Title
The Role of Implicit Information in Choice Architecture

Author
Leong, Lim

Publication Date
2020

UNIVERSITY OF CALIFORNIA SAN DIEGO

The Role of Implicit Information in Choice Architecture

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Experimental Psychology

by

Lim Man Leong

Committee in charge:

Professor Craig R. M. McKenzie, Chair
Professor Lera Boroditsky
Professor Adena Schachner
Professor Edward Vul
Professor Piotr Winkielman

2020

The Dissertation of Lim Man Leong is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____

Chair

University of California San Diego

2020

DEDICATION

To my friends, mentors, and family

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and support of numerous professors, colleagues, and friends – and without the cancellation and postponement of sports, theater, and social life in general. I am deeply grateful to everyone who has helped made these past years more enjoyable than binge-watching eight seasons of Game of Thrones.

I would like to thank my advisor Craig McKenzie. Your patience, wisdom, and compassion provided me the salubrious environment to thrive, and your mentorship has left an indelible mark not only in my research, but also my life. Thank you for pointing me in the right directions, inspiring me to step into the spotlight, and encouraging me to take risks. I would like to also thank the rest of my dissertation committee – Adena Schachner, Piotr Winkielman, Edward Vul, and Lera Boroditsky – for their constructive comments and thoughtful feedback.

I would like to acknowledge the contributions of my current and former lab mates and research assistants in the McKenzie Lab and the Mind & Development Lab. My research program would not have progressed without your support. In particular, I would like to thank Johannes Müller-Trede and Shlomi Sher for their delightful presence and insightful discussions.

To my esteemed colleagues and beloved friends – thank you for tolerating my wry sense of humor (among other quirks) and for indulging me to plan my work around play. I will always cherish the fond memories of us learning together, playing together, and most importantly, eating together.

Finally, I would like to thank my family for their unwavering support and wholehearted acceptance of my pursuits in life.

Chapter 1, in full, is a reprint of the material as it appears in the Journal of Behavioral Decision Making, 2017, Leong, Lim M.; McKenzie, Craig R. M.; Sher, Shlomi; Müller-Trede, Johannes. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is under review for publication of the material. Leong, Lim M.; McKenzie, Craig R. M. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Psychonomic Bulletin & Review, 2020, Leong, Lim M.; Yin, Yidan; McKenzie, Craig R. M. The dissertation author was the primary investigator and author of this paper.

VITA

2013 Bachelor of Arts, International Studies, University of California San Diego

2013 Bachelor of Science, Psychology, University of California San Diego

2016 Master of Arts, Psychology, University of California San Diego

2020 Doctor of Philosophy, Experimental Psychology, University of California San Diego

PUBLICATIONS

McKenzie, C. R. M., Leong, L. M., & Sher, S. (in press). Default sensitivity in attempts at social influence. *Psychonomic Bulletin & Review.*

Leong, L. M., Yin, Y., & McKenzie, C. R. M. (2020). Exploiting asymmetric signals from choices through default selection. *Psychonomic Bulletin & Review, 27*, 162-169.

Leong, L. M., McKenzie, C. R. M., Sher, S., & Müller-Trede, J. (2019). Illusory inconsistencies in judgment: Evoked reference sets and between-subject designs. *Psychonomic Bulletin & Review, 26*(2), 647-653.

McKenzie, C. R. M., Sher, S., Leong, L. M., & Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics, 5*(3-4), 337-360.

Leong, L. M., McKenzie, C. R. M., Sher, S., & Müller-Trede, J. (2017). The role of inference in attribute framing effects. *Journal of Behavioral Decision Making, 30*(5), 1147-1156.

ABSTRACT OF THE DISSERTATION

The Role of Implicit Information in Choice Architecture

by

Lim Man Leong

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2020

Professor Craig R. M. McKenzie, Chair

Choice architecture, or the design of the context in which people make judgments and decisions, can influence people's behaviors in systematic ways. In this dissertation, I focus on two aspects of choice architecture – framing (which description to use) and defaults (which option to preselect). Because the options under consideration remain the same regardless of which frame is used or which option is set as the default, inconsistent responses are commonly viewed as irrational. This perspective of choice architecture, however, overlooks the subtle information that choice architects may convey to decision makers through their choice of frame and default. Building on this information leakage framework, I aim to provide a more

comprehensive understanding of how this implicit social interaction affects judgment and decision making. Chapter 1 provides a causal link between frames, inferences, and judgments, and demonstrates that inferences are sufficient to generate attribute framing effects. Chapter 2 examines whether attribute framing is malleable to the informativeness of the choice of frame. While framing effects remained unaltered when the uninformative frame selection process was described using hypothetical vignettes, participants ceased to respond systematically to frames when the incentive structure was manipulated in a two-player repeated game. Finally, Chapter 3 shows that when decision makers can self-select into different default settings, they strategically exploited the asymmetric signal that their choices convey. Overall, the current dissertation provides further support for choice architecture as an implicit social interaction, and offers new insights on factors that moderate the effectiveness of behavioral interventions.

GENERAL INTRODUCTION

Decades of psychological research have found that people's judgments and decisions are influenced in predictable ways by seemingly irrelevant and inconsequential changes to the choice context (e.g., Gilovich, Griffin, & Kahneman, 2002; Kahneman, 2011). In recent years, the "nudge" approach has sought to steer these cognitive "biases" toward better outcomes through choice architecture, or the design of the choice environment (Thaler & Sunstein, 2008). To qualify as a nudge, the behavioral intervention must preserve the decision maker's autonomy to choose and impose no significant changes to the economic incentives. Because nudges are relatively low-cost and easy to implement, they have become popular behavioral policy tools in many domains, including retirement savings, energy, and health (e.g., Benartzi et al., 2017).

The current dissertation focuses on two nudges: attribute framing and default options. Unlike risky choice framing (e.g., "Disease Problem" from Tversky & Kahneman, 1981), attribute framing effects occur when people respond differently to logically equivalent descriptions of an object or event (Levin, Schneider, & Gaeth, 1998). For example, ground beef can be described as either "75% lean" or "25% fat," but people respond more favorably to the positive frame than the negative frame (Levin & Gaeth, 1988). In a conceptually similar vein, default effects occur when people are more likely to choose an option when it is set as the default (Jachimowicz, Duncan, Weber, & Johnson, 2019). For example, people are more likely to be organ donors in countries where "organ donor" is the default than countries where "not an organ donor" is the default (Johnson & Goldstein, 2003; Steffel, Williams, & Tannenbaum, 2019). Because attribute framing and default options do not alter the menu of options or prevent people from choosing what they want, people's inconsistent responses are often interpreted as evidence for irrationality (e.g., Shafir & LeBoeuf, 2002; Thaler & Sunstein, 2008).

The conventional approach to choice architecture, however, neglects the implicit social interaction between choice architects, decision makers, and third-party observers (Krijnen, Tannenbaum, & Fox, 2017; Hilton, 1995; Schwarz, 1994). Crucially, a choice architect's decision between different designs – which description to use or which option to preselect – can convey information such as reference points and personal attitudes (McKenzie, Sher, Leong, & Müller-Trede, 2018). That is, the logical equivalence between different choice architectures does not guarantee information equivalence. If different choice contexts convey different task-relevant information, then responding differently on the basis of different information is not problematic.

In line with this perspective, the "information leakage" account argues that attribute framing effects (McKenzie & Nelson, 2003; Sher & McKenzie, 2006) and default effects (McKenzie, Liersch, & Finkelstein, 2006) occur because different information is tacitly transmitted from choice architects (e.g., speakers, policymakers) to decision makers (e.g., listeners, decision makers). According to this framework, speakers do not choose frames at random, but are instead guided by background conditions such as reference points. For example, speakers are more likely to describe a cup at the half-way mark as "half-empty" (rather than "half-full") when the reference point is high (e.g., a full cup). In turn, listeners are attuned to this regularity and draw different inferences from frames. For example, listeners are more likely to infer that a cup was previously full (e.g., a high reference point) when it is described as "half-empty" (rather than "half-full"). The reference points that influence speaker frame selection therefore match the inferences that listeners draw from frames.

Along the same lines, the default-setter's personal attitudes can be revealed through their choice of default. For instance, policymakers who personally believe that others ought to be organ donors are more likely to set "organ donor" as the default than "not an organ donor." In turn, decision makers are more likely to infer this implicit recommendation when presented with the "organ donor" default. Thus, people may stick with default options in part because they perceive them as recommended courses of action. Rather than exemplifying irrationality, framing and default effects may instead demonstrate people's remarkable sensitivity to subtle cues in the choice environment. A similar informational analysis can also explain other ostensibly puzzling behaviors such as joint-separate reversals (Sher & McKenzie, 2014) and rating dominated gambles as more attractive (McKenzie & Sher, 2020).

The main goal of this dissertation is to provide a more comprehensive understanding of the role that pragmatic and social inferences play in attribute framing and default effects. Across 11 experiments, I examine whether inferences are sufficient to generate attribute framing effects (Chapter 1), whether attribute framing effects are adaptable to the informativeness of the choice of frame (Chapter 2), and whether decision makers can exploit the asymmetric signal from choices under different default settings (Chapter 3). Overall, the results provide additional support for choice architecture as an implicit social interaction, and suggest that the inferences people draw drive their judgments and choices.

Chapter 1 fills the gap in the existing literature by examining whether inferences play a causal role in the generation of attribute framing effects. While past research has demonstrated that frames influence both reference point inferences and judgments, it has not established a causal link between these inferences and judgments. To address this, we used a yoked design in Experiment 1 to decompose the typical attribute framing paradigm: "Modeler" participants saw one of two frames and provided their inferences regarding the reference point, and "recipient" participants then received these inferences and evaluated the same fully described (i.e., unframed) target. We found that even though recipients were not exposed to different frames, the different inferences they received were sufficient to reproduce an attribute framing effect. In Experiment 2, we tested whether experts, who are less likely to draw different inferences from frames, would show a diminished framing effect. Indeed, we found that participants who were more knowledgeable about basketball exhibited a reduced framing effect in the basketball domain, but not the medical domain. Together, these results illustrate the role of inferences in the generation and attenuation of attribute framing effects.

Chapter 2 extends the information leakage framework by testing whether attribute framing effects are adaptable to changes to the task environment. While attribute framing effects reflect behavior well-suited to the typical communicative context, it is less clear whether people can cease to draw inferences, and exhibit reduced framing effects, when the speaker's choice of frame is uninformative. We tested this by employing two different paradigms: the first three studies used hypothetical vignettes while the final study used a two-player interactive game. In Study 1, we found that informing participants that the frame was selected randomly by a computer did not reduce the framing effect. In Studies 2a, and 2b, we introduced a constraint that manipulated how much choice the speaker had in using a particular frame, and again found that the framing effect remains unaltered. In Study 3, we used an interactive repeated game, and manipulated the incentive structure between speaker and listener participants. Contrary to our previous results, we now found that when their interests were misaligned, speakers no longer selected frames based on reference points, and listeners no longer responded systematically to frames. Taken together, our results caution against a simple answer as to whether attribute framing effects are adaptable and instead suggests that it may depend on how the change to the task environment is implemented.

Chapter 3 investigates whether decision makers could strategically self-select into different default settings when given the opportunity. Prior research has shown that choosing the same option can convey different social meanings depending on whether that option was obtained by staying with, or switching from, the default (Davidai, Gilovich, & Ross, 2012). Given these asymmetric signals, we placed participants in the role of decision makers to examine whether they can exploit this perception. In five experiments using hypothetical scenarios and incentivized economic games, we found that in the presence of observers, decision makers were more likely to choose defaults that require switching when sending a positive signal and defaults that require staying when sending a negative signal. These results suggest that people understand how their choices under different defaults are construed by others, and have potential implications for behavioral interventions in the real-world.

References

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., . . . Galing, S. (2017). Should governments invest more in nudging? *Psychological Science, 28*, 1041–1055.

Davidai, S., Gilovich, T., & Ross, L. D. (2012). The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences, 109*, 15201–15205.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York, NY: Cambridge University Press.

Hilton, D. J. (1995). The social context of reasoning: conversational inference and rational judgment. *Psychological Bulletin, 118*, 248–271.

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 1-28.

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science, 302*, 1338–1339.

Kahneman, D. (2011). *Thinking, fast and slow.* New York, NY: Farrar, Straus and Giroux.

Krijnen, J. M. T., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy, 3*(2), 1–18.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*, 374-378.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*, 149–188.

McKenzie, C. R. M., Liersch, M. K., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science, 17*, 414–420.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review, 10*, 596–602.

McKenzie, C. R. M., & Sher, S. (2020). Gamble evaluation and evoked reference sets: Why adding a small loss to a gamble increases its attractiveness. *Cognition, 194*, 104043.

McKenzie, C. R. M., Sher, S., Leong L. M., & Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics, 5*, 337–360.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. *Advances in Experimental Social Psychology, 26*, 123–162.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53*(1), 491–517.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101*, 467–494.

Sher, S., & McKenzie, C. R. M. (2014). Options as information: Rational reversals of evaluation and preference. *Journal of Experimental Psychology: General, 143*(3), 1127–1143.

Steffel, M., Williams, E. F., & Tannenbaum, D. (2019). Does changing defaults save lives? Effects of presumed consent organ donation policies. *Behavioral Science & Policy, 5*(1), 69–88.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge. Improving decisions about health, wealth and happiness.* London, UK: Penguin Books.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.

Chapter 1:

The role of inference in attribute framing effects

Lim M. Leong, Craig R. M. McKenzie, Shlomi Sher, & Johannes Müller-Trede

As it appears in

Journal of Behavioral Decision Making

2017

Volume 30, Issue 5, Pages 1147-1156

Journal of Behavioral Decision Making, J. Behav. Dec. Making (2017)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.2030

The Role of Inference in Attribute Framing Effects

LIM M. LEONG,[1]* CRAIG R. M. MCKENZIE,[1] SHLOMI SHER[2] and JOHANNES MÜLLER-TREDE[1]
[1]*University of California, San Diego, La Jolla, CA USA*
[2]*Pomona College, Claremont, CA USA*

ABSTRACT

Previous research has shown that a speaker's choice between logically equivalent frames is influenced by reference point information, and that listeners draw accurate inferences based on the frame. Less clear, however, is whether these inferences play a causal role in generating attribute framing effects. Two experiments are reported, which suggest that frame-dependent inferences are sufficient to generate attribute framing effects, and that blocking such inferences may block framing effects. Experiment 1 decomposed the typical framing design into two parts: One group of participants saw a target described in one of two attribute frames and reported their estimates (inferences) of the typical attribute value. These estimates were then given to a second group of yoked participants, who evaluated the target. Although this latter group was not exposed to different attribute frames, they nevertheless exhibited a "framing effect" as a result of receiving systematically different inferences. In contrast, Experiment 2 shows that experts—who are familiar with an attribute's distribution and are therefore less likely to draw strong frame-based inferences—exhibit a diminished framing effect. Together, these findings underscore the role of inferences in the generation and attenuation of attribute framing effects. Copyright © 2017 John Wiley & Sons, Ltd.

