Evidence of Heterogeneity in the Direction and Magnitude of Narrative Effects on Transportation and Counterarguing Across Three Independent Samples

YIWEI XU[1]
Cornell University, USA

LIANA B. WINETT
Oregon Health & Science University–Portland State University
School of Public Health, USA

JEFF NIEDERDEPPE
Cornell University, USA

We report on 3 studies that reexamine two well-studied mechanisms of narrative persuasion, transportation and counterarguing, with two secondary data analyses and a preregistered follow-up experiment featuring identical messages and measures. This project employed internal replication, an undervalued method in communication research, to explore the stability of findings across three samples—a large national sample of U.S. adults (Study 1, secondary data analysis), a sample of U.S. state legislators (Study 2, secondary data analysis), and a convenience sample from Prolific, an online crowdsourcing platform (Study 3, preregistered follow-up study). Results show substantial evidence of treatment effect heterogeneity: a narrative message produced significantly less transportation than a nonnarrative message in Study 1, but the same narrative produced significantly more transportation in Study 3. The narrative also produced more counterarguing in Study 1 and Study 2, but not in Study 3. Indirect effects models revealed different patterns across the 3 studies. Such heterogeneity of the exact same treatment across different samples and populations highlights the need to qualify broad claims of generalizability of seemingly well-established narrative mechanisms.

Keywords: heterogeneity treatment effect, replication, narrative persuasion, transportation, counterarguing

Yiwei Xu: yx497@cornell.edu
Liana B. Winett: lwinett@pdx.edu
Jeff Niederdeppe: jdn56@cornell.edu
Date submitted: 2021-02-22

It is well established that narratives (stories) can persuade audiences about various issues (Braddock & Dillard, 2016). Some scholars approach narrative persuasion research by comparing different kinds of narratives with one another (e.g., de Graaf, Sanders, & Hoeken, 2016), while others compare narratives to nonnarrative messages. While both approaches can yield relevant insights into narrative processing, the current article compares narratives to nonnarrative messages to better understand the conditions under which narratives invite different patterns of cognitions that enhance or impede persuasion. This is a critical question for strategic communication. Transportation and limiting counterarguing are often cited as critical mechanisms that explain the process of narrative persuasion (Green & Brock, 2000). Transportation is associated with message-targeted outcomes like beliefs and attitudes (meta-analytic $r = .32$; Tukachinsky & Stokunaga, 2013) and narratives tend to produce less counterarguing than nonnarrative messages (meta-analytic $r = -.13$; Ratcliff & Sun, 2020), though no published studies to date report an average effect size of narratives on transportation.

While meta-analyses can help to synthesize previous empirical research, recent studies emphasize the degree of unresolved heterogeneity in communication research (Rains, Matthes, & Palomares, 2020). There remains a critical need for replication studies as the field grapples with publication biases in favoring statistically significant results (Keating & Totzkay, 2019; Sun & Pan, 2020). The current article uses internal replication (Dienlin et al., 2020; Lewis, 2020) with two secondary data analyses and a preregistered follow-up experiment using identical messages and measures, to reexamine two assumed mechanisms of narrative persuasion, transportation and counterarguing. We test the stability of findings across three samples—a large national sample of U.S. adults (Study 1, a secondary analysis of variables not analyzed in the original published study), a sample of U.S. state legislators (Study 2, another secondary analysis of variables not analyzed in the original published study), and a convenience sample from an online crowdsourcing platform (Study 3, an original preregistered follow-up study).

Replication and Heterogeneous Treatment Effects in Communication Research

Scholars across academic disciplines have called for replication and open science practices, which aim at improving reproducibility, replicability, and generalizability of findings in social science (Dienlin et al., 2020). Conceptual replication refers to the repetition of previous hypotheses/studies with different operationalizations or methods. Direct replication refers to the repetition of a previous study with the same operationalizations and data analysis methods (Dienlin et al., 2020; Schmidt, 2009). Direct replication studies are not often published in social science because the value of such replication is rarely overtly acknowledged: Social sciences value the discovery of new facts more than replication (Lindsay & Ehrenberg, 1993; Schmidt, 2009). Indeed, a publication trends analysis of studies from 2007–16 revealed that only one in every seven published communication journal articles was framed as a replication attempt, most of which were conceptual replications (Keating & Totzkay, 2019). Direct replications appear to be underused and undervalued in communication, but they can fill a significant gap by helping to increase the verifiability of existing knowledge in our field (McEwan, Carpenter, & Westerman, 2018).

There is a long and robust history of meta-analysis in the communication field, particularly in the area of message effects research (see O'Keefe, 2013; Rains, Levine, & Weber, 2018). A pair of recent summaries of 60 years of quantitative communication research revealed that most manipulations have

generated small-to-medium-sized effects (O'Keefe & Hoeken, 2021; Rains et al., 2018). These analyses have revealed substantial between-study treatment heterogeneity, which often goes unexplained (Levine & Weber, 2020). Furthermore, some scholars go so far as to argue that the predicted direction of message effects are highly unstable, such that 95% prediction intervals from meta-analyses of message effect treatments almost always include a significant likelihood of message effects in the *opposite* direction of the average treatment effect (O'Keefe & Hoeken, 2021). Understanding the degree of treatment effect heterogeneity is essential for estimating the replicability of results (Coppock, Leeper, & Mullinix, 2018). There is a critical need to explore potential sources of between-study treatment heterogeneity. One strategy is to hold many of these factors constant (e.g., use the same research practices, measures, and research designs) in direct replications of studies across different populations and sampling strategies. In this article we help to fill these gaps through replication studies that shed light on heterogeneity in message effects in the context of narrative persuasion.

Narrative Persuasion and Hypothesized Mechanisms in Communication Research

A narrative message has been defined as "a representation of connected events and characters that has an identifiable structure, is bounded in space and time, and contains implicit or explicit messages about the topic being addressed" (Kreuter et al., 2007, p. 222). It is well established that narratives can achieve persuasive goals among audiences. This includes both long-form narratives (TV programs, films) and short-form narratives employed in strategic communication (product advertisements; 30-second health public service announcements (PSAs)). For example, Shen, Sheer, and Li's (2015) meta-analysis showed that health narratives are, on average, more effective than nonnarratives, especially when the messages advocate detection and prevention behaviors. The average effect size in the analysis, however, was very small ($r = .06$), and several types of stories (print-based narratives, for example) did not show an average persuasive advantage. Furthermore, in a subsequent reanalysis of Shen and colleagues' (2015) meta-analytic data, O'Keefe and Hoeken (2021) concluded there was substantial heterogeneity in narrative versus nonnarrative message effects, such that the 95% prediction interval (which specifies "the range of plausible future individual effect sizes"; p. 3) ran from $r = -.11$ to $r = .24$. Another meta-analysis concluded that narratives can promote story-consistent beliefs, attitudes, intentions, and behaviors relative to no message at all (Braddock & Dillard, 2016). Again, however, the average effect size was relatively small, depending on whether the outcome was beliefs ($r = .17$), attitudes ($r = .19$), intentions ($r = .17$), or behavior ($r = -.23$).

