470221129608

Seiter, C. R., Zhao, X., & Rossheim, M. E. (in press). Valence of YouTube comments and college student reactions towards electronic nicotine product counter-marketing messages. *Health Communication*. https://doi.org/10.1080/10410236.2023.2177795

Severson, A. W., & Coleman, E. A. (2015). Moral frames and climate change policy attitudes. *Social Science Quarterly*, *96*(5), 1277–1290. https://doi.org/10.1111/ssqu.12159

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth.

Shen, F., Sheer, V. C., & Li, R. (2015). Impact of narratives on persuasion in health communication: A meta-analysis. *Journal of Advertising*, *44*(2), 105–113. https://doi.org/10.1080/00913367.2015.1018467

Simpson, A. (2017). The misdirection of public policy: Comparing and combining standardised effect sizes. *Journal of Education Policy*, *32*(4), 450–466. https://doi.org/10.1080/02680939.2017.1280183

Simpson, B., White, K., Laran, J., Morwitz, V., & Shavitt, S. (2018). When public recognition for charitable giving backfires: The role of independent self-construal. *The Journal of Consumer Research*, *44*(6), 1257–1273. https://doi.org/10.1093/jcr/ucx101

Skurka, C. (2019). Communicating inequalities to enhance support for obesity-prevention policies: The role of social comparisons, age frames, and emotion. *Health Communication*, *34*(2), 227–237. https://doi.org/10.1080/10410236.2017.1405477

Slater, J. S., Parks, M. J., Nelson, C. L., & Hughes, K. D. (2018). The efficacy of direct mail, patient navigation, and incentives for increasing mammography and colonoscopy in the Medicaid population: A randomized controlled trial. *Cancer Epidemiology Biomarkers & Prevention*, *27*(9), 1047–1056. https://doi.org/10.1158/1055-9965.EPI-18-0038

Slater, M. D., Peter, J., & Valkenburg, P. M. (2015). Message variability and heterogeneity: A core challenge for communication research. *Annals of the International Communication Association*, *39*(1), 3–31. https://doi.org/10.1080/23808985.2015.11679170

Sontag, J., Manderski, M. T. B., Hammond, D., & Wackowski, O. A. (2019). US young adults' perceived effectiveness of draft pictorial e-cigarette warning labels. *Tobacco Control*, *28*(e1), e49–51. https://doi.org/10.1136/tobaccocontrol-2018-054802

Sopory, P., & Dillard, J. P. (2002). The persuasive effects of metaphor: A meta-analysis. *Human Communication Research*, *28*(3), 382–419. https://doi.org/10.1111/j.1468-2958.2002.tb00813.x

Steinmetz, H., Knappstein, M., Ajzen, I., Schmidt, P., & Kabst, R. (2016). How effective are behavior change interventions based on the theory of planned behavior? A three-level meta-analysis. *Zeitschrift für Psychologie*, *224*(3), 216–233. https://doi.org/10.1027/2151-2604/a000255

Stibe, A., & Cugelman, B. (2016). Persuasive backfiring: When behavior change interventions trigger unintended negative outcomes. In A. Meschtscherjakov, B. D. Ruyter, V. Fuchsberger, M. Murer, & M. Tscheligi (Eds.), *Persuasive technology: 11th international conference, PERSUASIVE 2016* (pp. 65–77). Springer. https://doi.org/10.1007/978-3-319-31510-2_6

Tan, A. S. L., Lee, C. -J., Nagler, R. H., & Bigman, C. A. (2017). To vape or not to vape? Effects of exposure to conflicting news headlines on beliefs about harms and benefits of electronic cigarette use: Results from a randomized controlled experiment. *Preventive Medicine*, *105*, 97–103. https://doi.org/10.1016/j.ypmed.2017.08.024

Tannenbaum, M. B., Hepler, J., Zimmerman, R. S., Saul, L., Jacobs, S., Wilson, K., & Albarracín, D. (2015). Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. *Psychological Bulletin*, *141*(6), 1178–1204. https://doi.org/10.1037/a0039729