Additional Supporting Information may be found online in the supporting information tab for this article.

KEY WORDS framing effect; inference; information leakage; rationality; expertise

Framing effects occur when people's judgments or choices systematically depend on which logically equivalent description of outcomes or objects is presented to them. The present article focuses on attribute framing, in which a single attribute of an object is described in one of two ways. One frame is usually positive and one negative, and a robust finding is that the object is evaluated more favorably in the positive frame than the negative frame (a "valence-consistent shift" in preference; Levin, Schneider, & Gaeth, 1998). Ground beef is rated as better tasting and less greasy when described as "75% lean" rather than "25% fat" (Levin & Gaeth, 1988), a basketball player's performance is rated higher when described in terms of the percentage of shots "made" rather than "missed," and a medical treatment is more likely to be recommended when described in terms of "survival" rather than "mortality" rate (see Table 3 in Levin et al., 1998).

Several competing explanations for these intriguing effects have been proposed. Levin and colleagues suggested an associative account (Levin, 1987; Levin et al., 1998). Positive frames are assumed to evoke positive associations, negative frames evoke negative associations, and these associations influence the evaluation of the object. Thus, ground beef described as "75% lean" is evaluated more favorably because "lean" evokes positive associations, which in turn color the perception of the ground beef. A second account, query theory, posits that frames influence the order in which people retrieve supporting evidence (Hardisty, Johnson, & Weber, 2010). According to this account, the initial query generates more retrievals, and hence different query orders result in a different balance of evidence. People evaluating ground beef in a "lean" frame, for example, begin by retrieving favorable evidence before considering unfavorable evidence, and this order results in more favorable evidence being retrieved overall. Both the associative account and query theory are consistent with the common view of attribute framing effects as irrational biases, because surface associations and query order are unrelated to the value of the evaluated item.

An alternative, rational account of attribute framing focuses on the information content of frames (McKenzie & Nelson, 2003; Sher & McKenzie, 2006, 2008). According to this "information leakage" account, a speaker's choice among logically equivalent frames can "leak" relevant information beyond the chosen frame's literal content. For example, comparisons to a known reference point (the initial, typical, or expected level of an attribute) may influence a speaker's frame selection. In particular, speakers are more likely to frame options in terms of attributes that exceed a salient reference point. In one demonstration, McKenzie and Nelson (2003) found that "speaker" participants were more likely to describe a cup with liquid at the halfway mark as "half empty" rather than "half full" when the cup had initially been full (and was therefore relatively empty). "Listener" participants, in turn, consciously or unconsciously "absorbed" the information leaked by the speaker's choice of frame and were more likely to infer that a cup was originally full (rather than empty) when it was described as "half empty" (rather than "half full"). That is, listeners' inferred reference points matched the actual reference points that guide speakers' frame selection. Logically equivalent frames can thus implicitly convey different information. This speaker–listener framework has been used to help explain behavior in other framing contexts such as medical treatment outcomes (McKenzie & Nelson, 2003), time and work on a

*Correspondence to: Lim M. Leong, University of California, San Diego, La Jolla, CA, USA. E-mail: lmleong@ucsd.edu

project (Teigen & Karevold, 2005), and ground beef advertisements (Keren, 2007).

In the information leakage framework, attribute framing effects can arise from the different inferences drawn by listeners exposed to different frames. In particular, high evaluations in a positive frame reflect comparisons with an inferior inferred reference point (e.g., ground beef described as "75% lean" is good because typical ground beef is inferred to be less lean), whereas low evaluations in a negative frame reflect comparisons with a superior inferred reference point (e.g., ground beef described as "25% fat" is bad because typical ground beef is inferred to be less fatty). Note that while the listener's updated beliefs reflect her or his attunement to a subtle linguistic cue, this basis for her or his inference need not be consciously accessible to her or him. Indeed, as Sher and McKenzie (2006) point out, the inferential processes at play are likely to be largely implicit: If the non-equivalence of the information contained in logically equivalent attribute frames was self-evident, these framing effects would hardly have been regarded as problematic for rational models of decision making.

By showing that reference points influence speakers' frame selection, and that frames influence the inferences that listeners draw regarding a speaker's reference point, previous research offers considerable evidence that is consistent with this account. It has not, however, established a causal role of inferences in generating attribute framing effects. In this article, we provide strong evidence for this missing link between frames, inferences, and evaluations by demonstrating that reference point inferences are sufficient for generating attribute framing effects, and that when inferences are likely weaker or absent, framing effects are weaker or absent.

To demonstrate the sufficiency of reference point inferences, Experiment 1 employs a yoking procedure recently developed by Sher and McKenzie (2014) to examine how changes in context lead to different beliefs and how these different beliefs subsequently affect evaluations. We used this procedure to break a standard attribute framing design into two parts. In the first part, we presented "modeler" participants with the target attribute in either one of two frames. They were not asked to form evaluations, however, but to state inferences about the typical value of the attribute. We expected that the positive frame would yield lower estimates of the typical value, in line with findings of McKenzie and Nelson (2003) and Sher and McKenzie (2006). In the second part, each modeler participant was individually yoked to a "recipient" participant who was presented with the modeler's inference about the typical attribute value as part of the background information. The recipients then evaluated the target attribute which, crucially, was always described by *both* frames. In other words, the target attribute was not selectively framed for recipients. Information leakage predicts that these "unframed" recipients should nevertheless exhibit a framing effect in their evaluations: A positive frame presented to a modeler should, by way of the modeler's inference, lead the yoked recipient to provide a more favorable evaluation. Because recipients are provided with different (modeler) inferences, but not different attribute frames, such an effect would indicate that frame-based inferences are sufficient to generate an attribute framing effect.

Whereas Experiment 1 asks whether inferences are sufficient to generate attribute framing effects, in Experiment 2, we examine the converse prediction that blocking inferences should block framing effects. To this end, we measured participants' knowledge in a specific content domain (basketball) and investigated their reactions to attribute frames both in that domain and in an unrelated domain (medical treatments). Those who know more about the content domain should be less influenced by framing in that domain, as their stronger prior beliefs about the typical attribute value limit the scope of frame-dependent inferences. At the same time, knowledge in a specific domain should not preclude participants from being influenced by framing in the unrelated domain. Together, the experiments indicate that the presence of frame-based inferences can generate an attribute framing effect, while their absence can greatly attenuate the effect.

EXPERIMENT 1

Participants read a framing scenario about recruiting a basketball player. "Modeler" participants were presented with the target player's performance framed as either shots "made" or shots "missed" and then reported their estimates (i.e., their inferred "models") of the typical player's performance. "Recipient" participants then received these reference point inferences as part of their background information, and they evaluated the target player described in a neutral ("unframed") manner. To establish a baseline for the framing effect, we also included control conditions, in which participants simply provided evaluations after receiving one frame. The study was designed to test two main predictions. First, we expected that, replicating prior findings (McKenzie & Nelson, 2003; Sher & McKenzie, 2006), different frames would lead modelers to draw systematically different inferences about typical performance. Second, the critical question is then whether the different inferences drawn from the different frames are sufficient to reproduce the "framing effect" among recipients, who all receive the same, neutrally framed description. Finally, although this experiment was not specifically tailored to test the role of knowledge, we asked participants to report their level of general basketball knowledge, expecting those with more knowledge to show a reduced framing effect.

Method

The participants were 414 University of California, San Diego, undergraduate students ($M_{\text{age}} = 20.3$, one participant did not report age; 68% female) who received partial course credit. This sample was obtained by collecting data for a pre-determined period of time (the duration of an academic quarter). The experiment was part of a larger series of unrelated experiments lasting less than an hour. Participants were run at individual computer stations in groups of up to six.

DOI: 10.1002/bdm

Participants were randomly assigned to one of six conditions. In the two Control conditions (Figure 1a), participants read a scenario involving the performance of a basketball player (based on Levin et al., 1998). In the "made" frame condition, participants were asked to

> Imagine that you are a recruiter for a college basketball team. Your job is to search for promising high school basketball players and try to recruit them to your college. You are looking through files for players from local high schools, and you are especially interested in players who can score many points.
>
> The file you are currently looking at shows a player whose performance is quite unusual. This player made 40% of his shots last season.

In the "missed" frame condition, the last sentence instead stated "This player missed 60% of his shots last season." Afterward, participants were asked "How valuable do you think this player would be to your basketball team?" and answered by adjusting a continuous slider scale with a low anchor (*Not at all valuable*) and a high anchor (*Extremely valuable*) (Figure 1a). The numerical value corresponding to a slider position was not visible to participants, but their responses were recorded from 0 to 10 to two decimal places.

The four remaining conditions comprised the yoked design. The two Modeler conditions were exactly the same as the two Control conditions, except that participants reported an inference about the typical player's performance, rather than rating the target player (Figure 1b). Specifically, modelers in the "made" frame condition completed the statement "Typical high school basketball players on average make _____ % of their shots." Modelers in the "missed" frame condition completed the same statement except that the word "make" was replaced with "miss." Thus, the frame condition determined whether the estimate was elicited in terms of "make" or "miss," and this frame matched the one used to describe the target player's performance in the scenario. Participants responded by typing a number ranging from 0 to 100 and were instructed to make their best guess if they were unsure.

The target player's performance was described as "quite unusual" in the scenario because we wanted to discourage modelers from simply restating that player's performance (made 40% or missed 60%) for their inference. The "quite unusual" phrase might also amplify any effect of frame, in both the Modeler and Control conditions, because it indicates that a typical player is far from making 40%/missing 60% of his shots, and the frame is expected to influence the direction in which the inferred reference point is displaced (e.g., if the "made" frame suggests above-average performance, the "quite unusual" phrase would further suggest well-above-average performance).

The two Recipient conditions were identical to the two Control conditions except for two differences (Figure 1c). Each recipient was now provided with a modeler's estimate as part of the background information, and this estimate, along with the target player's performance, was described in a "double frame." Specifically, the sentence "Typical high school basketball players on average make ___ % of their shots and miss ___ % of their shots" was inserted into the background scenario just before the sentence that mentions the target player's performance as "quite unusual," and the target player's performance was described as "This player made 40% of his shots and missed 60% of his shots last

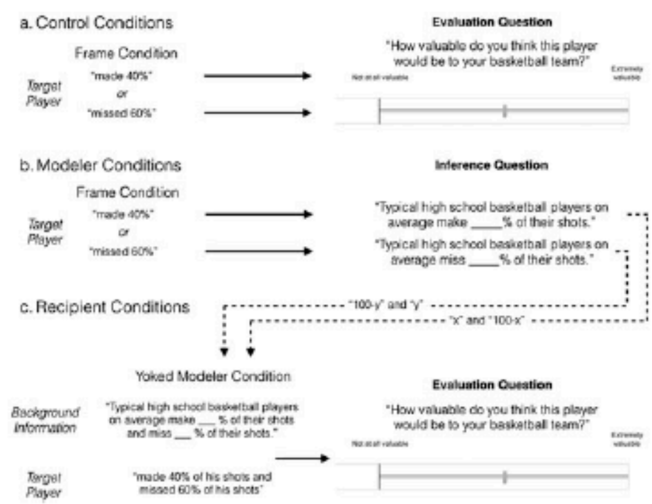

Figure 1. In the Control conditions (a), participants saw the target player described either in the "made" or "missed" frame and evaluated the player. In the Modeler conditions (b), participants saw the target player described in one of the two frames and made an inference about a typical player's performance rather than evaluating the target player. These inferences were then given to yoked participants in the Recipient conditions (c) as part of their background information. For recipients, both the typical and target player's performance were now described in double frames (in the order "made," then "missed"). Recipient participants then evaluated the target player in the same way as in the Control conditions

DOI: 10.1002/bdm

season." The blanks were filled in with the yoked modeler's estimate. For example, if a modeler were in the "made" frame condition and reported that the typical player makes 25% of his shots, then, for the recipient, the respective blanks would be filled in with "make 25%" and "miss 75%." The yoking was implemented such that each modeler's estimate was provided to the next recipient who completed the experiment on the same computer. Thus, the source of the inference (i.e., the frame condition of the yoked modeler) was the only difference between the two Recipient conditions. Recipients were asked to judge how valuable the target player is by using the same slider scale as in the Control conditions.

Demographic information was collected at the end of the experiment. At this stage, participants were also asked "In general, how knowledgeable are you about basketball?" and selected one of four answers (*Not at all knowledgeable*, *Slightly knowledgeable*, *Somewhat knowledgeable*, or *Very knowledgeable*).

Results

Figure 2a shows the mean ratings in the Control conditions by frame. We obtained a standard valence-consistent shift, with the target player judged as more valuable when his or her performance was described in the "made" frame than in the "missed" frame, *M*s = 4.93 and 3.42, $t(136) = 4.75$, $p < .001$, $d = 0.81$.

Next, we analyzed the inferences that modelers drew from the different frames. Although we attempted to discourage modelers from restating the target player's performance by describing it as "quite unusual," eight participants nevertheless provided that as their estimate of the typical player's performance. Their data were excluded from the following analyses because their judgments created an inconsistency in the background blurbs provided to their yoked recipients (i.e., the resulting blurbs described the target player's performance as both typical and unusual). Figures 3a and 3b show the distributions and boxplots of the modelers' estimates of the typical player's performance in the "made" and "missed" frame condition respectively. To facilitate comparisons between the two frame conditions, we transformed the estimates in the "missed" frame condition into estimates of shots *made* by subtracting them from 100. As predicted by the information leakage account, 55% (38/69) of the modelers in the "made" frame condition provided an estimate lower than the target player's performance of 40% shots made, compared with only 28% (17/61) of those in the "missed" condition, $\chi^2(1, N = 130) = 8.73$, $p = .003$, $\varphi = .260$. The mean estimates also exhibited the predicted pattern, with a higher mean estimate in the "missed" frame condition than in the "made" frame condition, *M*s = 49.03 and 42.06, $t(128) = 2.02$, $p = .046$, $d = 0.36$.

For the recipient analyses, the data for those yoked to the eight modelers who were excluded in the previous analyses were also excluded. We first confirmed that recipients were affected by the modeler estimates they received. Collapsing across the two modeler conditions, higher estimates of typical shooting performance led to lower recipient evaluations of the player, $r(128) = -.767$, $p < .001$.