Theorists have long argued that narrative engagement is one potential mechanism of narrative persuasion (Moyer-Gusé, 2008; Slater & Rouner, 2002; Tukachinsky & Stokunaga, 2013). Transportation is among the most commonly studied types of narrative engagement (Ratcliff & Sun, 2020; Tukachinsky & Stokunaga, 2013). Transportation is defined by Green and Brock (2000) as "a convergent process, where all the person's mental systems and capacities become focused on the events occurring in the narrative" (p. 701). Meta-analyses have shown transportation is positively associated with story-consistent beliefs, attitudes, and intentions (average $r = .33$; van Laer, de Ruyter, Visconti, & Wetzels, 2014; average $r = .32$; Tukachinsky & Stokunaga, 2013), though no studies have estimated an average effect size for narratives versus nonnarratives in generating transportation. Other authors have argued that transportation is one of many potential mechanisms involved in narrative persuasion (e.g., Niederdeppe, Heley, & Barry, 2015).

Theorists further argue that narratives often achieve persuasion goals by reducing readers' resistance to a message (Dal Cin, Zanna, & Fong, 2004; Moyer-Gusé, 2008; Slater & Rouner, 2002). Resistance is a common reaction to a message in which a receiver perceives pressure for a change in attitudes or beliefs (Knowles & Linn, 2004). Investigating how narratives can overcome resistance to persuasion can be particularly helpful in understanding narrative processing (Moyer-Gusé & Nabi, 2009). Counterarguing is a common form of resistance to a message that has been well studied in the domain of narrative persuasion (Ratcliff & Sun, 2020). Counterarguing is characterized by the "generation of thoughts that dispute or are inconsistent with the persuasive argument" (Slater & Rouner, 2002, p. 180).

Narrative persuasion theories suggest that when audiences are absorbed into a narrative, they are less likely to counterargue against the message's intended persuasive goals (Green, 2006; Moyer-Gusé, 2008; Slater & Rouner, 2002). A recent meta-analysis found that narrative messages generated less counterarguing than nonnarrative messages ($r = -.13$), though there was substantial heterogeneity in effect sizes across studies (Ratcliff & Sun, 2020). Transportation was also associated with lower levels of persuasion resistance, an index that included measures of counterarguing ($r = -.15$). Recent evidence shows that narratives can also reduce resistance to health-related public policies, not just the promotion of behaviors (Bandara, McGinty, & Barry, 2020; Niederdeppe et al., 2015).

Both theory and evidence highlight transportation and counterarguing as two oft-cited mechanisms in studies that seek to understand the conditions under which narratives may or may not persuade. However, it is also clear that there is substantial heterogeneity in effect sizes of studies that compare narratives to nonnarrative messages, both in their overall persuasiveness and in the centrality of transportation and counterarguing as explanatory mechanisms. Furthermore, much of that heterogeneity remains unexplained by moderating variables that have been proposed and tested by researchers (O'Keefe & Hoeken, 2021). It may also be that narratives can persuade without substantial levels of transportation or counterarguing. After all, narratives are a very heterogeneous category of messaging to begin with, spanning different forms, formats, structures, and persuasive (or not) intent. Efforts to better understand the degree to which message factors, research practices, measures, research designs, samples, and study populations may influence narrative treatment effect heterogeneity and narrative processing are essential for better understanding of both the conditions under which narrative persuasion occurs and the reliability of expectations of a narrative advantage across various messaging contexts. We explore these questions in the context of short narrative messages designed to increase support for childcare policies, holding several of these factors constant across three separate data collections to isolate the potential roles of study populations and sampling strategies.

The Study Context

This article focuses on the issue of state-level childcare policies, as investments in high-quality childcare and early education for children ages zero to five years influence various health, educational, and social outcomes across the life course (Chandra et al., 2016). Comprehensive early childhood development programs (including childcare centers) benefit children (Phillips et al., 2017), their parents (Morrissey, 2017), and childcare providers (Otten et al., 2019). There nevertheless remain major challenges to access, affordability, and quality of early childcare programs in the United States. State investments in such

programs have strong potential to improve the health and well-being of families and communities. However, media portrayals of early childhood development tend to focus more on children beating the odds (i.e., individual efforts) than society changing the odds (via policy interventions that enhance access and affordability) (Berkeley Media Studies Group, 2018). Childcare policy advocates emphasize the importance of telling stories about families that place childcare within the broader social context and illustrate systemic issues that reduce accessibility, affordability, and quality of childcare programs.

For the analyses reported in the current article, we identified three targeted policies (in consultation with an advisory panel of childcare policy experts) and developed narrative and nonnarrative messages designed to increase support for these policies. We then tested message effects on policy support in three different samples across a period of several months, with a focus on how narrative mechanisms may operate in shifting support for these policies.

Hypotheses and Research Questions

This article offers a secondary analysis of data from two published studies (Study 1 and Study 2), focusing on variables not analyzed or reported in the original studies (transportation and counterarguing), and an original, preregistered follow-up Study 3 reported for the first time here. The original Studies 1 and 2 included preregistered hypotheses that predicted exposure to a narrative would increase public (Study 1) and state legislator (Study 2) support for increased investment in childcare policies, relative to a nonnarrative message. We chose these two audiences because policy changes can occur through various pathways that include (a) members of the public voting directly on policy propositions (which occurs in several U.S. states), (b) members of the public applying pressure on elected officials to enact policies, and/or (c) direct persuasion to enact policies among elected officials themselves. We tested the same message, at around the same time, using the same measures, across samples of the general public and state legislators.

The first two studies produced opposing findings. The narrative message produced *greater* policy support among members of the general public who were initially opposed to these policies (in a pretest) in the national sample of U.S. adults. The narrative produced *lower* intentions to engage in policy advocacy in the state legislator sample, a result seemingly driven by Republican lawmakers who were initially opposed to these policies. In short, the narrative achieved its intended persuasive effect in the general public sample while backfiring in that effort among a sample of state legislators. While there are examples of studies either (a) finding that a single narrative message was more or less transporting or persuasive among different subgroups (e.g., Green et al., 2008; see van Laer et al., 2014), or (b) observing variation in the size of effects between different narrative formats (e.g., van Laer, Feiereisen, & Visconti, 2019), we are unaware of studies finding diametrically opposed patterns of effects (persuasive for one, backfiring for another) on outcomes when holding constant the narrative's form and content.

In response to these seemingly contradictory patterns of findings, we conducted Study 3 as a new preregistered, follow-up direct replication study using a convenience sample of the general public (recruited from Prolific) in an effort to better understand the pattern of results observed in Studies 1 and

2 and to explore the degree of treatment effect heterogeneity across two distinct study populations and three distinct samples. The use of a convenience sample in Study 3 enabled direct comparisons between Study 3 and Study 1, which employed a more systematic sampling approach to better characterize the demographic distribution of the U.S. population. Here, we describe a subset of Study 3 hypotheses that either (a) test assumptions or (b) directly replicate analyses from Studies 1 and 2 related to transportation and counterarguing.