Thacker, I., Sinatra, G. M., Muis, K. R., Danielson, R. W., Pekrun, R., Winne, P. H., & Chevrier, M. (2020). Using persuasive refutation texts to prompt attitudinal and conceptual change. *Journal of Educational Psychology*, *112*(6), 1085–1099. https://doi.org/10.1037/edu0000434

Thorson, E., Wicks, R., & Leshner, G. (2012). Experimental methodology in journalism and mass communication research. *Journalism & Mass Communication Quarterly*, *89*(1), 112–124. https://doi.org/10.1177/1077699011430066

Van Stee, S. K. (2018). Meta-analysis of the persuasive effects of metaphorical vs. literal messages. *Communication Studies*, *69*(5), 545–566. https://doi.org/10.1080/10510974.2018.1457553

Van Zant, A. B., & Moore, D. A. (2015). Leaders' use of moral justifications increases policy support. *Psychological Science*, *26*(6), 934–943. https://doi.org/10.1177/0956797615572909

Vereen, R. N., Kurtzman, R., & Noar, S. M. (2023). Are social media interventions for health behavior change efficacious among populations with health disparities? A meta-analytic review. *Health Communication*, *38*(1), 133–140. https://doi.org/10.1080/10410236.2021.1937830

Villanti, A. C., West, J. C., Mays, D., Donny, E. C., Cappella, J. N., & Strasser, A. A. (2019). Impact of brief nicotine messaging on nicotine-related beliefs in a U.S. sample. *American Journal of Preventive Medicine*, *57*(4), e135–142. https://doi.org/10.1016/j.amepre.2019.05.015

Walter, N., Cody, M. J., Xu, L. Z., & Murphy, S. T. (2018). A priest, a rabbi, and a minister walk into a bar: A meta-analysis of humor effects on persuasion. *Human Communication Research*, *44*(4), 343–373. https://doi.org/10.1093/hcr/hqy005

Wang, L., & Miller, L. C. (2020). Just-in-the-moment adaptive interventions (JITAI): A meta-analytical review. *Health Communication*, *35*(12), 1531–1544. https://doi.org/10.1080/10410236.2019.1652388

Weiler, B., Moyle, B. D., Wolf, I. D., de Bie, K., & Torland, M. (2017). Assessing the efficacy of communication interventions for shifting public perceptions of park benefits. *Journal of Travel Research*, *56*(4), 468–481. https://doi.org/10.1177/0047287516646472

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality & Social Psychology Bulletin*, *25*(9), 1115–1125. https://doi.org/10.1177/01461672992512005

White, B. X., & Albarracín, D. (2018). Investigating belief falsehood: Fear appeals do change behavior in experimental laboratory studies: A commentary on Kok et al. *Health Psychology Review*, *12*(2), 147–150. https://doi.org/10.1080/17437199.2018.1448292

Winett, L. B., Niederdeppe, J., Xu, Y., Gollust, S. E., & Fowler, E. F. (2021). When "tried and true" advocacy strategies backfire: Narrative messages can undermine state legislator support for early childcare policies. *Journal of Public Interest Communications*, *5*(1), 45–77. https://doi.org/10.32473/jpic.v5.i1.p45

Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health programs. *Health Education & Behavior*, *27*(5), 591–615. https://doi.org/10.1177/109019810002700506

Yang, Q. (2017). Are social networking sites making health behavior change interventions more effective? A meta-analytic review. *Journal of Health Communication*, *22*(3), 223–233. https://doi.org/10.1080/10810730.2016.1271065

Yang, Q., Millette, D., Zhou, C., Beatty, M., Carcioppolo, N., & Wilson, G. (2020). The effectiveness of interactivity in improving mediating variables, behaviors and outcomes of web-based health interventions: A meta-analytic review. *Health Communication*, *35*(11), 1334–1348. https://doi.org/10.1080/10410236.2019.1631992