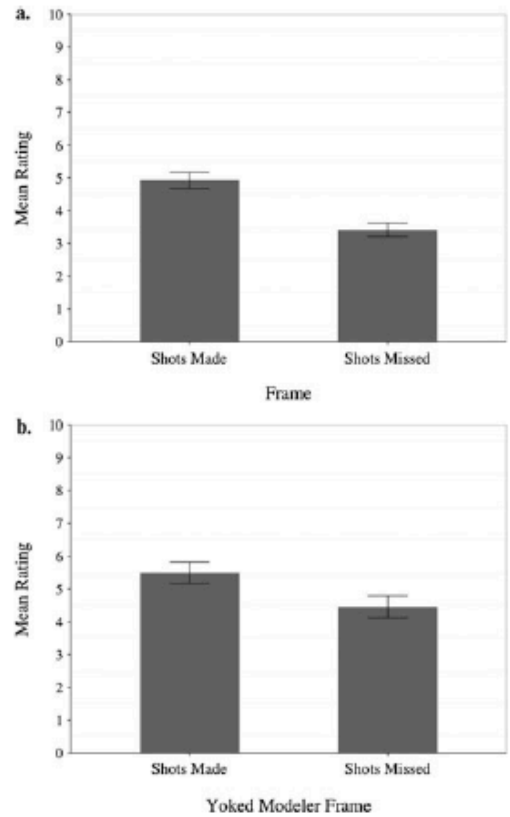

Figure 2. Mean rating of the target player (a) as a function of frame in the control conditions and (b) as a function of the frame condition of the yoked modeler in the recipient conditions. Standard error bars are shown

Thus frames influenced modeler estimates, and modeler estimates influenced recipient evaluations. Putting these two effects together, Figure 2b shows that recipients yoked to modelers in the "made" condition on average judged the target player to be more valuable than recipients yoked to modelers in the "missed" condition, *M*s = 5.50 and 4.46, $t(128) = 2.26$, $p = .025$, $d = 0.40$. Even though the information they received was not subjected to the typical attribute framing manipulation, recipients nevertheless exhibited a "framing effect." This novel effect was somewhat smaller ($d = 0.40$) than was the standard framing effect observed in the control conditions ($d = 0.81$). To more directly compare the two effects, we performed a 2 (Condition: control vs. recipient) by 2 (Frame: made vs. missed) analysis of variance. This analysis revealed a main effect of condition, $F(1, 264) = 8.42$, $p = .004$, $\eta_p^2 = .031$, with higher overall ratings in the recipient conditions. This may in part be due to the use of a double frame for recipients, which has been found to

DOI: 10.1002/bdm

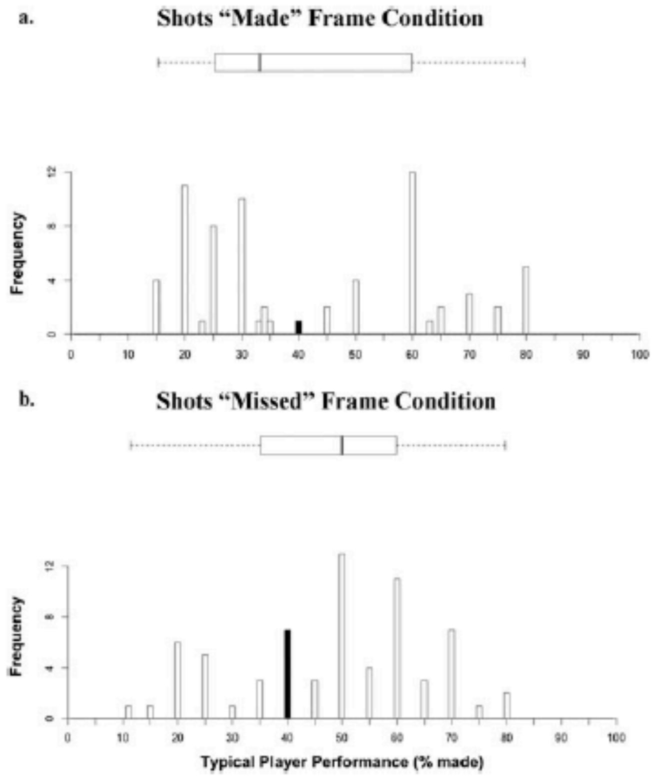

Figure 3. The distribution of modelers' estimates of the typical player's performance in (a) the shots "made" frame condition and (b) the shots "missed" frame condition. The boxplots do not include the eight estimates that are equal to 40

lead to relatively favorable evaluations (Kreiner & Gamliel, 2016). There was also a main effect of frame, $F(1, 264) = 21.21, p < .001, \eta_p^2 = .074$. Compared with their shots "missed" frame counterparts, participants gave higher ratings in the "made" frame (control conditions) or when the inferences came from a modeler presented with the "made" frame (recipient conditions). Importantly, however, the Condition × Frame interaction was not significant, $F(1, 264) = .71, p = .402, \eta_p^2 = .003$. That is, the effect resulting from frame-based inferences is not significantly different from the effect resulting from the frames themselves.[1]

We also analyzed whether the participant's general basketball knowledge interacts with their judgments and inferences. The percentage of participants in the overall sample who self-reported their knowledge as *Not at all knowledgeable*, *Slightly knowledgeable*, *Somewhat knowledgeable*, and *Very knowledgeable* are respectively 26%, 41%, 24%, and 9%. A 2 (Frame: made vs. miss) by 4 (Knowledgeable: not at all vs. slightly vs. somewhat vs. very) analysis of variance was performed separately for the control conditions, modeler conditions, and recipient conditions. In each of these analyses, the interaction between frame and knowledge was not significant (*ps* > .20).

Discussion

We replicated the typical attribute framing effect in our basketball scenario: Participants in the two Control conditions judged the target player as more valuable when his performance was described in the "made" frame than in the "missed" frame. Crucially, participants in the two Recipient conditions, who all received the same wording, also exhibited the framing effect. The only difference between the two Recipient conditions was the source of the inferences: Each recipient saw the inference from a modeler who had seen the target player described in either the "made" frame or the "missed" frame. As predicted by information leakage, modelers who saw the "missed" frame inferred a higher

[1]Including the data from the eight modelers and the recipients yoked to these modelers in the analyses do not qualitatively change any of the results except for one. The difference in the mean estimate of the typical player's performance for modelers in the two frame conditions changed from being significant to being marginally significant, $M_{missed} = 48.10$ vs. $M_{made} = 42.03$, $t(136) = 1.86$, $p = .065$, $d = 0.32$.

DOI: 10.1002/bdm

reference point, or typical performance level, than do modelers who saw the "made" frame. Compared with recipients who were provided with inferences from modelers in the "missed" frame, recipients provided with inferences from modelers in the "made" frame evaluated the target player as more valuable. In sum, the different reference point inferences drawn from the different frames were sufficient to reproduce the attribute framing effect.

Experiment 1 tested for an effect of basketball knowledge on (basketball) attribute framing, and we did not find evidence for such an effect. However, our ability to detect this effect, if it does exist, was limited, as only very few of our participants reported being "Very knowledgeable" about basketball (35/398, or 9%). Moreover, the high school context and the description of the target player's performance as "quite unusual" may have discouraged knowledgeable participants from applying their knowledge to our scenario. These limitations are addressed in our next experiment.

EXPERIMENT 2

While Experiment 1 demonstrated that frame-based inferences are sufficient for attribute framing effects, the goal of Experiment 2 was to examine whether these inferences are necessary for the effect. The prediction is that "turning off" frame-based inferences will attenuate or even abolish attribute framing effects. Experiment 1 attempted to address this question by examining self-reported levels of basketball knowledge, with the expectation that greater prior knowledge would limit the scope of frame-based inferences and hence the size of the framing effect. Though there was no evidence that knowledge affected inferences or target player evaluations, the categorical self-report measure we employed was crude, there were very few "Very knowledgeable" participants, and, as noted above, the special context and "unusual" background description may have discouraged those participants from applying their general knowledge.

Experiment 2 overcomes these shortcomings to provide a proper test of a moderating role of expertise in attribute framing effects. In particular, we recruited participants with varied degrees of basketball knowledge, including a sizeable subset of highly knowledgeable participants. Furthermore, we made two contextual changes to our basketball scenario. First, we adapted the scenario to an NBA context because we assumed that people generally know more about basketball statistics in that setting. Second, we now described the target player's performance in terms of free throw shooting, because we expected that free throw shooting percentages would be more readily interpretable to experts than generic shooting percentages.[2] Those who are knowledgeable should recognize the target player's performance as poor relative to the actual distribution of free throw performance, while those who are not knowledgeable may not.[3] We also measured basketball knowledge via an NBA trivia quiz, which provided an objective measure of knowledge in place of the self-report method used in Experiment 1. Finally, participants were presented with two attribute framing scenarios, one in which basketball knowledge is relevant and another in which it is not. This allowed us to test the specificity of the role of knowledge across different domains.

We expected more knowledgeable participants to show an attenuated framing effect in their domain of expertise. Those who know more about basketball should both score higher on our quiz and have a better idea of what constitutes a typical free throw shooting percentage. They should then be less likely to draw different inferences—and by extension form different evaluations—when performance is described with different frames. In particular, knowledgeable participants should recognize the specific free throw percentage we used to describe the target player as very low for the NBA, regardless of the frame. Finally, knowledge should only be associated with a reduced framing effect in the relevant domain (basketball) and not in an irrelevant domain (medical treatments).

Method

Participants were recruited on Amazon Mechanical Turk in two batches, with a target sample size of 200 for each. After excluding 45 participants with missing responses and duplicate IP addresses, we were left with a final sample of 364. In the first batch ($N = 198$, $M_{age} = 34.8$, one participant did not report age; 35% female), we specifically targeted those who are knowledgeable about NBA basketball. In particular, we requested that "We are looking for NBA fans to read some short scenarios and answer questions about them. Do not accept this HIT if you do not watch the NBA or know about basketball." In the second batch ($N = 166$, $M_{age} = 36.7$, 58% female), we removed this request and did not target any particular population. Recruitment was conducted in this way to help obtain a larger sample of participants knowledgeable in NBA basketball. Data collection for the second batch started 2 days after data collection was completed for the first batch, and those who participated in the first batch were not allowed to participate in the second batch.

Each participant was presented with two framing scenarios in counterbalanced order, one about NBA basketball and the other about a medical treatment. The frame conditions in the two scenarios were orthogonally manipulated (i.e., participants were randomly assigned to one of the four combinations of frames across the two scenarios). For the

[2]The generic shooting percentage previously used may be difficult to interpret owing to differences in playing positions (e.g., guard vs. center), which relate to how often players attempt shots and the distance they shoot from the basket. Using free throw percentage mitigated these problems because all players shoot under the same circumstances (namely, from the free throw line). Also note that the number in the shooting percentage was changed to "made 60%" and "missed 40%" for the two frames.

[3]In the NBA regular season 2015–2016 (during which Experiment 2 was run), only six out of 122 players had a free throw percentage below 60% (http://espn.go.com/nba/statistics/player/_/stat/free-throws/sort/freeThrowPct/seasontype/2/order/false). If participants are knowledgeable about the true underlying distribution of free throw percentage, then they should recognize that a free throw percentage of made 60% or missed 40% is very poor.

DOI: 10.1002/bdm

basketball scenario, participants in the "made" condition were instructed to

> Imagine that you are a scout for an NBA team. Your job is to search for promising basketball players and to draft them to your team. You are looking through the files for potential players in the upcoming draft, and you are only interested in players who are good free throw shooters.
>
> The file you are currently looking at shows a player who, last season, made 60% of his free throws.[4]

In the "missed" condition, the last sentence instead stated that the player "missed 40% of his free throws." Afterward, participants evaluated how valuable the target player is in the same way as the Control conditions in Experiment 1. Also note that, in contrast to Experiment 1, the target player's performance was no longer described as "quite unusual" in this scenario.

The medical treatment scenario we used was based on McKenzie and Nelson (2003). Participants in the "survive" frame condition were instructed to

> Imagine a rare disease that leads to many unpleasant symptoms and can even cause death. The method by which this disease is contracted has been studied, but scientists have yet to identify the exact cause. For the past 20 years, the same treatment has been used in patients with the disease.
>
> A new experimental treatment has been tested, and it has several advantages and disadvantages. In terms of outcome, 85% of patients undergoing this new treatment survive at least 5 years.

In the "die" frame condition, the last sentence instead stated "In terms of outcome, 15% of patients undergoing this new treatment die within 5 years." Afterward, participants rated the effectiveness of the new treatment using the same slider scale as in the basketball scenario except that the low anchor was changed to *Not at all effective* and the high anchor to *Extremely effective*.

After providing a rating for each of the two framing scenarios, participants answered an NBA trivia quiz with six multiple-choice questions, three regarding aspects of the league and three regarding the rules of the game (see Supporting Information). To discourage participants from looking up the answers, they had 10 seconds to respond for each question.

Results and discussion

We first checked whether our targeted recruitment was successful in obtaining a larger proportion of participants knowledgeable about NBA basketball. Figure 4 shows the percentage of participants who received each of the seven possible scores on our quiz in each recruitment batch. As expected, participants recruited in the first batch scored higher on the quiz than did those in the second batch, $Ms = 3.91$ and 2.72, $t(362) = 6.37$, $p < .001$, $d = 0.67$.

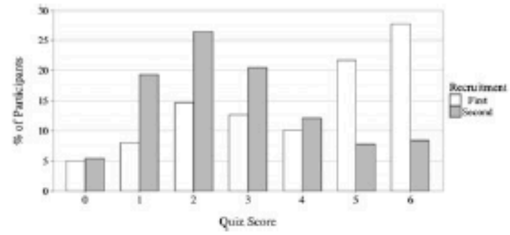

Figure 4. Percentage of participants in the two recruitment batches who received each of the seven possible quiz scores. The sample sizes for the first recruitment and second recruitment are 198 and 166, respectively. The question format on the quiz was multiple choice with four possible answers, and the expected number of correct answers by chance alone is 1.50

Next, we analyzed how basketball knowledge affected the framing effect in the two scenarios. Starting with the basketball scenario, we regressed the ratings of how valuable the target player is on frame and quiz score. Frame was dummy coded with 0 indicating the "missed" frame and 1 indicating the "made" frame, and quiz score indicates the number of correct answers. We predicted that participants with the least basketball knowledge would rate the target player as more valuable in the "made" frame than in the "missed" frame,

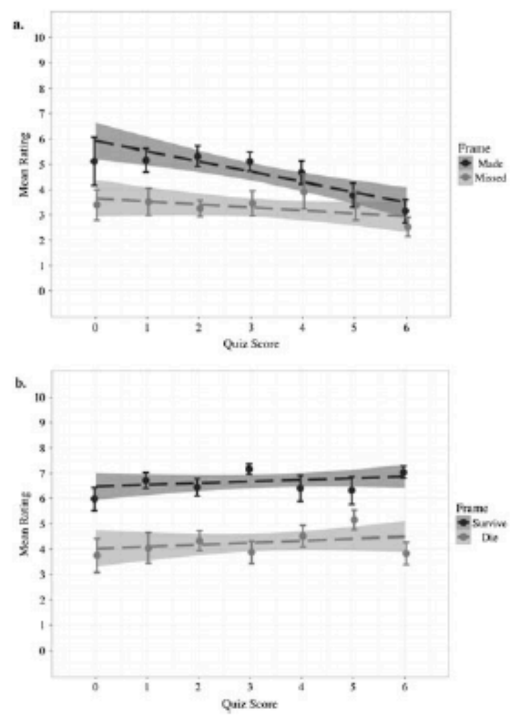

Figure 5. Fitted regression lines and group means as a function of frame and quiz score for (a) the basketball scenario and (b) the medical treatment scenario. Shaded regions represent the 95% confidence interval of the regression lines, and standard error bars for the mean ratings are shown

[4]Note that we assume participants would evaluate the target player's free throw percentage according to NBA standards even though that player has yet to play in the NBA.