We began with a preregistered hypothesis 1 (PH1):

PH1:   *(manipulation check): Compared with a simple [nonnarrative] propolicy message, a narrative message will be more likely to be perceived as a personal story.*

We had not included such a manipulation check in the first two studies and wanted to confirm that respondents indeed perceived the narrative message as a personal story. Our second PH stemmed from preliminary analysis of Studies 1 and 2 in which, contrary to theoretical predictions, the narrative produced greater counterarguing than the nonnarrative. From these observations, we hypothesized (and preregistered) the following:

PH2:   *Compared with a simple propolicy message, a narrative message will increase counterarguing among respondents.*

Since our preliminary analysis of Studies 1 and 2 found inconsistent patterns of narrative versus nonnarrative effects on transportation, we did not preregister a hypothesis about it. Instead, we offer a nonpreregistered hypothesis (H) informed by theory:

H3:   *Compared with a simple propolicy message, a narrative message will produce greater transportation among respondents.*

Consistent with theory and research, we offered PH4:

PH4:   *Greater policy support will be associated with less counterarguing and more transportation.*

Finally, we offered a preregistered research question (PRQ) that combines predictions from PH2, H3 and PH4 into an indirect effects model:

PRQ1:   *Compared with a simple propolicy message, will the narrative message deliver indirect effects on policy support through increased counterarguing and/or increased transportation?*

It should be noted that while many narrative theories and models predict a stepwise sequence between the two mechanistic variables, such that greater transportation reduces counterarguing (e.g., EORM, E-ELM), However, our data (and most studies in the published literature) cannot tease out that sequence because we concurrently measured transportation and counterarguing in the same battery of

postmessage survey items. Therefore, we did not indicate any sequential relationships between these two variables in any of our hypotheses or research questions.

Methods

Study Procedure and Message Conditions

Study 1 (August and September 2019) sampled respondents from an established web-based research panel (M.S.G. Panel) and used quotas to ensure that respondent demographics were roughly comparable to U.S. Census estimates. The original Study 1 included seven message conditions and a two-week follow-up, though in the current secondary data analysis, we focus on the Time 1 comparison between narrative and nonnarrative messages (termed a "simple propolicy message" in the preregistration materials; n = 2,605 in these two conditions) that was constant across all three studies. Study 1's original preregistration, including all item wordings and response categories, can be found at the Open Science Framework (OSF) website (https://osf.io/jzyps).

Study 2 (August through December 2019) participants are state legislators recruited via direct e-mail contact using a database from the National Conference of State Legislators (NCSL), which includes contact information for all 7,500 current (at the time of the study) state legislators (https://osf.io/mg4zk). The original Study 2 included four total message conditions. In this secondary data analysis, we again focus on the narrative versus nonnarrative messages that are directly comparable across all three studies (n = 417 in these two conditions).

For the follow-up preregistered Study 3 (May 2020), we recruited participants from Prolific (https://www.prolific.co/), an online crowdsourcing platform. We restricted the sample to adults living in the United States (https://osf.io/5grnq). Study 3 had two conditions, the narrative and the nonnarrative (n = 525).

Each study featured random assignment to the same messages (comparing a narrative, 671 words, to a nonnarrative, 521 words) arguing that effects of quality childcare endure into adulthood and are essential for children from low-income families. We used identical language where possible, though the nonnarrative did not include information about plot, setting, or characters. The narrative told a story about Alisha and Jason, working parents living in Denver, who have struggled with finding affordable, high-quality childcare. The story argued it is an increasingly common challenge to access childcare while both parents work to support the family, described their efforts to offset them, placed Alisha and Jason's struggles in a broader context ("parents and families nationwide recognize the challenge of . . ."), and described how specific targeted policies could help to address the problem. The nonnarrative message argued for the necessity of investing in childcare and explained why state and local leaders should consider policies that support families and the professionals who provide high-quality childcare.

We removed respondents who dedicated fewer than 20 seconds on any study messages (a preregistered decision). This reduced the analytic sample to 1,747 respondents in Study 1 (67% of the original unfiltered sample), 297 respondents in Study 2 (71% of the original sample), and 500 respondents

in Study 3 (95% of the original sample). All three studies were deemed exempt by the Institutional Review Board at our institutions (see Table 1 for demographic composition and randomization checks; Table 2 for time participants spent to complete the study and read stimuli).

Table 1. Demographic Composition of Analytic Samples.

	Study 1 (n = 1,747)		Study 2 (n = 297)		Study 3 (n = 500)	
	Frequency (or mean)	% (or SD)	Frequency (or mean)	% (or SD)	Frequency (or mean)	% (or SD)
Age	$F(1, 1742) = 2.72, p = .100$		$F(1, 248) = 0.17, p = .685$		$F(1, 498) = 0.18, p = .668$	
	48.1	15.3	58.4	12.3	32.4	12.6
Education	$\chi^2(4) = 3.35, p = .502$		$\chi^2(3) = 5.9, p = .117$		$\chi^2(4) = 6.74, p = .15$	
Less than high school	66	3.9	0	0	8	1.6
High school diploma/GRE	375	22.4	5	2	66	13.2
Some college/ technical	669	40	19	7.4	166	33.2
Bachelor	402	24	87	34	189	37.8
Advanced degrees	161	9.6	145	56.6	71	14.2
Household income	$\chi^2(5) = 7.47, p = .188$		$\chi^2(3) = 6.1, p = .107$			
$0–$24,999	410	24.5	0	0		
$25,000–$49,999	471	28.2	20	8.4		
$50,000–$74,999	345	20.6	25	10.5	Not measured	
$75,000–$99,999	191	11.4	47	19.7		
100,000 or more	226	13.5	147	61.5		
Declined to respond	30	1.8	N/A	N/A		
Sex/gender	$\chi^2(2) = 0.79, p = .672$		$\chi^2(2) = 0.96, p = .618$		$\chi^2(2) = 0.7, p = .703$	
Female	922	52.8	116	47.3	280	56
Male	819	46.9	128	52.2	201	40.2
Transgender/ nonbinary	5	0.3	1	0.4	19	3.8
Political party	$\chi^2(3) = 0.66, p = .881$		$\chi^2(3) = 3.37, p = .338$		$\chi^2(3) = 2.67, p = .446$	
Democrat	610	34.9	148	57.8	246	49.2
Republican	546	31.3	99	38.7	67	13.4
Independent	441	25.2	6	2.3	133	26.6
Another party/no preference	150	8.6	3	1.2	54	10.8
Political party (with leaners)	$\chi^2(1) = 0.21, p = .646$		$\chi^2(1) = 1.36, p = .243$		$\chi^2(1) = 0.56, p = .453$	
Democrat	918	53.1	151	59	380	76
Republican	812	46.9	105	41	120	24

Political ideology	$\chi^2(2) = 0.05, p = .977$		Measured in terms of both social and fiscal ideology; see below		$\chi^2(2) = 0.73, p = .693$	
Conservative	566	33.9			84	16.8
Liberal	511	30.6			328	65.6
Moderate	594	35.5			88	17.6
Social ideology	Measured overall, not by social or fiscal ideology;		$\chi^2(2) = 4.6, p = .10$		Measured overall, not by social or fiscal ideology;	
Conservative			61	23.8		
Liberal	see above		116	45.3	see above	
Moderate			79	30.9		
Fiscal ideology	see above		$\chi^2(2) = 2.04, p = .361$		see above	
Conservative			89	34.8		
Liberal	see above		50	19.5	see above	
Moderate			117	45.7		
Race						
White	$\chi^2(1) = 0.05, p = .817$		$\chi^2(1) = 2.14, p = .143$		$\chi^2(1) = 0.65, p = .42$	
	1357	81.1	214	83.6	399	79.8
Black	$\chi^2(1) = 0.23, p = .635$		$\chi^2(1) = 0, p = 1$		$\chi^2(1) = 0.14, p = .711$	
	219	13.1	21	8.2	32	6.4
Hispanic/ Latinx	$\chi^2(1) = 0, p = .989$		$\chi^2(1) = 0, p = .988$		$\chi^2(1) = 0.54, p = .461$	
	158	9.5	13	5.2	43	8.6
Another race	$\chi^2(1) = 0.11, p = .735$		$\chi^2(1) = 0.25, p = .616$		$\chi^2(1) = 0.11, p = .736$	
	143	8.5	15	5.9	91	18.2
Parents (with children ≤ 5)	$\chi^2(1) = 0.41, p = .524$		$\chi^2(1) = 0.02, p = .884$		$\chi^2(1) = 0.14, p = .711$	
	252	15.1	59	19.9	48	9.6
Parents (any age)	$\chi^2(1) = 0.03, p = .857$		$\chi^2(1) = 0.2, p = .658$		$\chi^2(1) = 0.65, p = .419$	
	591	35.3	50	19.9	131	26.3