J. Behav. Dec. Making (2017)
DOI: 10.1002/bdm

and that this framing effect would decrease for participants with higher quiz scores. Figure 5a shows the fitted regression lines, as well as mean ratings, as a function of quiz score and frame. For participants who did not answer any of the quiz questions correctly, we found a framing effect, $b_{frame} = 2.27$, $t(360) = 4.35$, $p < .001$, and this framing effect decreased as performance on the quiz increased, $b_{frame \times quiz} = -0.28$, $t(360) = -2.10$, $p = .036$. That is, the model predicts a difference of 2.27 in rating between the two frames for participants with the least basketball knowledge, but a difference of only 0.56 for those with the most knowledge. As predicted by information leakage, the participants with little basketball knowledge thus exhibited a sizeable framing effect, and the effect was greatly attenuated for those participants knowledgeable about basketball.[5]

We analyzed the role of basketball knowledge in the medical treatment scenario in the same way. Our dependent variable was the new treatment's rated effectiveness, and frame was dummy coded with 0 indicating the "die" frame and 1 indicating the "survive" frame. We predicted that participants would rate the treatment as more effective in the "survive" condition than in the "die" condition, and that this framing effect would be independent of quiz score. Figure 5b illustrates that participants in the "survive" frame consistently provided higher effectiveness ratings than those in the "die" frame, regardless of their basketball knowledge. As predicted, we found a framing effect for participants who did not answer any of the quiz questions correctly, $b_{frame} = 2.46$, $t(360) = 6.20$, $p < .001$, and the framing effect did not change depending on the level of basketball knowledge, $b_{frame \times quiz} = -0.01$, $t(360) = -0.11$, $p = .91$. The regression model thus predicts a difference of 2.46 in rating between the two frames for participants with the least basketball knowledge, and it predicts a similar difference of 2.38 for those with the most knowledge. Participants more knowledgeable about NBA basketball exhibited a reduced framing effect in the basketball framing scenario, but an unaltered, sizable framing effect in the medical treatment framing scenario.[6] This indicates that it is their basketball expertise, and not something else about the knowledgeable participants, that attenuates the framing effect they exhibit in the basketball scenario.

Finally, we note that the attenuation of the framing effect in the basketball scenario could also be explained "mechanistically" if participants with more basketball knowledge were simply extremely consistent in their judgments: If knowledge constrained the range of experts' judgments, it would also constrain the range of a potential framing effect, whether or not this framing effect is caused by inferences. According to this potential alternative explanation, the responses of participants with higher quiz scores should be less variable than the responses of participants with lower quiz scores, which would lead to heteroskedastic errors in our regression model. We tested for this possibility by performing a White test but did not find evidence against the homogeneity of variance, $\chi^2(4, N = 364) = 1.93$, $p = .75$. Alternative Breusch–Pagan tests that directly assessed heteroskedasticity due to linear or quadratic effects of quiz scores led to the same conclusion ($ps > .25$). These results suggest that the reduction in the framing effect is not merely due to less variability in the responses of the more knowledgeable participants.[7] However, an inferential account naturally explains the full pattern of results.

GENERAL DISCUSSION

In the information leakage framework, attribute framing effects occur because people draw systematically different inferences from different frames. While previous research has demonstrated that frames influence inferences, the causal connection between inferences and framing effects has not been established. In this article, we report two experiments that provide evidence for such a causal relation by establishing that frame-dependent inferences are sufficient to produce an attribute framing effect (Experiment 1), and that expertise, which presumably renders the inferences unnecessary, reduces the effect (Experiment 2). The results of our experiments are not readily explained by an associative account or query theory. Instead, they implicate inferences in the generation and attenuation of attribute framing effects.

Experiment 1 showed, using a yoked design, that inferences from frames are sufficient to generate a standard attribute framing effect. Modeler participants presented with a target player described in the "made" frame, rather than the "missed" frame, inferred that the typical shooting percentage was lower. These results replicate and extend previous findings (e.g., McKenzie & Nelson, 2003; Teigen & Karevold, 2005). Yoked recipient participants then received these inferences as part of their background information, and those yoked to modelers who saw the "made" frame evaluated the fully described (i.e., unframed) target player as more valuable than those yoked to modelers who saw the "missed" frame.

While the results of Experiment 1 provide strong support for an inferential explanation, they are not necessarily inconsistent with an associative account or query theory. For example, if one makes the ancillary assumption that positive associations in the "made" frame lead participants to infer that the typical performance levels are below those of the "positively tagged" target player, then affective associations would be contributing to the inferences that participants draw. However, because strong reference point inferences have been demonstrated in non-evaluative domains (such as rolls of a die or the level of a cup of water; McKenzie & Nelson, 2003; Sher & McKenzie,

[5] At the suggestion of a reviewer, we conducted additional, unplanned analyses to examine the effect of gender (see Supporting Information for details). Male participants in our sample on average scored higher on the basketball quiz than did female participants, and when gender and its interactions were added to the regression model, we found a significant three-way interaction. This suggests that the effect of knowledge on the framing effect differed between women and men. Additional analyses revealed that the predicted pattern was obtained for women but not for men, who did not show a framing effect regardless of quiz performance. Importantly, these gender effects do not affect our theoretical conclusions—the information leakage account predicts that, if there are group differences in framing, the group that exhibits the framing effect should also exhibit the frame by knowledge interaction, which is what we find.

[6] See Supporting Information for additional regression analyses on recruitment batch.

[7] Also see Supporting Information for a table with the means and standard deviations as a function of frame and quiz score for the basketball scenario.

J. Behav. Dec. Making (2017)
DOI: 10.1002/bdm

2006), affective associations or query orders are unlikely to be essential for the pattern of reference point inferences found here. While contributions from other sources cannot be ruled out, the most parsimonious explanation of the full body of findings is that listeners are implicitly attuned to regularities in how speakers select frames: The reference point affects a speaker's choice of frame, and the frame accordingly affects a listener's beliefs about the reference point.

These results on the role of frame-based inferences in attribute framing complement recent findings on the role of sample-based inferences from options in joint–separate reversals (JSRs). JSRs occur when an option that is superior on a difficult-to-evaluate attribute receives high ratings when judged jointly with the alternative option and low ratings when judged in isolation (Hsee, 1996). In a study resembling Experiment 1, Sher and McKenzie (2014) presented modeler participants two options either separately or jointly and then asked them to estimate the mean and range of the difficult-to-evaluate attribute. Modelers drew very different inferences across joint and separate evaluation conditions, and these different inferences were sufficient to reproduce the JSR in recipients, all of whom evaluated only a single option. These results provided support for an "options-as-information" model, according to which JSRs occur not because of different attribute weighting in different "evaluation modes" (joint vs. separate) but because of the different inferences that are drawn from different option samples (Sher & McKenzie, 2014). Sample-based inference likely also contributes to the asymmetric dominance effect (Prelec, Wernerfelt, & Zettelmeyer, 1997; Sher, Müller-Trede, & McKenzie, 2016) and can lead to intransitive behavior in multi-attribute choice (Müller-Trede, Sher, & McKenzie, 2015). When prior knowledge is limited, people appear to draw inferences both from the set of available options and from the way in which those options are framed. Such inferences can, in turn, generate both context effects and framing effects.

Information leakage also predicts that framing effects should be attenuated when frame-based inferences are eliminated. Experiment 2 showed that expertise abated a framing effect in the relevant content domain (NBA basketball) but did not alter a second framing effect in an irrelevant content domain (medical treatments). Decision makers should only draw frame-based inferences about reference points insofar as their prior knowledge of the relevant attribute distribution is limited. Expertise concerning specific attributes reduces framing effects for those attributes. The results of Experiment 2 cannot be easily explained by the associative account or query theory, as it is not clear why frame-based associations or query orders should depend on basketball knowledge.

We note that while expertise reduced the relevant framing effect in Experiment 2, other research has found mixed effects of expertise on judgment and decision making tasks. Some researchers have argued that experts rely on the same heuristics and exhibit the same biases as non-experts (e.g., Tversky & Kahneman, 1971), while others have found that relevant knowledge attenuates biases (e.g., Wilson, Houston, Etling, & Brekke, 1996). Note, however, that from an information leakage perspective, the question is not who exhibits more or less bias, as both non-experts (who fill in the gaps in their imperfect knowledge via frame-based inferences) and experts (who, thanks to their prior knowledge, need not rely as much on frame-based inferences) are behaving reasonably. Seemingly more relevant are studies showing that medical students and even physicians are affected by how treatment outcomes are framed in choice under uncertainty (e.g., McNeil, Pauker, Sox, & Tversky, 1982; McNeil, Pauker, & Tversky, 1988). However, the hypothetical scenarios used in these studies provide minimal context, making it difficult for physicians to apply their domain-specific knowledge. Furthermore, whenever they lack detailed relevant knowledge about specific attributes within the domain, experts, like novices, may rely on the provided frame to fill in the gaps. Domain expertise thus need not attenuate all framing effects broadly related to that domain. Instead, expertise should only attenuate framing effects when specific prior knowledge pre-empts specific inferences that would otherwise be drawn from a speaker's choice of frame.

We further note that, from an information leakage perspective, knowledge of an attribute's distribution should reduce, but need not completely eliminate, attribute framing effects. Related work has generalized the information leakage framework from signaling a speaker's reference point (e.g., whether ground beef is relatively lean) to signaling a speaker's attitude toward the object—a type of implicit recommendation (Sher & McKenzie, 2006). For instance, speakers were more likely to describe a research and development (R&D) team in terms of its "failure" rather than its "success" rate when the team was obviously inept rather than stellar. Moreover, listeners are sensitive to this framing when making decisions about allocating R&D funds. Duchon, Dunegan, and Barton's (1989) participants allocated fewer funds to R&D teams described in terms of their number of unsuccessful projects rather than their number of successful projects. The notion that frames signal implicit recommendations has also been used to explain default effects, because people expect policy makers to select their favored course of action as a default (McKenzie, Liersch, & Finkelstein, 2006). Because frames may signal implicit recommendations in addition to reference points, even experts who are highly familiar with the distribution of a framed attribute may be sensitive to a speaker's choice of frame. Echoing the preceding discussion, this observation supports the general prediction that domain-relevant expertise should often reduce, but not necessarily eliminate, attribute framing effects.

Framing effects have often been regarded as compelling evidence for incoherence and irrationality in human decision making. The underlying assumption is that an option or outcome is the same regardless of how it is described, and thus decision makers should not make different choices or judgments when different descriptions are used. But subtle changes in wording and context may provide task-relevant information, particularly when prior knowledge is limited, and decision makers have been shown to be sensitive to these implicit cues (e.g., Hilton, 1995; Payne, Bettman, & Johnson, 1993; Schwarz, 1994). The experiments reported here suggest that frame-based inferences can account for both the generation and attenuation of the valence-consistent shift. These findings, together with recent work on JSRs (Sher &

DOI: 10.1002/bdm

McKenzie, 2014) and intransitive choice behavior (Müller-Trede et al., 2015), point to an important role for inferences in context and framing effects.

ACKNOWLEDGEMENT

This research was supported by a Scholar Award from the James S. McDonnell Foundation to S. S.

REFERENCES

Duchon, D., Dunegan, K. J., & Barton, S. L. (1989). Framing the problem and making decisions: The facts are not enough. *IEEE Transactions on Engineering Management, February*, 25–27. https://doi.org/10.1109/17.19979.

Hardisty, D. J., Johnson, E. J., & Weber, E. U. (2010). A dirty word or a dirty world? Attribute framing, political affiliation, and query theory. *Psychological Science, 21*, 86–92. https://doi.org/10.1177/0956797609355572.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118*, 248–271. https://doi.org/10.1037//0033-2909.118.2.248.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67*, 247–257. https://doi.org/10.1006/obhd.1996.0077.

Keren, G. (2007). Framing, intentions, and trust–choice incompatibility. *Organizational Behavior and Human Decision Processes, 103*, 238–255. https://doi.org/10.1016/j.obhdp.2007.02.002.

Kreiner, H., & Gamliel, E. (2016). Looking at both sides of the coin: Mixed representation moderates attribute-framing bias in written and auditory messages. *Applied Cognitive Psychology, 30*, 332–340. https://doi.org/10.1002/acp.3203.

Levin, I. P. (1987). Associative effects of information framing. *Bulletin of the Psychonomic Society, 25*, 85–86. https://doi.org/10.3758/BF03330291.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*, 374–378. https://doi.org/10.1086/209174.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior & Human Decision Processes, 76*, 149–188. https://doi.org/10.1006/obhd.1998.2804.

McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science, 17*, 414–420. https://doi.org/10.1111/j.1467-9280.2006.01721.x.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review, 10*, 596–602. https://doi.org/10.3758/BF03196520.

McNeil, B. J., Pauker, S. G., Sox, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine, 30*, 1259–1262. https://doi.org/10.1056/NEJM198205273062103.

McNeil, B. J., Pauker, S. G., & Tversky, A. (1988). On the framing of medical decisions. In D. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making: Descriptive, normative, and prescriptive interactions* (pp. 562–568). New York: Cambridge University Press.

Müller-Trede, J., Sher, S., & McKenzie, C. R. M. (2015). Transitivity in context: A rational analysis of intransitive choice and context-sensitive preference. *Decision, 2*(4), 280–305. https://doi.org/10.1037/dec0000037.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.

Prelec, D., Wernerfelt, B., & Zettelmeyer, F. (1997). The role of inferences in context effects: Inferring what you want from what is available. *Journal of Consumer Research, 24*, 118–126. https://doi.org/10.1086/209498.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. *Advances in Experimental Social Psychology, 26*, 123–162.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101*(3), 467–494. https://doi.org/10.1016/j.cognition.2005.11.001.

Sher, S., & McKenzie, C. R. M. (2008). Framing effects and rationality. In M. Oaksford, & N. Chater (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 79–96). Oxford: Oxford University Press.

Sher, S., & McKenzie, C. R. M. (2014). Options as information: Rational reversals of evaluation and preference. *Journal of Experimental Psychology: General, 143*(3), 1127–1143. https://doi.org/10.1037/a0035128.

Sher, S., Müller-Trede, J., & McKenzie, C. R. M. (2016). Asymmetric information in asymmetric dominance: The informational value of worthless options. Manuscript in preparation.

Teigen, K. H., & Karevold, K. I. (2005). Looking back versus looking ahead: Framing of time and work at different stages of a project. *Journal of Behavioral Decision Making, 18*, 229–246. https://doi.org/10.1002/bdm.502.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110. https://doi.org/10.1037/h0031322.

Wilson, T. D., Houston, C., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General, 125*, 387–402. https://doi.org/10.1037/0096-3445.125.4.387.

Authors' biographies:

Lim Leong is a graduate student in the Department of Psychology at the University of California, San Diego. His research interests include decision making, rationality, production and perception of randomness, and language.

Craig R. M. McKenzie received his PhD in Psychology from the University of Chicago and is now a professor in the Rady School of Management and in the Department of Psychology at the University of California, San Diego. His research interests include decision making, inference, rationality, and creativity.

Shlomi Sher received his PhD in Psychology from Princeton University and is now an assistant professor in the Department of Psychology at Pomona College. His research interests include decision making, rationality, and consciousness.

Johannes Müller-Trede holds a PhD in Economics from Universitat Pompeu Fabra and is now assistant professor in the Department of Managerial Decision Sciences at IESE Business School, Barcelona. His research focuses on the performance, the psychology, and the rationality of people's decision making.

Authors' addresses:

Lim M. Leong, University of California, San Diego, La Jolla, CA USA.

Craig R. M. McKenzie, University of California, San Diego, La Jolla, CA USA.

Shlomi Sher, Pomona College, Claremont, CA USA.

Johannes Müller-Trede, University of California, San Diego, La Jolla, CA USA.

Chapter 1, in full, is a reprint of the material as it appears in the Journal of Behavioral Decision Making, 2017, Leong, Lim M.; McKenzie, Craig R. M.; Sher, Shlomi; Müller-Trede, Johannes. The dissertation author was the primary investigator and author of this paper.

Chapter 2:

Adaptability of attribute framing effects

Lim M. Leong, & Craig R. M. McKenzie

(manuscript under review for publication)

Abstract

Prior research has demonstrated that attribute framing effects arise from the different inferences that listeners draw from the speaker's choice of frame. While attribute framing is adaptive in the typical communicative environment in which speakers systematically choose frames to convey relative abundance, it is less clear whether listeners can appropriately change their behavior when the choice of frame is uninformative. In our first three studies, we used a hypothetical vignette that described how the frame was selected. In Study 1, we found that informing participants that a computer has randomly chosen a frame did not reduce the framing effect. In Studies 2a and 2b, we added an external constraint that made it harder for the speaker to use one frame over another. Despite manipulating how much choice the speaker had in using a particular frame, we again found that the framing effect remains unaltered. In our final study, we placed speaker and listener participants in a zero-sum repeated communication game and manipulated the incentive structure. In contrast to our previous results, we found that when their interests were misaligned, listeners no longer responded systematically to frames. Taken together, our results suggest that the adaptability of attribute framing depends on how the change in the task environment is implemented.

Keywords: framing effect, inference, information leakage, adaptability, rationality

Adaptability of Attribute Framing Effects

Imagine that you and your friend are grocery shopping. Your friend picks up a package of ground beef and tells you that it is "25% fat." Given this description, you are likely to judge it as fattier than other ground beef. Now imagine that your friend had instead described the same ground beef as "75% lean." In this case, you are likely to judge it as relatively lean. This example illustrates an attribute framing effect, in which people respond more favorably when a characteristic of an object is described in a positive frame than a negative frame (Levin & Gaeth, 1988; Levin, Schneider, & Gaeth, 1998). The traditional view is that framing effects are irrational because logically equivalent descriptions should not lead to divergent judgments and decisions (Shafir & LeBoeuf, 2002; Tversky & Kahneman, 1986).

The logical equivalence of frames, however, does not guarantee information equivalence. One account of attribute framing effects argues that frames can communicate information beyond their logical content (McKenzie & Nelson, 2003; Sher & McKenzie, 2006). In particular, a speaker's choice of frame can leak implicit information such as reference points, and listeners can absorb this information through the different inferences they draw from frames. For example, "speaker" participants are more likely to describe a cup with water at the half-way mark as "half-empty" (rather than "half-full") when the cup was previously full (McKenzie & Nelson, 2003). In turn, "listener" participants are more likely to infer that a cup of water was previously full when it was described as "half-empty" (rather than "half-full"). If a speaker's choice of frame leaks relevant information, then it is not irrational to respond differently to different frames. Thus, rather than exemplifying a flaw in the cognitive system, attribute framing effects may reflect people's remarkable ability to extract information from subtle linguistic regularities. This information leakage framework helps explain why people respond differently even to complementary frames presented immediately one after another (Aczel, Szollosi, & Bago, 2018), and why robust attribute framing effects are found in a variety of domains (e.g., Keren, 2007; Teigen & Karevold, 2005; also see Table 3 in Levin et al., 1998).

The preceding analysis illustrates that attribute framing effects are *adaptive* – that is, they reflect behavior well-suited to the typical environment in which speakers systematically select frames to signal relative abundance (e.g., relatively lean, relatively full). What happens, though, if frames are selected in a way that renders them uninformative? Will listeners be able to adjust their behavior appropriately? That is, are listeners' responses to frames *adaptable* (McKenzie, 2005)? In the present research, our main question is whether listeners can alter their inferences from frames, and exhibit reduced attribute framing effects, when the speaker's choice of frame is uninformative.

One simple and straightforward way to test this is to explicitly describe how the frame was chosen. By telling listeners that the speaker chose the frame at random or had no choice but to use a particular frame, the regularities that warrant frame-based inferences should be severed, rendering the choice of frame uninformative. Related studies in social cognition that employed similar manipulations (e.g., Hilton, 1995; Schwarz, 1994; Schwarz, Strack, Hilton, & Naderer, 1991) suggest that people are sensitive to changes to conversational norms and can alter their judgments appropriately. Thus, if listeners are responsive to descriptions about an uninformative frame selection process, then they should show attenuated framing effects.

Alternatively, listeners may fail to exhibit reduced framing effects in response to this type of manipulation. At first glance, this may seem inconsistent with information leakage. However, we believe this would simply illustrate that listeners are unable to combine their explicit knowledge about frame selection with the underlying inferential process. This possibility was raised by Sher & McKenzie (2006): “There is, perhaps, an analogy with sensitivity to subtle shifts in facial expression. If Bob knows that the almost imperceptible upward curls at the corners of Sue’s mouth are due to a congenital nervous disorder, does this knowledge suppress Bob’s impression, otherwise justified, that Sue is happy?” (p. 489). The inferential process that give rise to framing effects is likely fast and largely operates at an unconscious level. That is, listeners may not actively reason about why the speaker’s choice of frame is informative. If listeners are not consciously aware of how frames leak information in the typical environment, then it would not be surprising that explicit knowledge about uninformative frames in an atypical environment would have little effect.

The adaptability of framing effects may thus depend on how the uninformative frame selection process is implemented. In particular, listeners may be more responsive to manipulations that tap into their implicit knowledge. To address this possibility, we use a paradigm adapted from research that examined people’s ability to produce random sequences. When people are explicitly instructed to generate random sequences, they are poor at doing so (Nickerson, 2002; Bar-Hillel & Wagenaar, 1991). However, they are more successful at producing random sequences when placed in a two-person zero-sum repeated game (e.g., a “matching pennies” game) in which a mixed strategy of random responding is optimal (e.g., Rapoport & Budescu, 1992). Adapted to a communication game between speakers and listeners, the optimal strategy for speakers incentivized not to convey useful information is to select frames at random. In turn, the listener’s optimal strategy is to ignore the frame and respond at random. Rather than telling listeners that the speaker’s choice of frame is uninformative, this paradigm allows listeners to draw on their knowledge about the speaker’s uncooperative communicative intentions. By using a variety of different manipulations to render the choice of frame uninformative, we provide a more comprehensive test of the adaptability of framing effects.

In four studies, we examined the adaptability of attribute framing effects with respect to how frames are selected. Specifically, we tested whether listeners could alter their inferences and judgments when the speaker's choice of frame is rendered uninformative. To that end, we employed two different types of paradigms: Our first three studies used hypothetical vignettes that described how the frame was chosen while our last study used a two-player interactive game. Study 1 and Study 2b were preregistered. We report all relevant measures and exclusions, and all data, materials, and preregistrations are available at https://osf.io/pqxfu/?view_only=7a47fbde47f5400f8836df0b1ae2e8d7.

Study 1

Study 1 employed a straightforward way to test whether people are sensitive to how frames are chosen: Participants were told that rather than a human speaker, a computer had randomly selected a frame to use. Because the frame was selected randomly, the choice of frame does not convey any relevant information. Adapting the basketball scenario from Leong, McKenzie, Sher, and Müller-Trede (2017), we manipulated the frame (positive vs. negative) and the frame selection process (nonrandom vs. random). If listeners are sensitive to how frames are selected, then they should not show a framing effect when the frame was selected randomly.

Method

Four hundred and thirty-nine UCSD undergraduates (M_{age} = 21.03; 57% female) participated for partial course credit and were randomly assigned to one of four conditions. Participants were asked to imagine being a recruiter for a basketball team. In the "nonrandom" conditions, they read that a fellow recruiter had decided how to describe the performance of each player. Specifically, participants in the "positive frame" condition read: "For each player, a fellow recruiter has chosen how to phrase the player's performance – either in terms of '% of shots made' or '% of shots missed.' The file you are currently looking at shows that the player 'made 40%' of his shots last season." In the "negative frame" condition, they instead read that the player "missed 60%" of his shots. In the "random" conditions, participants read the same vignette except that "a computer has randomly determined how to phrase the player's performance" instead of "a fellow recruiter has chosen how to phrase the player's performance." Afterward, they were asked "How valuable do you think this player would be to your basketball team?" on a scale from "not at all valuable" (1) to "extremely valuable" (10). Finally, participants provided basic demographic information.

Results & discussion

Figure 1 shows the mean ratings for each condition. A two-way ANOVA revealed a main effect of frame, $F(1, 435) = 23.68$, $p < .001$, $\eta_p^2 = .052$, no main effect of selection process, $F(1, 435) = 0.02$, $p = .88$, $\eta_p^2 < .001$, and no frame-by-selection interaction, $F(1, 435) = 0.47$, $p = .50$, $\eta_p^2 = .001$. Although participants judged the basketball player as more valuable when his performance was described in the "made" frame than the "missed" frame, they did so regardless of whether the frame was selected nonrandomly by a human speaker or randomly by a computer. This suggests that listeners are unable to incorporate their knowledge about frame selection into their judgments and simply continued to respond as if the frame was still informative.

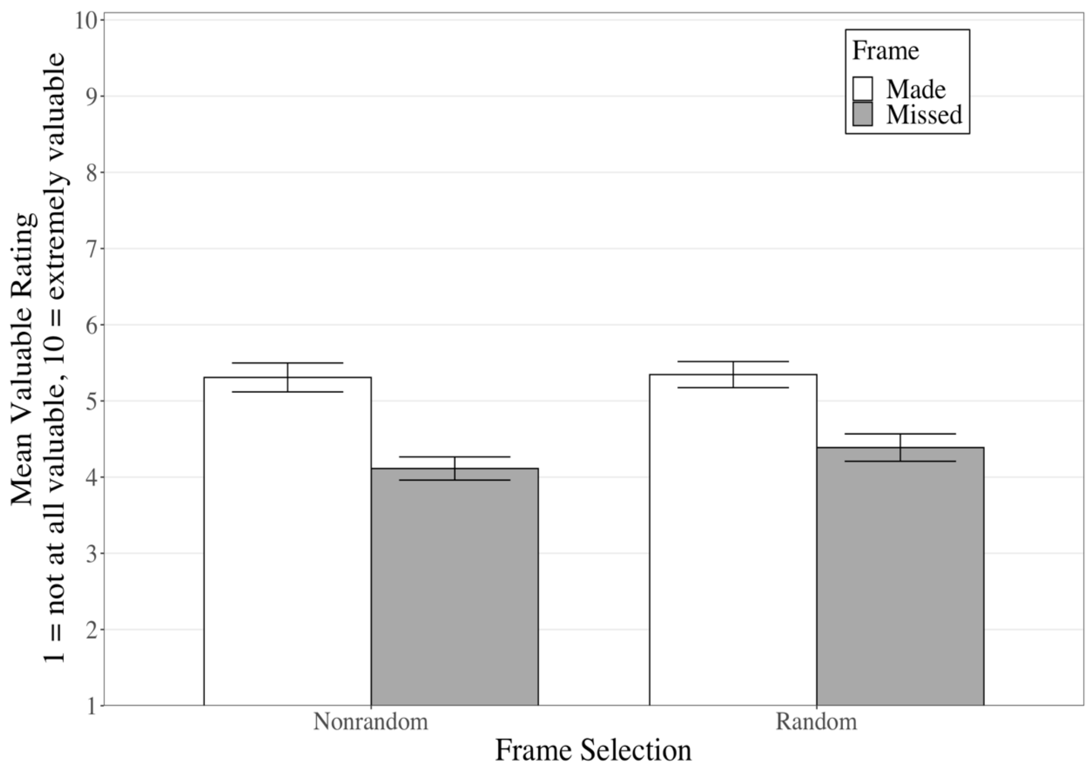

Figure 2.1. Mean ratings of target player as a function of frame and frame selection process.

Study 2a

In the next two studies, we manipulated the informativeness of frames by introducing an external constraint that made it harder for the speaker to use one frame over another. In Study 2a, the speaker chose the frame that was made easier to use by the constraint, which reduces the frame's informativeness. We manipulated the frame that was used (positive vs. negative) and the frame selection process (less informative vs. control). If listeners understand that the constraint reduces the informativeness of the speaker's choice of frame, then they should show an attenuated framing effect.

Method

Four hundred and thirty-six UCSD undergraduates (M_{age} = 19.89; 68% female) participated for partial course credit and were randomly assigned to one of four conditions. Participants were asked to imagine that they are grocery shopping with their friend, and that their friend “picks up a package of ground beef that happens to have a partially torn label.” They were shown an accompanying image of what the partially torn label looks like (see Figure 2). In the control conditions, the tear on the label does not prevent the friend from reading its lean/fat description. In the “positive frame” condition, the friend describes the ground beef as “85% lean” while in the “negative frame” condition, the friend describes it as “15% fat.” In the “less informative” conditions, participants read the same vignette but the accompanying image shows that the tear precludes the friend from reading one of the two descriptions. Specifically, in the “positive frame” condition, the “fat” description is missing from the label (Figure 2c) and the friend describes the ground beef as “85% lean.” Likewise, in the “negative frame” condition, the “lean” description is missing (Figure 2d) and the friend describes the ground beef as “15% fat.” Afterward, participants were asked “Given how your friend described the ground beef, how *lean/fat* do you think the ground beef is, *relative to other ground beef on the market*?” on a scale from “much leaner than most” (0) to “much fattier than most” (10). Finally, participants provided basic demographic information.

Results & discussion

Figure 3 shows the mean lean/fat ratings for each condition. A two-way ANOVA revealed a main effect of frame, $F(1, 432) = 16.59$, $p < .001$, $\eta_p^2 = .037$, no main effect of selection process, $F(1, 432) = 0.07$, $p = .79$, $\eta_p^2 < .001$, and no frame-by-selection interaction, $F(1, 432) = 0.64$, $p = .42$, $\eta_p^2 = .001$. While participants judged the ground beef as fattier when it was described in the “fat” frame than the “lean” frame, they did so regardless of whether a tear on the label reduces the amount of choice the speaker had in using a particular frame. This suggests that people are not sensitive to how the informativeness of frames is reduced when the choice of frame is constrained.