Note. We used χ^2 and analysis of variance (ANOVA) to test whether measured demographic characteristics were comparable between the narrative and nonnarrative conditions. *F* tests and χ^2 tests in the table show that randomization produced balanced groups on each of these measured variables.

Table 2. Time Spent on the Study and Stimuli Across the Three Direct Replications.

	Study 1		Study 2		Study 3	
	Mean (minutes)	Median (minutes)	Mean (minutes)	Median (minutes)	Mean (minutes)	Median (minutes)
Study completion time	15.3	11.7	5.2 (days)	18.6	14.3	10.0
Stimuli reading time	2.3	1.2	2.0	1.2	2.3	1.7

Measures

Perceived Narrativity

We asked participants the extent to which they perceived the message to be a personal story (from 1 = *not a personal story* to 7 = *a personal story*). This single-item measure was only used in Study 3 (M = 4.19, SD = 2.11).

Transportation

We measured transportation (Green & Brock, 2000) using six items that we adapted to be relevant to both narrative and nonnarrative messages (i.e., changing references from a "story" to a "message"). Each item was measured on a 7-point Likert-type scale (1 = *strongly disagree*; 7 = *strongly agree*). These items comprised a reliable composite measure in all three studies (see preregistered materials for full measures, and see Table 3 for means, standard deviations, and Cronbach's alphas for all studies).

Counterarguing

We asked respondents four questions about the degree to which they engaged in counterarguing after reading the messages (Nabi, Moyer-Gusé, & Byrne, 2007; Silvia, 2006). We used the same 7-point Likert scales. These items formed a reliable measure in all three studies (Table 3).

Targeted Policy Support

We asked participants how much they supported or opposed (ranging from 1 = *strongly oppose* to 7 = *strongly support*) seven policies for increasing affordable and high-quality childcare, three of which were explicitly targeted by the messages. These three items on targeted policy support comprised a reliable composite measure in all three studies (see Table 3). This outcome measure was also examined and reported elsewhere (Niederdeppe, Winett, Xu, Fowler, & Gollust, 2021; Winett, Niederdeppe, Xu, Gollust, & Fowler, 2021). The current article only uses this variable in indirect effects models that explore the potential roles of transportation and counterarguing in message processing.

Table 3. Scale Reliability.

	Study 1			Study 2			Study 3		
	M	*SD*	*a*	*M*	*SD*	*a*	*M*	*SD*	*a*
Perceived narrativity	N/A	N/A	N/A	N/A	N/A	N/A	4.19	2.11	N/A
Transportation	5.11	1.12	0.79	4.79	1.17	0.85	5.12	1.03	0.80
Counterarguing	2.98	1.30	0.80	3.21	1.67	0.93	2.71	1.43	0.82
Targeted policy support	5.52	1.47	0.93	4.92	1.93	0.95	5.73	1.36	0.93

Notes. M denotes mean; *SD* denotes standard deviation; *a* denotes Cronbach's alpha.

Analytic Approach

We conducted the same statistical analyses for each study using R Studio (Version 1.2.5019). We conducted independent-samples *t* tests to test effects of narrative and nonnarrative messages on transportation and counterarguing. We computed Pearson correlations to test whether and how the outcome variable was associated with transportation and counterarguing. We estimated indirect effect models with the *lavaan* package to assess indirect effects of the narrative message on policy support, through transportation and counterarguing, with the bootstrapping technique recommended by Preacher and Hayes (2004, 2008) with 10,000 iterations.

Results

Effects of the Narrative on Perceived Narrativity (Manipulation Check)

Participants in Study 3's narrative condition were far more likely to perceive the message as a personal story (*M* = 5.54, *SD* = 1.54) than participants in nonnarrative (simple propolicy) condition (*M* = 2.76, *SD* = 1.65; *t* = 19.49, *p* < .001, Cohen's *d* = 1.75). PH1 was thus supported.

Effects of the Narrative on Counterarguing (PH2) and Transportation (H3)

For counterarguing, Study 1 found that the narrative (*M* = 3.05, *SD* = 1.30) produced more counterarguing than the nonnarrative condition (*M* = 2.83, *SD* = 1.29; *t* = −3.12, *p* = .002, *d* = 0.16), which is contrary to theoretical predictions. Similarly, Study 2 found that the narrative (*M* = 3.45, *SD* = 1.69) produced more counterarguing than the nonnarrative condition (*M* = 2.93, *SD* = 1.58, *t* = −2.51, *p* = .013, *d* = 0.32), again contrary to theory. Thus, both Studies 1 and 2 supported PH2. Study 3 found no differences in counterarguing between narrative (*M* = 2.63, *SD* = 1.41) and nonnarrative messages (*M* = 2.73, *SD* = 1.44, *t* = −0.75, *p* = .455, *d* = 0.07), which was inconsistent with previous two studies, contrary to theory, and contrary to PH2.

For transportation, Study 1 found that the narrative produced significantly less transportation (*M* = 5.07, *SD* = 1.11) than the simple propolicy (nonnarrative) message (*M* = 5.21, *SD* = 1.12, *t* = 2.30, *p* = .02, *d* = 0.12), contrary to theory and H3. Study 2 found no differences in transportation between narrative

($M = 4.81$, $SD = 1.19$) and nonnarrative messages ($M = 4.82$, $SD = 1.06$, $t = 0.09$, $p = .929$, $d = 0.01$). Study 3 found that the narrative produced significantly more transportation ($M = 5.30$, $SD = 0.94$) than the nonnarrative message ($M = 4.95$, $SD = 1.09$, $t = 3.79$, $p < .001$, $d = 0.34$), consistent with theory and supporting H3.

Correlations Between Counterarguing, Transportation, and Policy Support (PH4)

Pearson correlation tests showed that greater policy support was associated with greater transportation, $r(1679) = 0.42$, $p < .001$, Study 1; $r(246) = 0.34$, $p < .001$, Study 2; $r(498) = 0.28$, $p < .001$, Study 3, and less counterarguing, $r(1680) = -0.66$, $p < .001$, Study 1; $r(249) = -0.88$, $p < .001$, Study 2; $r(498) = -0.72$, $p < .001$, Study 3, across all three studies. These findings offer unequivocal support for PH4 in three separate and direct replications.