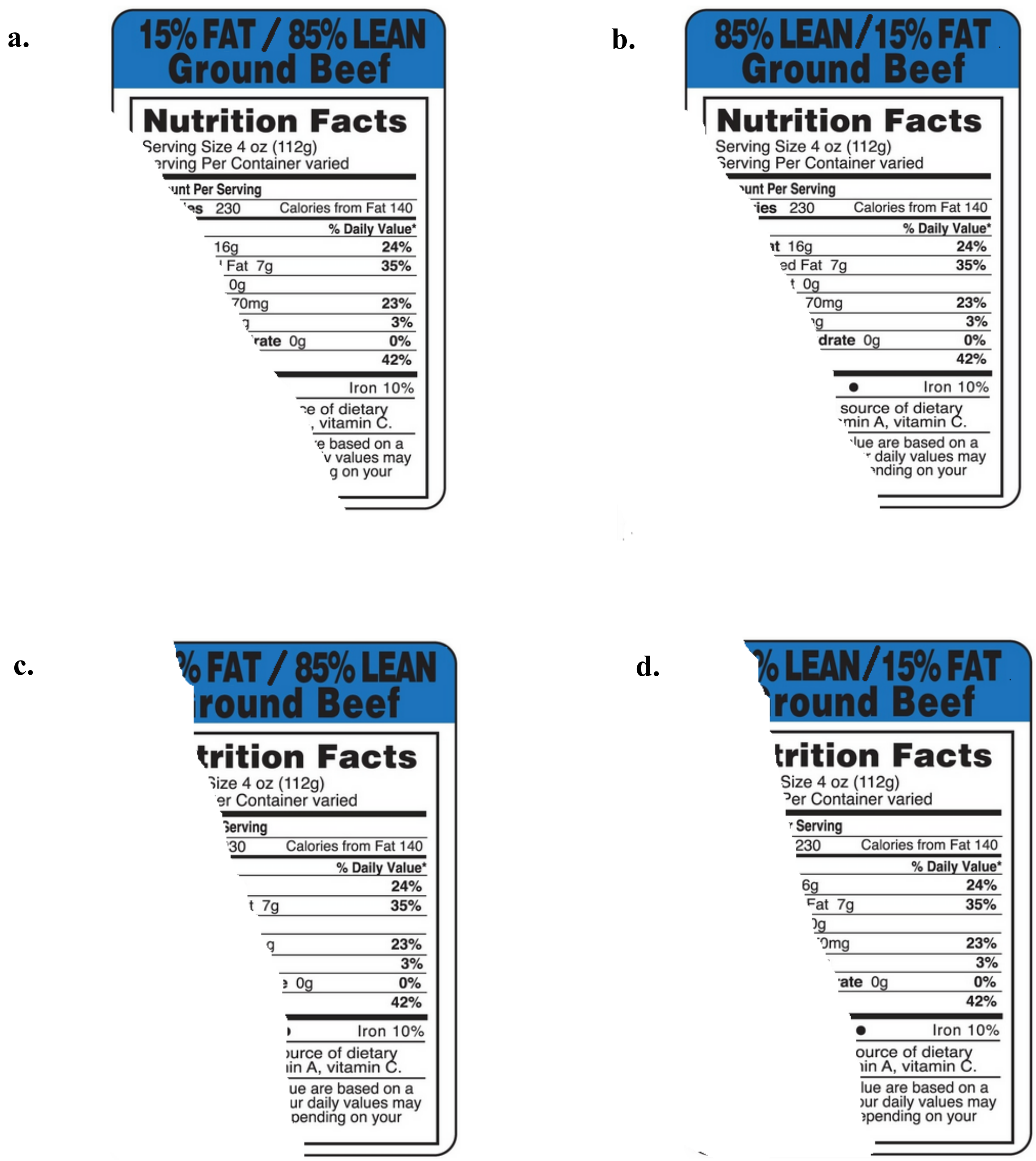

Figure 2.2. In the control conditions in Study 2a, participants saw label (a) in the positive frame condition and label (b) in the negative frame condition. In the “less informative” conditions in both Study 2a and 2b, participants saw label (c) in the positive frame condition and label (d) in the negative frame condition. In the “more informative” conditions in Study 2b, participants saw label (d) in the positive frame condition and label (c) in the negative frame condition.

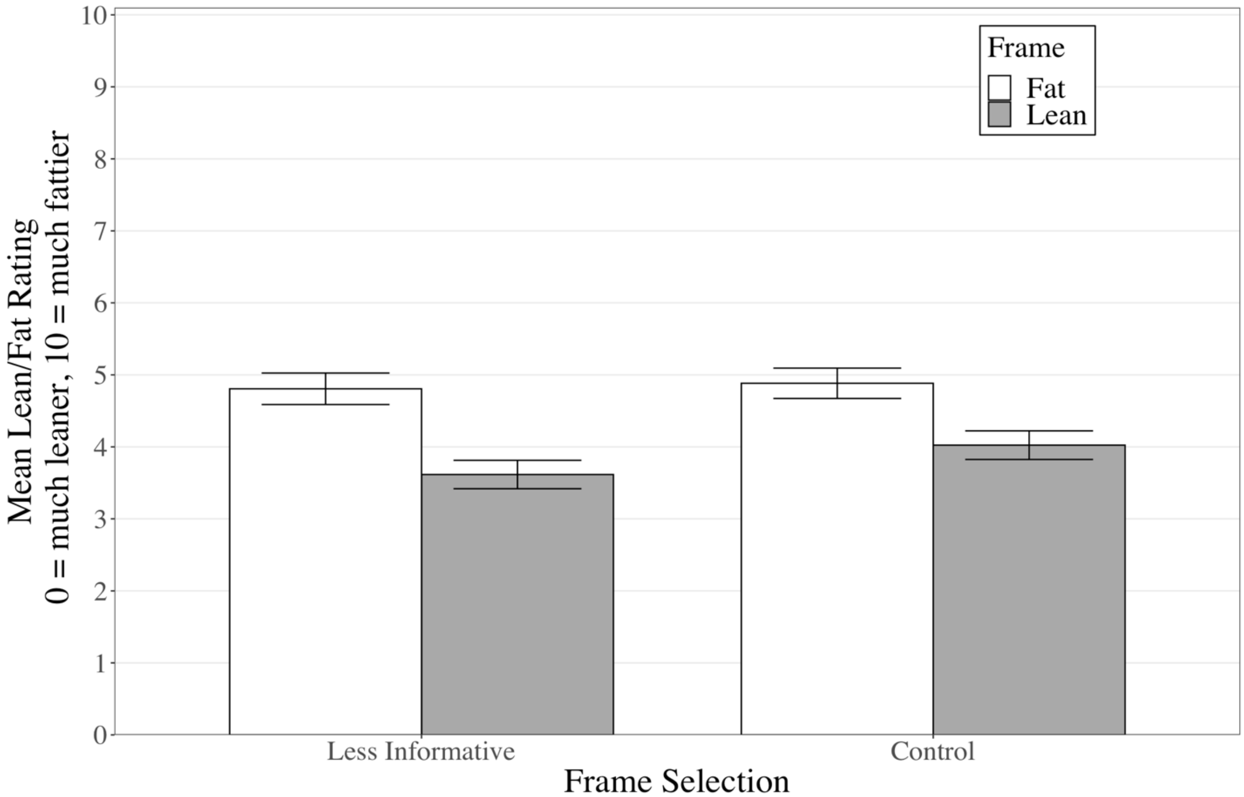

Figure 2.3. Mean ratings of ground beef as a function of frame and frame selection.

Study 2b

While Study 2a tested whether framing effects would be attenuated when a constraint reduces the informativeness of the frame, Study 2b examined the opposite: The speaker chooses the frame that is made harder to use by the constraint, which increases the frame's informativeness. Similar to the design of Study 2a, we manipulated the frame that was used (positive vs. negative) and the frame selection process (less informative vs. more informative). In the new "more informative" condition, if listeners understand that the constraint increases the informativeness of the speaker's choice of frame, then they should show a larger framing effect.

Method

Three hundred and eighty-one UCSD undergraduates (M_{age} = 20.02, two did not report age; 73% female) participated for partial course credit and were randomly assigned to one of four conditions.[1] Participants read the same ground beef vignette used in Study 2a, and the "less informative" conditions were the same as before. In the new "more informative" conditions, the tear again precluded the friend from reading one of the two descriptions, yet the friend nevertheless chooses to the use the missing frame. Specifically, in the "positive frame" condition, the "lean" description is missing from the label (Figure 2d), yet the friend describes the ground beef as "85% lean." Likewise, in the "negative frame" condition, the "fat" description is missing from the label (Figure 2c), yet the friend describes the ground beef as "15% fat." Afterward, participants were asked "How healthy do you think this ground beef is?" on a scale from "not at all healthy" (1) to "extremely healthy" (10). Note that we also changed how we asked this question from Study 2a to avoid repeating the framed descriptors (i.e., lean/fat) in the question. Finally, participants provided basic demographic information.

Results & discussion

Figure 4 shows the mean healthy ratings for each condition. A two-way ANOVA revealed a main effect of frame, $F(1, 377) = 8.01$, $p = .005$, $\eta_p^2 = .021$, no main effect of selection process, $F(1, 377) = 0.18$, $p = .67$, $\eta_p^2 < .001$, and no frame-by-selection interaction, $F(1, 377) = 0.08$, $p = .78$, $\eta_p^2 < .001$. While we again replicated the basic framing effect, this effect did not vary depending on the presence of a constraint in frame selection. Replicating and extending the results of Study 2a, we observed neither a reduced nor enlarged framing effect when the choice of frame is made, respectively, less and more informative by an external constraint.

[1] The target sample size in our preregistration was 400 total participants. Unfortunately, due to a pandemic, we did not reach our target before the end of the academic term.

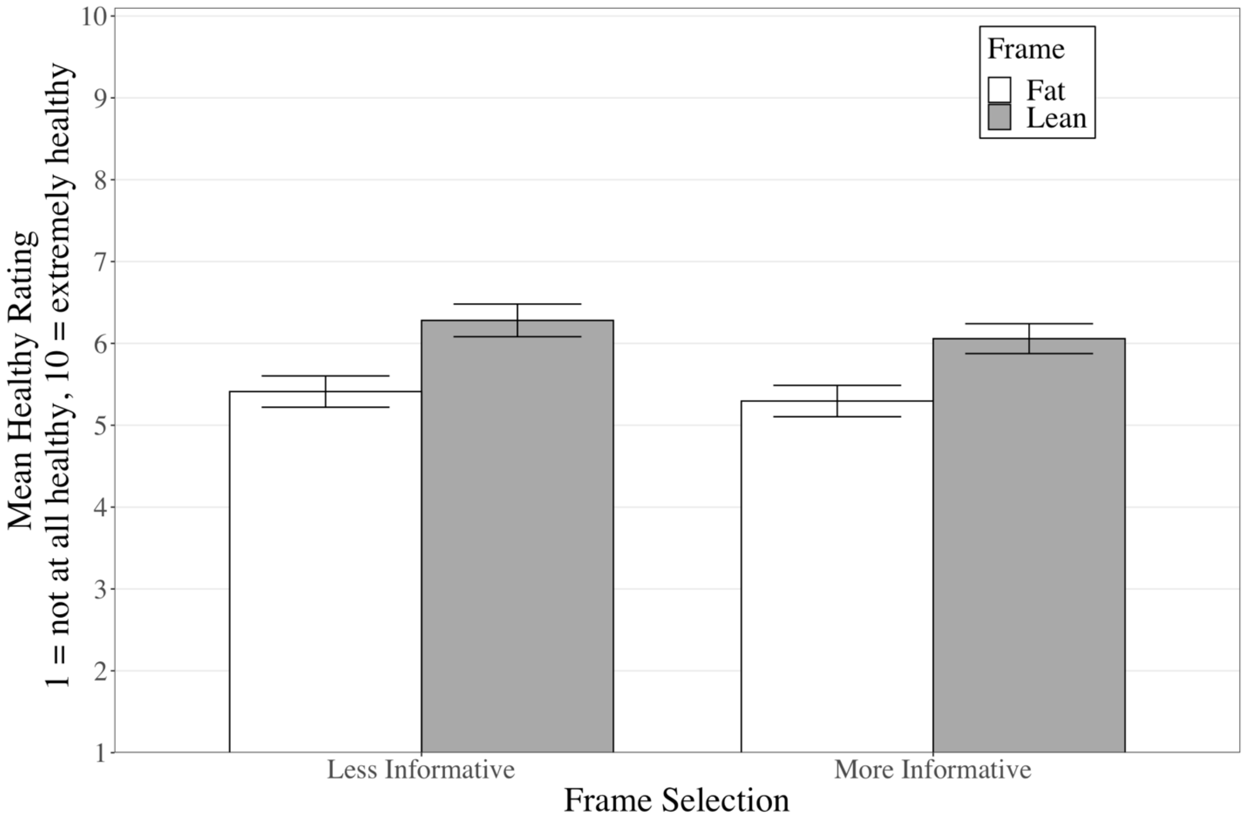

Figure 2.4. Mean ratings of ground beef as a function of frame and frame selection process.

Study 3

The results of our studies so far suggest that attribute framing effects are not adaptable, at least to manipulations that explicitly describe how the frame was selected. In Study 3, we used a different paradigm: Rather than telling participants how the frame was selected, we manipulated whether the speakers' and listeners' interests are aligned or misaligned in a communication game. When their interests are aligned, as in typical communicative contexts, we expect speakers to choose frames in an informative manner and listeners to respond systematically to frames. In contrast, when their interests are misaligned, listeners should recognize that speakers would strive to make their choice of frame uninformative, and listeners should therefore cease to respond systematically to different frames.

Furthermore, our design includes a number of additional features that might make listeners more sensitive to the informativeness of frames. First, listeners interact in real-time in speaker-listener pairs rather than respond to hypothetical vignettes. We believe that placing listeners in a real communicative environment would better tap into their intuitions about how they should react. Second, the interaction is repeated over many trials with feedback after each trial. This affords listeners the opportunity to learn and to update their beliefs about the structure of the environment. Third, we use a within-subjects design rather than a between-subjects design. This makes our manipulation of the incentive structure, and how speakers might change their frame selection strategy, more salient to listeners. By using a vastly different design than our previous ones, we provide a broader test of whether framing effects are adaptable.

Method

One hundred and forty-six UCSD undergraduates (M_{age} = 19.82; 71% females) participated for partial course credit. In groups of 2, 4, or 6, they were randomly assigned to the role of "speaker" or "listener" (referred to as "messenger" and "guesser" in the materials), and each speaker was randomly paired with a listener. These roles and pairings remained the same throughout the session.

Speakers and listeners played a communication game regarding the relation between two cups in real-time via a computer interface (see Figure 5). At the beginning of each trial, speakers were presented with complete information about the fullness/emptiness of a "target cup" and a "hidden cup." They are asked to choose which one of two "messages" to send the listener: one that describes the target cup in the "full" frame or one that describes it in the "empty" frame. The hidden cup thus serves as a potential reference for framing the target cup. After receiving their respective speaker's frame, which was their only information about the cups, listeners were asked to complete the statement “The target cup is ___ than the hidden cup” with either “fuller” or “emptier.” At the end of each trial, both speakers and listeners were reminded of the message (frame) and guess (inference), and received feedback as to whether the guess was correct.

Participants first played two guided practice trials to familiarize themselves with their roles. Afterward, they played the communication game in two blocks with 16 trials each for a total of 32 trials. The first block constituted the “aligned interests” condition, and the second block constituted the “misaligned interests” condition. In the aligned interests condition, both speakers and listeners earned 1 point if the listener guessed correctly. In the misaligned interests condition, speakers earned 1 point if listeners guessed incorrectly, while listeners earned 1 point if they guessed correctly. The order of presentation for the trials were fixed (see OSF page for full materials). For half of the trials, the hidden cup was fuller than the target cup, so the correct answer was “emptier.” For the remaining half, the hidden cup was emptier than the target cup, so the correct answer was “fuller.” At the end of each block, participants were asked to explain the strategy they used in the game and the strategy they think their partner used. Finally, at the end of both blocks, participants provided basic demographic information.

Results & discussion

We first analyzed the overall performance in the communication game. In the aligned interests condition, listeners were on average correct 77.6% of the time (12.41 out of 16), with 32.9% (24/73) of them getting perfect scores. In contrast, in the misaligned interests condition, listeners were on average only correct 54.3% of the time (8.68 out of 16), with none of them getting a perfect score. The higher performance in the aligned interests condition compared to the misaligned interests condition was statistically significant, $M_{diff} = 3.73$, $t(72) = 8.93$, $p < .001$, 95% CI [2.89, 4.56].

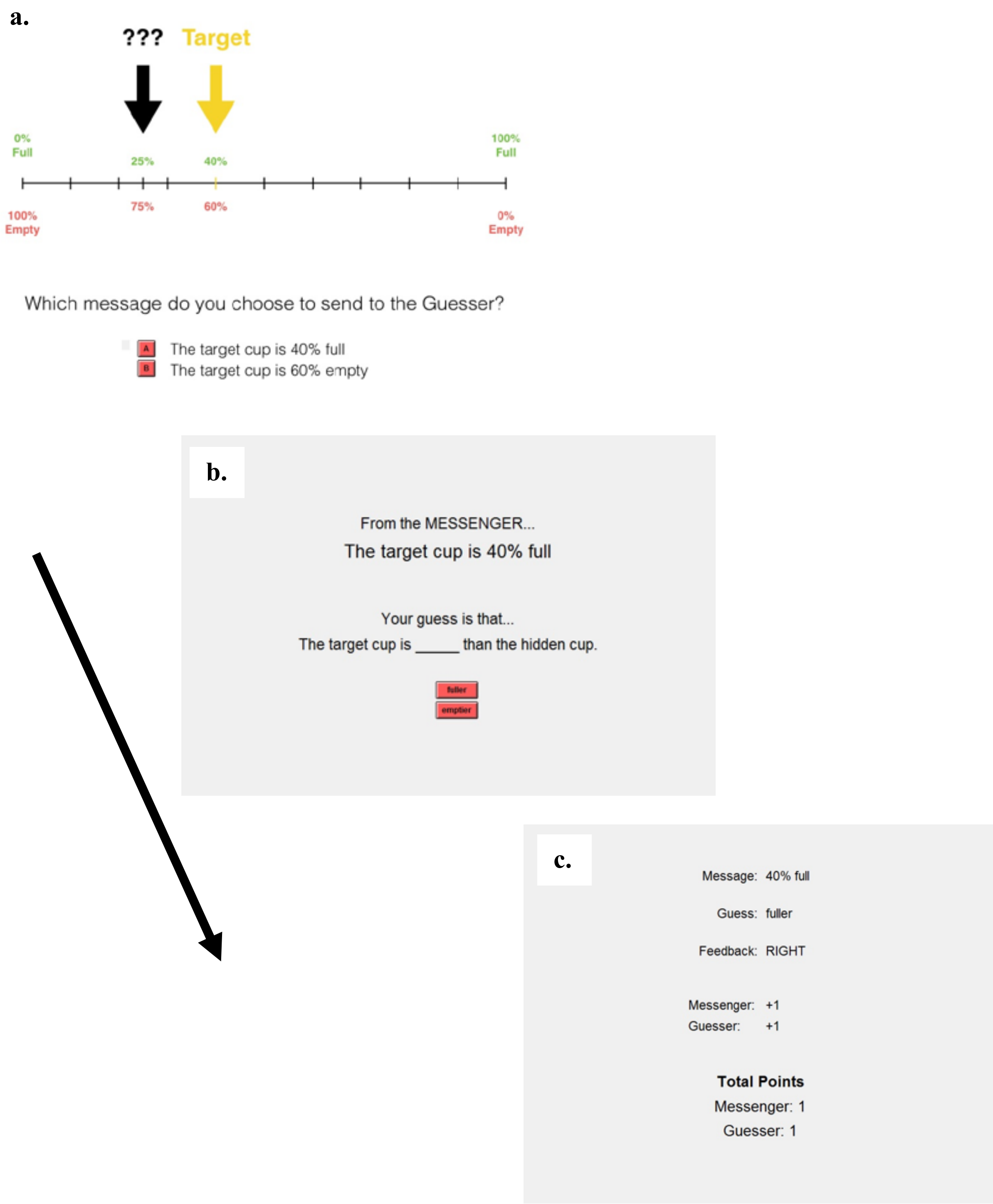

Figure 2.5. Example of one trial in the communication game with the speaker screen (a), the listener screen (b), and the joint feedback screen (c).

Next, we examined whether this higher performance reflects how speakers chose frames. Figure 6a shows the percentage of trials speakers selected the "full" frame as a function of reference point (whether the hidden cup was fuller or emptier than the target cup) for each condition. In the aligned interests condition, speakers used the "full" frame 76% of the time when the reference point was low (hidden cup was emptier) and 27% of the time when it was high (hidden cup was fuller; $p < .001$; Fisher's exact test). In contrast, in the misaligned interests condition, speakers used the "full" frame 58% of the time when the reference point was low and 60% of the time when it was high ($p = 0.44$; Fisher's exact test). Overall, speakers selected frames consistent with information leakage 75% of the time in the aligned interests condition, but only 49% of the time in the misaligned interests condition, $\chi^2(1, N = 2336) = 163.02$, $p < .001$. Together, this demonstrates that speakers can transmit implicit information to listeners when their interests are aligned and, crucially, they can also stop doing so when their interests are misaligned.

Next, we examined how listeners responded to the speaker's frame. Figure 6b shows the percentage of trials listeners guessed "fuller" as a function of frame for each condition. In the aligned interests condition, listeners guessed that the target cup was "fuller" than the hidden cup 76% of the time when given the "full" frame and 33% of the time when given the "empty" frame ($p < .001$; Fisher's exact test). In contrast, in the misaligned interests condition, listeners guessed "fuller" 51% of the time when given the "full" frame and 50% of the time when given the "empty" frame ($p = .72$; Fisher's exact test). Overall, listeners responded consistent with information leakage 71% of the time in the aligned condition, but only 51% of the time in the misaligned condition, $\chi^2(1, N = 2336) = 105.47$, $p < .001$. Mirroring the speaker results, this

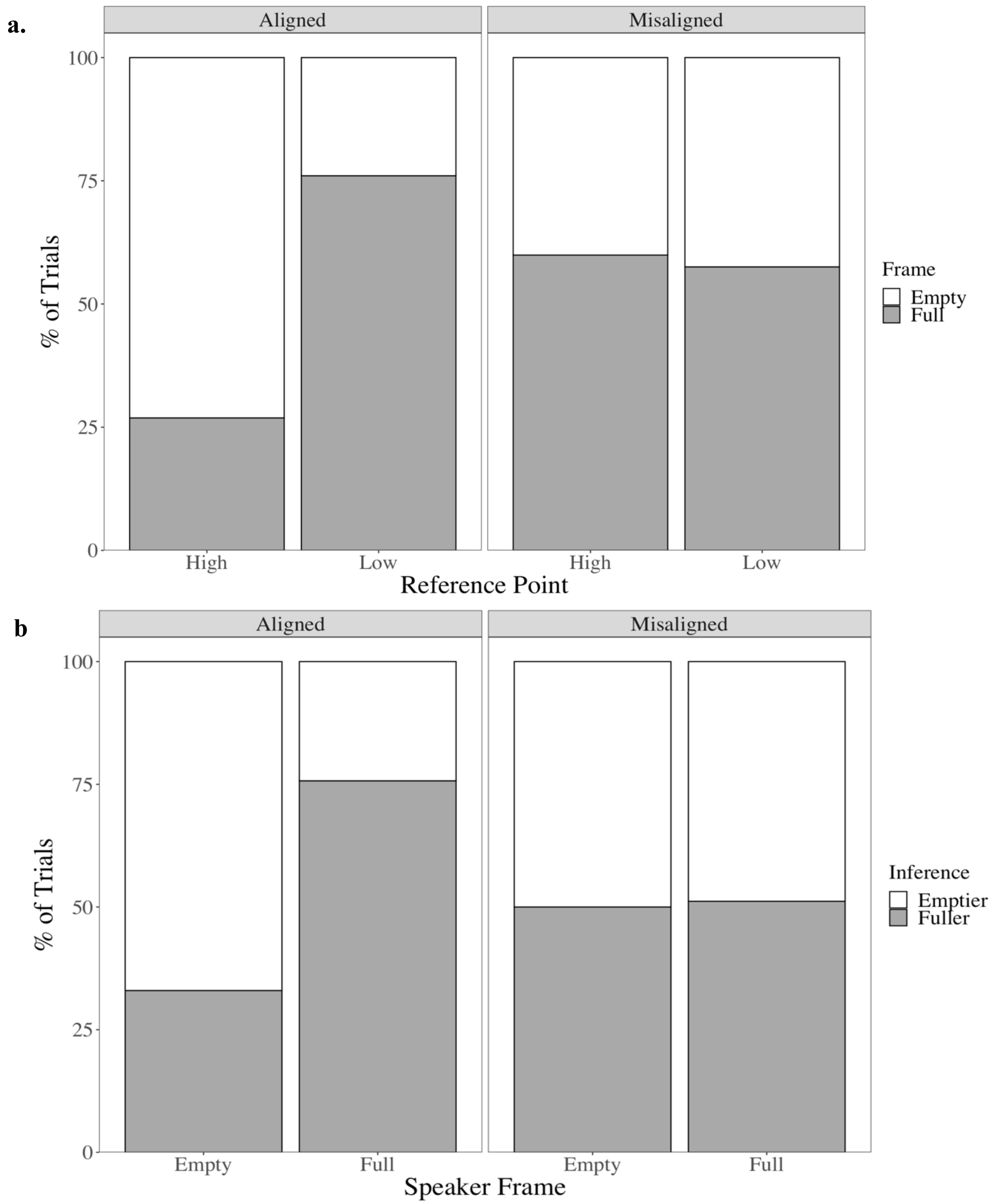

Figure 2.6. The percentage of trials that speakers chose the "full" frame as a function of reference point in each condition (a), and the percentage of trials that listeners guessed "fuller" as a function of speaker frame in each condition (b).

demonstrates that listeners can cease to respond systematically to frames when their interests are misaligned with the speakers' interests.

Finally, we analyzed the joint responses in speaker-listener pairs. Figure 7a shows how we classified the four possible combinations of speaker message and listener guess as "Information Leakage," "Opposite," "Wrong Frame," and "Wrong Inference." According to information leakage, speakers tend to use a particular label when that label's percentage exceeds a reference point, and listeners draw this corresponding inference from the frame. Thus, we classified the joint response of the speaker using the "full" ("empty") frame when the reference point is low (high) and the listener guessing "fuller" ("emptier") from the "full" ("empty") frame as "Information Leakage." This joint response, however, is not the only one that results in listeners making the correct guess. When speakers and listeners both respond opposite of information leakage (e.g., speaker selecting "full" frame when the reference point is high and listener guessing "emptier" from the "full" frame), we classified that joint response as "Opposite." For "Wrong Frame," speakers choose frames inconsistent with information leakage while listeners' guesses are consistent with information leakage, and for "Wrong Inference," speakers choose frames consistent with information leakage while listeners' guesses are inconsistent with information leakage. As shown in Figure 7b, the most common joint response is Information Leakage in the aligned interests condition while all possible joint responses were more evenly distributed in the misaligned interests condition.[2] Figures 7c and 7d further breaks

[2] While both Information Leakage and Opposite are both equally effective ways for speakers and listeners to coordinate their responses, we observed that they strongly preferred the former. To investigate whether this was due to how we phrase the inference question, we ran an additional study (N = 98, or 49 speaker-listener pairs) that reversed how we asked the question. This study is identical to Study 3, except we only ran the aligned interest condition and listeners were instead asked to complete the statement "The hidden cup is ___ than the target cup." When we reversed the inference question, we found that speakers and listeners did not prefer Opposite but were equally likely to coordinate on Information Leakage and Opposite (40% vs. 40%). When averaged across the two

down the percentage of trials that fit the four joint response profiles by individual speaker-listener pairs.

Our results in the aligned interests condition replicate previous findings that the speaker's choice of frame can convey reference point information, and that listeners are attuned to this regularity in drawing different inferences. More importantly, our results in the misaligned interests condition suggest that this behavior is adaptable: When incentivized not to communicate useful information, speakers no longer selected frames based on reference points, and listeners ceased to respond systematically to the frame.

---

ways of asking the inference question in these studies, speakers and listeners coordinated on Information Leakage 53% of the time and Opposite 26% of the time (and 10% for Wrong Frame and 11% for Wrong Inference).