Indirect Effects of Narrative (Versus Nonnarrative) Messages on Policy Support (PRQ1)

In Study 1, the indirect effect estimate between the narrative message and targeted policy support through counterarguing was -0.15, and the 95% confidence interval (CI) was -0.24 to -0.06. The estimate through transportation was -0.02, 95% CI [-0.04, -0.004]. The fact that both 95% CIs do not overlap zero indicates significant indirect associations between the narrative message and targeted policy support via transportation and counterarguing, though both in the direction of reduced policy support. We also observed residual direct effects of the narrative message on targeted policy support (estimate $= 0.24$, $p < .001$), which suggests the presence of other unexplained mechanistic variables (see Figure 1).
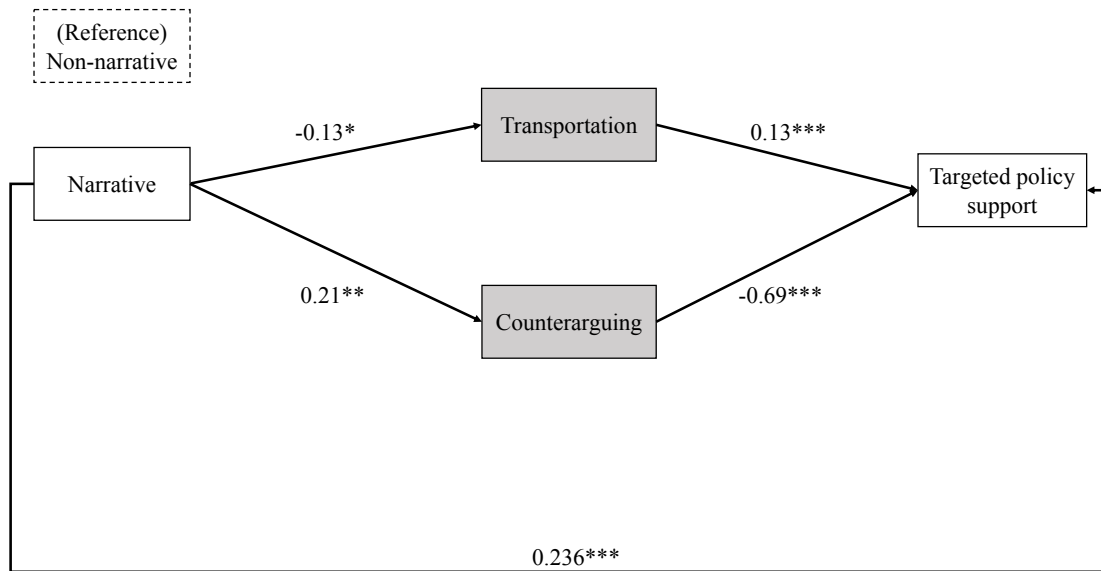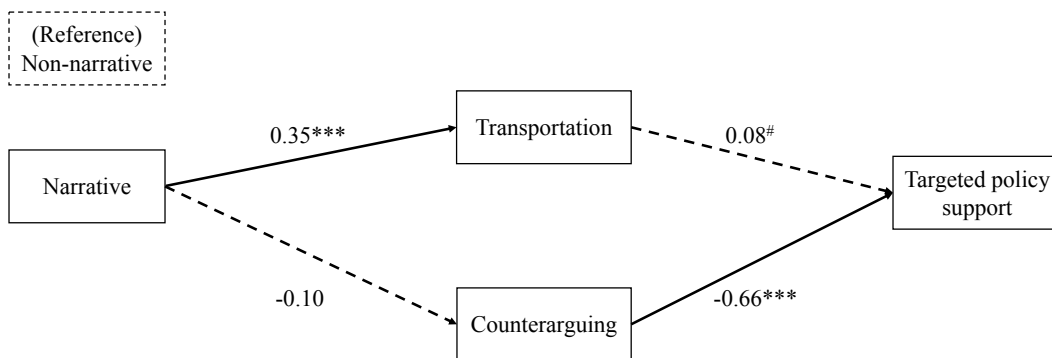

Figure 1. Indirect effects models (Study 1).

In Study 2, the indirect effect estimate through counterarguing was −0.55, 95% CI [−0.99, −0.15], which does not overlap zero. The indirect effect through transportation was −0.001, 95% CI [−0.03, 0.02]. The fact that the 95% CI overlaps zero means there was no significant indirect effect. The residual direct effect in this model was not significant (see Figure 2).
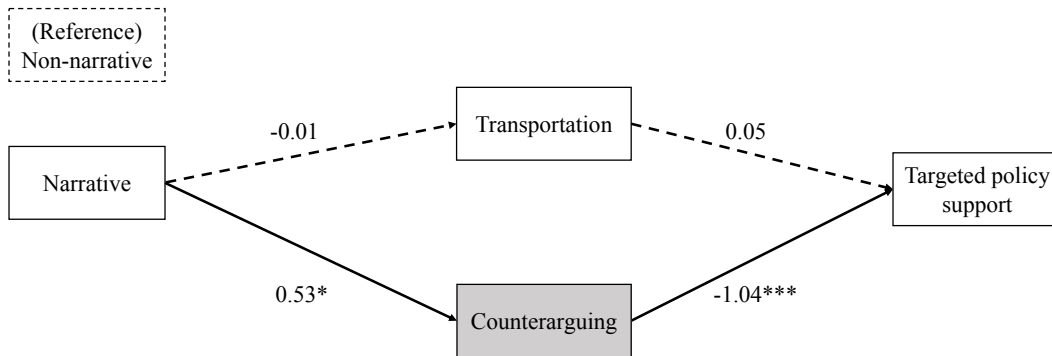
Figure 2. Indirect effects models (Study 2).

In Study 3, the indirect effect estimates were 0.06, 95% CI [−0.10, 0.23] (through counterarguing) and 0.03, 95% CI [−0.003, 0.07] (through transportation), which means there were no significant indirect associations via these mechanisms, although the direction of these coefficients were consistent with theoretical predictions for a narrative versus nonnarrative persuasive advantage via reduced counterarguing and increased transportation (see Figure 3).

Figure 3. Indirect effects models (Study 3).

Discussion

This internal replication project revealed treatment heterogeneity in terms of narrative message effects on counterarguing and transportation. In short, we found no evidence that either transportation or (reduced) counterarguing played any role in the strong observed persuasive effects of a strategic narrative in Study 1; in fact, both indicators suggested that that the narrative was persuasive *despite* counterproductive effects on these indicators. Study 2 offered evidence that increased counterarguing of the narrative likely played a role in its backfiring among legislators. Study 3, using a convenience sample that most closely resembles that typical of prior research in message effects and narrative persuasion (which typically relies on student or online convenience samples; see McCullock, Hildenbrand, Schmitz, & Perrault, 2021, for example), offered evidence of increased transportation most consistent with prior narrative persuasion theory. The lack of correspondence between a purposeful (Study 1) and convenience sample (Study 3) suggests the possibility that knowledge claims based on convenience samples may produce a false sense of certainty about the direction and predictability of effects in narrative persuasion research. This evidence, alongside recent analyses documenting substantial heterogeneity of communication message effects across numerous message considerations (O'Keefe & Hoeken, 2021), will be critical for efforts to assess the magnitude of replication concerns in communication research, and perhaps to temper claims about the invariability of "well-established" message effects.

In the pages that follow, we discuss—and in most cases, rule out—various potential causes of heterogeneous findings, leaving a few plausible explanations that warrant further inquiry and exploration.

Methodological Artifacts

One possible cause of treatment heterogeneity stems from issues about measurement (un)reliability, (poor stimulus) manipulations, and (variable) range restriction (Levine & Weber, 2020). We think such an explanation is unlikely for three reasons. First, each measure had high inter-item reliability, values of Cronbach's alpha of 0.79 or higher, across all studies. Second, the weak manipulation explanation seems implausible in light of tests of PH1 which showed that participants were far more likely to perceive the narrative message as a personal story compared with the nonnarrative condition (a mean 3 points higher on a 7-point scale; Cohen's d = 1.75, p < .001), as well as the fact that the narrative indeed produced significantly different persuasive outcomes than the nonnarrative in both Study 1 (greater persuasion among those with initial opposition to the policy) and Study 2 (a backfire effect among those inclined to oppose the policy). Third, there is limited evidence for scale range restriction, as there was sizable spread around mean values for each measure (all SDs > 1.00, many > 1.50 using 7-point Likert scales).

Construct Invalidity

Construct (in)validity is unlikely to explain our pattern of results (or lack thereof). We used the same scales across three studies and we employed measures that have either been previously validated (transportation; Green & Brock, 2000) or have been evaluated as having strong face validity (counterarguing; Nabi et al., 2007). In the narrative meta-analysis (Ratcliff & Sun, 2020), all the studies they included used Green and Brock's (2000) transportation scale, sometimes with adaptation (k = 33);

most studies adopted or modified counterarguing scales from Nabi and colleagues (2007) and Silvia (2006; *k* = 23).

Questionable Research Practices

Conducting questionable research practices is another concern, including issues regarding (insufficient) sample sizes, strategic case removal and deletion (Levine & Weber, 2020; Matthes et al., 2015; Vermeulen & Hartmann, 2015). To address these potential concerns, we employed open science practices and preregistered our data analysis plans for Study 3 on OSF for the sake of transparency. We preregistered (and followed) our sample sizes and case removal criteria in all three Studies on OSF.

Lab or Research Team Bias

The same research group used identical messages, measures, and sequences, which minimized procedural differences and reduced the possibility of lab/team bias.

Publication Bias

Publication bias (the avoidance of publishing statistically insignificant and theory-inconsistent findings; Sun & Pan, 2020) can lead to inflated effects and more false positives, which are associated with skewed scientific literature and reduced falsifiability of theories (Levine, 2013; Sun & Pan, 2020). This is not a plausible explanation for the observed pattern of heterogeneity here, however, because this article reports findings (some null) from all three of the preregistered studies conducted by the team.

Heterogeneity of Effects Across Different Samples/Populations; Potential Moderators and Other Unknown Moderators

Having argued that the previous explanations are unlikely explanations for our observed treatment effect heterogeneity, we arrive at the most plausible culprit: sample/population heterogeneity and the resulting possibility of potential or other unknown moderators.

The evidentiary base for many of the conclusions in narrative persuasion studies has relied on either college student samples or small convenience samples via online platforms like MTurk. The challenges of college student samples are well documented; they likely differ from other populations in personality traits (Corker, Donnellan, Kim, Schwartz, & Zamboanga, 2017), and even student samples from different disciplines can be quite different (Meltzer, Naab, & Daschmann, 2012). Similar critiques have been levied against studies that use crowdsourcing platforms like MTurk (see Chandler & Shapiro, 2016; Walters, Christakis, & Wright, 2018), though some scholars argue that MTurk samples can produce findings that are comparable to probability samples (Mullinix, Leeper, Druckman, & Freese, 2015).

Study 3 produced results that most closely resemble empirical and theoretical predictions from previous research. Notably, Study 3 used an online crowdsourcing platform (Prolific). Prolific itself has not

been studied extensively by researchers, but it is very similar to other, more well-studied platforms, like MTurk. Studies 1 and 2, however, produced results that were widely discrepant from theoretical predictions.

Several features of Study 1 likely improve the generalizability of the findings relative to the Prolific sample. Most notably, we were able to recruit a very large sample (Study 1 *n* = 1,747; Study 3 *n* = 500) and, through purposeful quota sampling, the demographic characteristics more closely reflect the U.S. national distribution in terms of age, gender, education, and political partisanship and ideology.

Study 2 was the most distinct from previous studies of narrative in that it involved a very specific set of stakeholders (state legislators) who have vested authority over the allocation of resources being advocated for in the message. Childcare was a "live" issue in most states when the study was being conducted—at least one such policy was actively being debated in each U.S. state. While the demographics of this sample are quite distinct from the distribution of the U.S. national population, they also likely represent a population that is more motivated by partisan ideologies and the financial realities of state investments in social programs than the general population. Study 2 participants also engaged with the survey in very different ways—a significant proportion of the state legislator sample completed the survey over a manner of days, not minutes (*M* = 4.3 days in Study 2, versus 17.3 minutes in Study 1 and 14.3 minutes in Study 3). While we did not observe any meaningful differences in the size or direction of narrative versus nonnarrative message effects between respondents who completed the survey in one sitting versus those who returned to it later, these dissimilar patterns of overall response are most likely to reflect larger differences in various potential moderators of narrative effects.

Previous narrative studies have identified various moderators of narrative transportation or persuasion: self-rated transportability (e.g., Mazzocco, Green, Sasota, & Jones, 2010), mental imagery ability (e.g., Zheng, 2014), need for cognition (NFC; e.g., Green et al., 2008; Zwarun & Hall, 2012), environmental distraction level (e.g., Zwarun & Hall, 2012), attention, education (e.g., van Laer et al., 2014; Williams, Green, Kohler, Allison, & Houston , 2011), sex/gender (e.g., van Laer et al., 2014; Williams et al., 2011). We did not measure most of these moderators since we did not have a priori expectations that these variables would lead to directionally opposed patterns of findings in our two samples. We did test for effect moderation by education and sex/gender in Study 1 and found no evidence that these variables moderated narrative processing or effects (Niederdeppe et al., 2021). Yet we think it implausible that the three samples differ so meaningfully in transportability, mental imagery ability, or need for cognition to be able to explain the levels of treatment effect heterogeneity observed here, particularly because we held the stimulus, measures, and analytic method constant across the three studies. This suggests the likelihood of other potential moderators of narrative processing and effects in the context of public and legislative debates around public policy.

The most notable divergence in Study 2 results was a strong (*B* = .50, *p* < .001) and positive effect of the narrative message on counterarguing—a finding that runs counter to meta-analytic evidence (e.g., Ratcliff & Sun, 2020). Perhaps related to this finding was that the narrative did not transport state legislators (indeed, the narrative was slightly less likely than a nonnarrative message to do so, albeit not by a statistically significant margin). More challenging to the evidence base, however, is that the very same message promoted transportation among a convenience sample (Study 3), and (as reported elsewhere)

that it produced higher levels of support for the targeted policy among respondents who were inclined to oppose it (Study 1; Niederdeppe et al., 2021). These findings, combined with strong evidence that respondents indeed perceived the message as a personal story (PH1), belie the idea that this interpretation can be dismissed out of hand due to the use of an ineffective or invalid narrative.

Study Limitations

We nevertheless cannot rule out Time × Treatment interactions as a possible explanation. While we collected data for Studies 1 and 2 on a similar timeline in late 2019, Study 3 took place in May 2020. Notably, March 2020 witnessed the onset of a global pandemic (COVID-19) with far-reaching health, economic, and social implications, including schools and childcare centers closure. We note, however, that Study 3 results most clearly reflect longstanding theoretical predictions and meta-analytic syntheses of previous findings. We also note that patterns of results of Studies 1 and 2 were different in magnitude from one another. We thus think that timing alone could not explain variability in observed effects, but we cannot rule out this possibility. In addition, while using a single pair of messages in one context, the same measures, procedures, and team allow us to explore sample differences as a potential source of treatment effect heterogeneity, they also limit the broader generalizability. There remains considerable value to multiteam replication studies in communication, and conceptual replications (which do not hold such details constant) offer different benefits in addressing broader questions of generalizability and reproducibility of research results.

Study Implications

While we can only speculate, plausible explanations for sample and population divergence in treatment effects include concepts like issue involvement, motivated reasoning, policy knowledge, perceptions of persuasive intent, respondents' history of transactional politics on the issues studied, or any other number of unmeasured variables that may have shifted message effects. Yet it may be difficult, a priori, to anticipate such a wide variety of potential explanations, and previous attempts to systematically identify reliable moderators of message effects (including narrative persuasion) have remained elusive (O'Keefe & Hoeken, 2021). Furthermore, one must establish the presence of treatment effect heterogeneity before one can begin to diagnose its origins. All told, this replication underscores concerns about treatment heterogeneity in communication research, highlights the need to consider diverse samples and populations in testing theoretical predictions, and suggests that common causes for treatment of heterogeneity proposed in previous work may not fully explain the magnitude of heterogeneity revealed in earlier meta-analyses of message effects.

Our findings suggest a need for significant caution and nuance in making knowledge claims about communication phenomena or, at the least, message design implications (see O'Keefe, 2015; O'Keefe & Hoeken, 2021). These findings give pause to those seeking to draw nonconditional conclusions about the effects of message design features. Meta-analytic results are certainly valuable because these studies often carefully weight their effect sizes for differences in sample sizes, methodologies, content stimuli, and other factors. However, conclusions might take care to emphasize that nonconditional meta-analytic differences are not definitive prescriptions for success. They might qualify that even evidence-based message effects

do not always occur and, under some conditions that we do not yet fully understand, can even backfire. Such cautions are needed, as communications research findings are imported into real-world contexts of great societal consequence, such as policy advocacy, public communication campaigns, and policy testimony. For example, it is commonly recommended in policy advocacy and campaign circles that communicators "tell the story" of their issues (Ganz, 2011). Our findings suggest that narrative communications in such contexts could cause social harm, either through ineffectiveness or even retrenchment into prior beliefs, if they do not appreciate or integrate a potential for heterogenous outcomes.

References

Bandara, S. N., McGinty, E. E., & Barry, C. L. (2020). Message framing to reduce stigma and increase support for policies to improve the wellbeing of people with prior drug convictions. *International Journal of Drug Policy, 76*, 102643. doi:10.1016/j.drugpo.2019.102643

Berkeley Media Studies Group. (2018). *From beating the odds to changing the odds: Recommendations for journalists covering early childhood*. Retrieved from http://www.bmsg.org/resources/publications/recommendations-for-journalists-covering-early-childhood/

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. doi:10.1080/03637751.2015.1128555

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12*, 53–81. doi:10.1146/annurev-clinpsy-021815-093623

Chandra, A., Acosta, J. D., Carman, K. G., Dubowitz, T., Leviton, L., Martin, L. T., . . . Plough, A. L. (2016). *Building a national culture of health: Background, action framework, measures, and next steps*. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR1199.html

Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences of the United States of America, 115*(49), 12441–12446. doi:10.1073/pnas.1808083115

Corker, K. S., Donnellan, M. B., Kim, S. Y., Schwartz, S. J., & Zamboanga, B. L. (2017). College student samples are not always equivalent: The magnitude of personality differences across colleges and universities. *Journal of Personality, 85*(2), 123–135. doi:10.1111/jopy.12224

Dal Cin, S., Zanna, M. P., & Fong, G. T. (2004). Narrative persuasion and overcoming resistance. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 175–191). Mahwah, NJ: Erlbaum.

de Graaf, A., Sanders, J., & Hoeken, H. (2016). Characteristics of narrative interventions and health effects: A review of the content, form, and context of narratives in health-related narrative persuasion research. *Review of Communication Research, 4*, 88–131.

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., . . . de Vreese, C. (2020). An agenda for open science in communication. *Journal of Communication, 71*(1), 1–26. doi:10.1093/joc/jqz052

Ganz, M. L. (2011). Public narrative, collective action, and power. In S. Odugbemi & T. Lee (Eds.), *Accountability through public opinion: From inertia to public action* (pp. 273–289). Washington, DC: The World Bank. doi:10.1596/9780821385050_ch18

Green, M. C. (2006). Narratives and cancer communication. *Journal of Communication, 56*(s1), S163–S183. doi:10.1111/j.1460-2466.2006.00288.x

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79*(5), 701–721. doi:10.1037/0022-3514.79.5.701

Green, M. C., Kass, S., Carrey, J., Herzig, B., Feeney, R., & Sabini, J. (2008). Transportation across media: Repeated exposure to print and film. *Media Psychology, 11*(4), 512–539. doi:10.1080/15213260802492000

Keating, D. M., & Totzkay, D. (2019). We do publish (conceptual) replications (sometimes): Publication trends in communication science, 2007–2016. *Annals of the International Communication Association, 43*(3), 225–239. doi:10.1080/23808985.2019.1632218

Knowles, E. S., & Linn, J. A. (2004). *Resistance and persuasion*. Mahwah, NJ: Erlbaum.

Kreuter, M. W., Green, M. C., Cappella, J. N., Slater, M. D., Wise, M. E., Storey, D., . . . Woolley, S. (2007). Narrative communication in cancer prevention and control: A framework to guide research and application. *Annals of Behavioral Medicine, 33*(3), 221–235. doi:10.1080/08836610701357922

Levine, T. R. (2013). A defense of publishing nonsignificant (ns) results. *Communication Research Reports, 30*(3), 270–274. doi:10.1080/08824096.2013.806261

Levine, T. R., & Weber, R. (2020). Unresolved heterogeneity in meta-analysis: Combined construct invalidity, confounding, and other challenges to understanding mean effect sizes. *Human Communication Research, 46*(2/3), 343–354. doi:10.1093/hcr/hqz019

Lewis, N. A. (2020). Open communication science: A primer on why and some recommendations for how. *Communication Methods and Measures, 14*(2), 71–82. doi:10.1080/19312458.2019.1685660

Lindsay, R. M., & Ehrenberg, A. S. C. (1993). The design of replicated studies. *The American Statistician, 47*(3), 217–228. doi:10.2307/2684982

Matthes, J., Marquart, F., Naderer, B., Arendt, F., Schmuck, D., & Adam, K. (2015). Questionable research practices in experimental communication research: A systematic analysis from 1980 to 2013. *Communication Methods and Measures, 9*(4), 193–207. doi:10.1080/19312458.2015.1096334

Mazzocco, P. J., Green, M. C., Sasota, J. A., & Jones, N. W. (2010). This story is not for everyone: Transportability and narrative persuasion. *Social Psychological and Personality Science, 1*(4), 361–368. doi:10.1177/1948550610376600

McCullock, S. P., M. Hildenbrand, G., Schmitz, K. J., & Perrault, E. K. (2021). The state of health communication research: A content analysis of articles published in *Journal of Health Communication* and *Health Communication* (2010–2019). *Journal of Health Communication, 26*(1), 28–38. doi:10.1080/10810730.2021.1879320

McEwan, B., Carpenter, C. J., & Westerman, D. (2018). On replication in communication science. *Communication Studies, 69*(3), 235–241. doi:10.1080/10510974.2018.1464938

Meltzer, C. E., Naab, T., & Daschmann, G. (2012). All student samples differ: On participant selection in communication science. *Communication Methods and Measures, 6*(4), 251–262. doi:10.1080/19312458.2012.732625

Morrissey, T. (2017). Child care and parent labor force participation: A review of the research literature. *Review of Economics of the Household, 15*(1), 1–24. doi:10.1007/s11150-016-9331-3

Moyer-Gusé, E. (2008). Toward a theory of entertainment persuasion: explaining the persuasive effects of entertainment-education messages. *Communication Theory, 18*(3), 407–425. doi:10.1111/j.1468-2885.2008.00328.x

Moyer-Gusé, E., & Nabi, R. L. (2009). Explaining the effects of narrative in an entertainment television program: Overcoming resistance to persuasion. *Human Communication Research, 36*(1), 26–52. doi:10.1111/j.1468-2958.2009.01367.x

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science, 2*(2), 109–138. doi:10.1017/XPS.2015.19

Nabi, R. L., Moyer-Gusé, E., & Byrne, S. (2007). All joking aside: A serious investigation into the persuasive effect of funny social issue messages. *Communication Monographs, 74*(1), 29–54. doi:10.1080/03637750701196896

Niederdeppe, J., Heley, K., & Barry, C. L. (2015). Inoculation and narrative strategies in competitive framing of three health policy issues. *Journal of Communication, 65*(5), 838–862. doi:10.1111/jcom.12162

Niederdeppe, J., Winett, L. B., Xu, Y., Fowler, E. F., & Gollust, S. E. (2021). Evidence-based message strategies to increase public support for state investment in early childhood education: Results from a longitudinal panel experiment. *The Milbank Quarterly*. Advance online publication. doi:10.1111/1468-0009.12534

O'Keefe, D. J. (2013). The relative persuasiveness of different message types does not vary as a function of the persuasive outcome assessed: Evidence from 29 meta-analyses of 2,062 effect sizes for 13 message variations. *Annals of the International Communication Association, 37*(1), 221–249. doi:10.1080/23808985.2013.11679151

O'Keefe, D. J. (2015). Message generalizations that support evidence-based persuasive message design: Specifying the evidentiary requirements. *Health Communication, 30*(2), 106–113. doi:10.1080/10410236.2014.974123

O'Keefe, D. J., & Hoeken, H. (2021). Message design choices don't make much difference to persuasiveness and can't be counted on—Not even when moderating conditions are specified. *Frontiers in Psychology, 12*, 2533. doi:10.3389/fpsyg.2021.664160

Otten, J. J., Bradford, V. A., Stover, B., Hill, H. D., Osborne, C., Getts, K., & Seixas, N. (2019). The culture of health in early care and education: Workers' wages, health, and job characteristics. *Health Affairs, 38*(5), 709–720. doi:10.1377/hlthaff.2018.05493

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., . . . Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects*. Washington, DC: Brookings Institute. Retrieved from https://www.brookings.edu/research/puzzling-it-out-the-current-state-of-scientific-knowledge-on-pre-kindergarten-effects/

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*(4), 717–731. doi:10.3758/BF03206553

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891. doi:10.3758/BRM.40.3.879

Rains, S. A., Levine, T. R., & Weber, R. (2018). Sixty years of quantitative communication research summarized: Lessons from 149 meta-analyses. *Annals of the International Communication Association, 42*(2), 105–124. doi:10.1080/23808985.2018.1446350

Rains, S. A., Matthes, J., & Palomares, N. A. (2020). Communication science and meta-analysis: Introduction to the special issue. *Human Communication Research, 46*(2–3), 115–119. doi:10.1093/hcr/hqaa007

Ratcliff, C. L., & Sun, Y. (2020). Overcoming resistance through narratives: Findings from a meta-analytic review. *Human Communication Research, 46*(4), 412–443. doi:10.1093/hcr/hqz017

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100. doi:10.1037/a0015108

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. doi:10.1080/00913367.2015.1018467

Silvia, P. J. (2006). Reactance and the dynamics of disagreement: Multiple paths from threatened freedom to resistance to persuasion. *European Journal of Social Psychology, 36*(5), 673–685. doi:10.1002/ejsp.309

Slater, M. D., & Rouner, D. (2002). Entertainment-education and elaboration likelihood: Understanding the processing of narrative persuasion. *Communication Theory, 12*(2), 173–191. doi:10.1111/j.1468-2885.2002.tb00265.x

Sun, Y., & Pan, Z. (2020). Not published is not perished: Addressing publication bias in meta-analytic studies in communication. *Human Communication Research, 46*(2–3), 300–321. doi:10.1093/hcr/hqz015

Tukachinsky, R., & Stokunaga, R. (2013). The effects of engagement with entertainment. *Annals of the International Communication Association, 37*(1), 287–322. doi:10.1080/23808985.2013.11679153

van Laer, T., de Ruyter, K., Visconti, L. M., & Wetzels, M. (2014). The extended transportation-imagery model: A meta-analysis of the antecedents and consequences of consumers' narrative transportation. *Journal of Consumer Research, 40*(5), 797–817. doi:10.1086/673383

van Laer, T., Feiereisen, S., & Visconti, L. M. (2019). Storytelling in the digital era: A meta-analysis of relevant moderators of the narrative transportation effect. *Journal of Business Research, 96*, 135–146. doi:10.1016/j.jbusres.2018.10.053

Vermeulen, I., & Hartmann, T. (2015). Questionable research and publication practices in communication science. *Communication Methods and Measures, 9*(4), 189–192. doi:10.1080/19312458.2015.1096331

Walters, K., Christakis, D. A., & Wright, D. R. (2018). Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? *PLOS ONE, 13*(6), 1–10. doi:10.1371/journal.pone.0198835

Williams, J. H., Green, M. C., Kohler, C., Allison, J. J., & Houston, T. K. (2011). Stories to communicate risks about tobacco: Development of a brief scale to measure transportation into a video story— The ACCE Project. *Health Education Journal, 70*(2), 184–191. doi:10.1177/0017896910373171

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. *The Journal of Public Interest Communications, 5*(1), 45–45. doi:10.32473/jpic.v5.i1.p45

Zheng, L. (2014). Narrative transportation in radio advertising: A study of the effects of dispositional traits on mental transportation. *Journal of Radio & Audio Media, 21*(1), 36–50. doi:10.1080/19376529.2014.891213

Zwarun, L., & Hall, A. (2012). Narrative persuasion, transportation, and the role of need for cognition in online viewing of fantastical films. *Media Psychology, 15*(3), 327–355. doi:10.1080/15213269.2012.700592