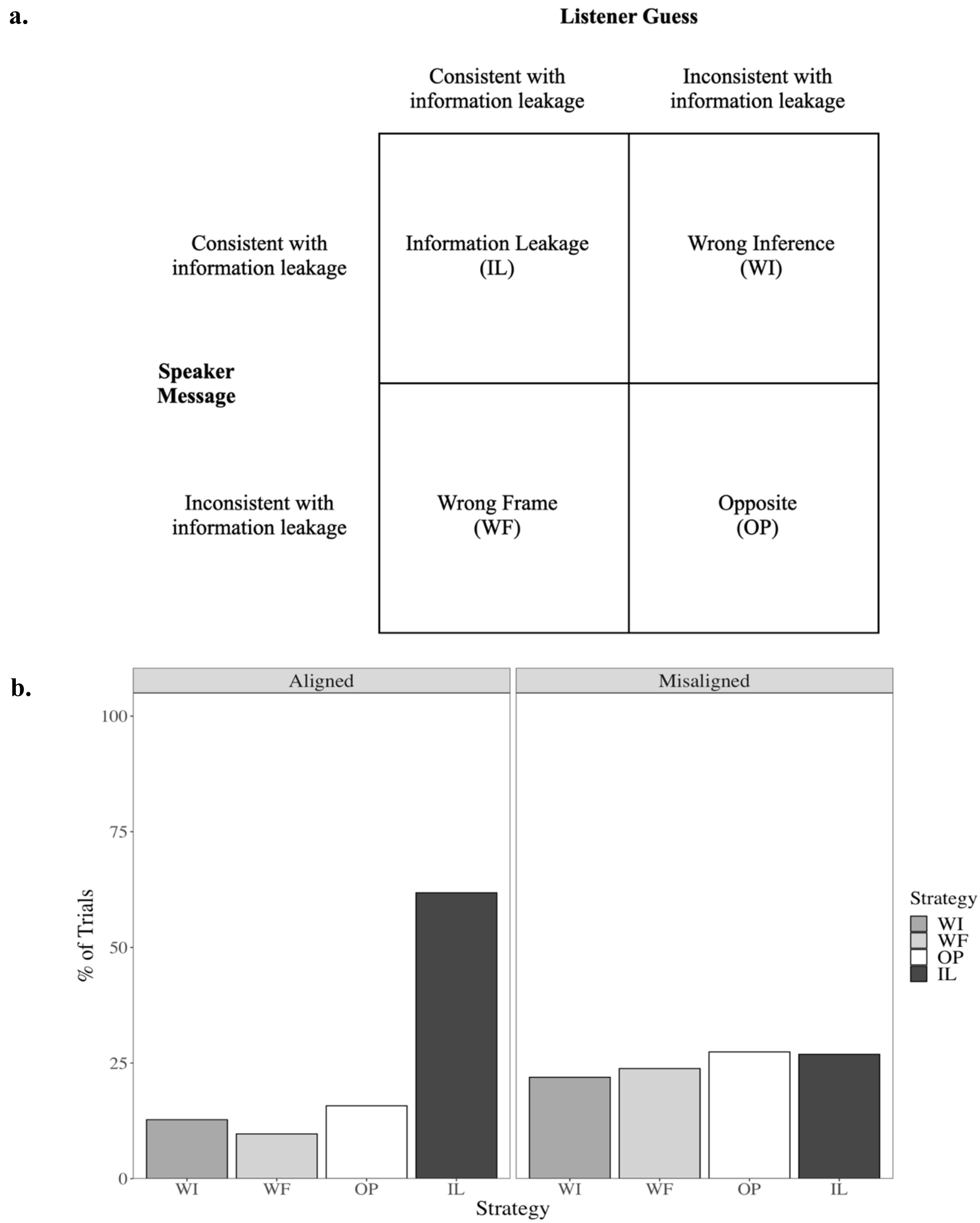

Figure 2.7. Our classification of the four possible combinations of speaker message and listener guess (a). The most common joint response is Information Leakage in the aligned interests condition while all possible joint responses were more evenly distributed in the misaligned interests condition (b). The distribution of joint responses for individual speaker-listener pair is shown for the aligned interests (c) and misaligned interests conditions (d).

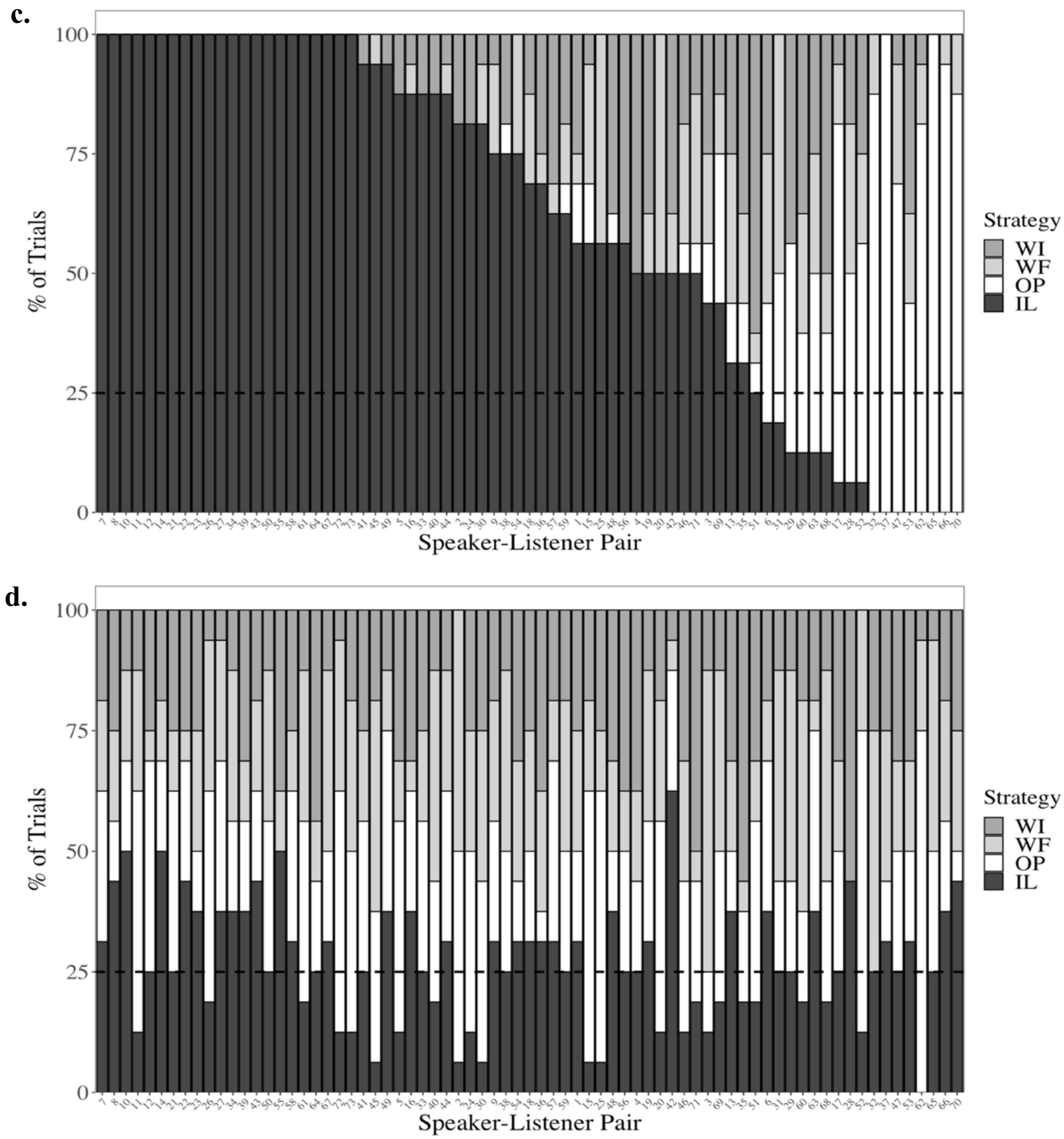

Figure 2.7. Classification of possible combinations of speaker message and listener guess, Continued.

General Discussion

According to information leakage, attribute framing effects arise from the different inferences that people draw from frames (Leong et al., 2017; McKenzie & Nelson, 2003; Sher & McKenzie, 2006). While this behavior is adaptive in the usual environment in which the speaker's choice of frame is informative, few studies have examined the adaptability of this behavior to environments in which the speaker's choice of frame is uninformative. In the present article, we report four studies that manipulated the informativeness of frames in two vastly different paradigms to test whether listeners can flexibly change their behavior.

In the first three studies, we presented participants with hypothetical vignettes that described the frame selection process. In Study 1, we informed participants that a computer has randomly selected a frame to use, which should render the choice of frame uninformative, yet we failed to find a reduced framing effect. In Studies 2a and 2b, we introduced an external constraint that made it more difficult for the speaker to use one frame over another. However, regardless of whether the speaker chose the frame that was made easier or more difficult to use by the constraint, we failed to find a reduced or enhanced framing effect. Finally, in Study 3, we manipulated the incentive structure of a communication game between speakers and listeners. Unlike our previous results, we found that when their interests were changed from being aligned to misaligned, speakers no longer selected frames in an informative manner and, more importantly, listeners ceased to respond systematically to frames.

Taken together, our results caution against a simple answer as to whether attribute framing effects are adaptable. Instead, the adaptability of framing effects seems to depend on how the change in task environment is implemented. When the uninformative frame selection was explicitly described to them, listeners did not modify their reactions to frames. Yet, when speakers are incentivized to be uncooperative in a communication game, listeners appear to recognize how that affects the informativeness of the speaker's frame and could respond appropriately. We suspect that people are unable to consciously combine their explicit knowledge about frame selection with the implicit inferential process that underly attribute framing. People are, however, sensitive to how speakers select frames when placed in a game that taps into their knowledge about cooperation and communication. One interesting question for future research is whether inferences based on other linguistic regularities such as active-passive form sentences (e.g., Johnson-Laird, 1968) and subject-complement structures (e.g., Chestnut & Markman, 2018) are adaptable, and whether their adaptability may similarly depend on how the change to the task environment is implemented.

Using a paradigm analogous to the one we used in Study 3, Altmann, Falk, & Grunewald (2020) examined the adaptability of a conceptually similar phenomenon known as default effects. The default refers to the option that is implemented in the absence of active choice, and default effects are said to occur when decision makers tend to stick with whichever option is set as the default (Jachimowicz, Duncan, Weber, & Johnson, 2019). Although defaults do not change the menu of options, the choice of default can signal an implicit recommendation (McKenzie, Liersch, & Finkelstein, 2006). For instance, participants were more likely to infer that policymakers are in favor of organ donation when they set "organ donor" (rather than "not an organ donor") as the default. In their default-setting game, Altmann et al. (2020) also manipulated the alignment of interests between default-setters and decision makers. When their interests were aligned, decision makers accepted the default 90% of the time, but when their interests were misaligned, they did so only 58% of the time. This provides convergent evidence that the subtle transmission of information through logically equivalent choice architecture is adaptable to different incentive structures in a simple repeated two-player game.

Our findings also provide additional evidence that, for at least some tasks, our cognitive system is malleable and our behaviors adaptable. Prior studies on reasoning and judgment tasks such as hypothesis testing (Klayman & Ha, 1987; McKenzie & Mikkelsen, 2000), the selection task (Oaksford & Chater, 1994), and covariation assessment (McKenzie & Mikkelsen, 2007) have demonstrated that people's behaviors can change in appropriate ways when it is clear the environmental structure has changed. To illustrate, consider the typical covariation task: Participants are asked to assess the relationship between two variables, each with levels of presence and absence. Traditionally, the view is that all four combinations of the resulting contingency matrix are equally informative, but empirically, participants' judgments are influenced most by joint presence observations and least by joint absence observations. This reasoning "bias," however, is adaptive from a Bayesian inferential perspective that assumes the presence of variables to be rare. Crucially, when it is clear that the absence of variables is rare, participants instead consider joint absence observations to be most informative (McKenzie & Mikkelsen, 2007). Furthermore, this effect was only found for variables that participants were familiar with. When presented with unfamiliar variables, participants relied on their default assumption about how labeling indicates what is rare. When presented with familiar variables, however, participants exploited their real-world knowledge about what is rare. In this way, people's sensitivity to rarity appears also to depend on how it is conveyed in the task.

When choice architecture is viewed as an implicit social interaction (Krijnen, Tannenbaum, & Fox, 2017; McKenzie, Sher, Leong, & Müller-Trede, 2018), factors that change the nature of that interaction should lead to corresponding changes to behavior. Although people's responses to frames are surprisingly insensitive to descriptions of the frame selection process, their behavior is adaptable in a communication game that manipulates the incentive for speakers to be cooperative. When listeners know that speakers are incentivized not to leak information, they no longer respond in a systematic manner to frames.

Chapter 2, in full, is under review for publication of the material. Leong, Lim M.; McKenzie, Craig R. M. The dissertation author was the primary investigator and author of this paper.

References

Altmann, S., Falk, A., & Grunewald, A. (2020). *Incentives and information as driving forces of default effects.* Unpublished manuscript.

Aczel, B., Szollosi, A., & Bago, B. (2018). The effect of transparency on framing effects in within-subject designs. *Journal of Behavioral Decision Making, 31*(1), 25–39.

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics, 12*(4), 428–454.

Chestnut, E. K., & Markman, E. M. (2018). "Girls are as good as boys at math" implies that boys are probably better: A study of expressions of gender equality. *Cognitive Science, 42*, 2229-2249.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3: Speech acts* (pp. 41–58). New York: Academic Press.

Hilton, D. J. (1995). The social context of reasoning: conversational inference and rational judgment. *Psychological Bulletin, 118*, 248–271.

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 1–28.

Johnson-Laird, P. N. (1968). The interpretation of the passive voice. *Quarterly Journal of Experimental Psychology, 20*, 69–73.

Keren, G. (2007). Framing, intentions, and trust–choice incompatibility. *Organizational Behavior and Human Decision Processes, 103*, 238–255.

Klayman, I., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Krijnen, J. M. T., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy, 3*(2), 1–18.

Leong, L. M., McKenzie, C. R. M., Sher, S., & Müller-Trede, J. (2017). The role of inference in attribute framing effects. *Journal of Behavioral Decision Making, 30*, 1147-1156.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*, 374-378.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior & Human Decision Processes, 76,* 149-188.

McKenzie, C. R. M. (2005). Judgment and decision making. In K. Lamberts & R. L. Goldstone (Eds.), *Handbook of cognition* (pp. 321–338). London: Sage Publications.

McKenzie, C. R. M., Liersch, M. K., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science, 17*, 414–420.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin and Review, 7*, 360–366.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology, 54*(1), 33–61.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review, 10*, 596–602.

McKenzie, C. R. M., Sher, S., Leong L. M., & Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics, 5*, 337–360.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review, 109*(2), 330–357.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608–631.

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General, 121*, 352–363.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*(2), 195–202.

Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition, 9*(1), 67-84.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53*, 491-517.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101*, 467–494.

Teigen, K. H., & Karevold, K. I. (2005). Looking back versus looking ahead: Framing of time and work at different stages of a project. *Journal of Behavioral Decision Making, 18*, 229–246.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59*, S251–278.

Chapter 3:

Exploiting asymmetric signals from choices through default selection

Lim M. Leong, Yidan, Yin, & Craig R. M. McKenzie

As it appears in

Psychonomic Bulletin & Review

2020

Volume 27, Issue 1, Pages 162-169

Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-019-01699-2

Exploiting asymmetric signals from choices through default selection

Lim M. Leong[1] · Yidan Yin[2] · Craig R. M. McKenzie[1,2]

Abstract
Setting defaults is an effective nudge, but few studies have examined situations where individuals can select their own default settings. Past research suggests that even when the final outcome is identical, observers perceive stronger signals from choices that switch from, rather than stay with, the default. In five experiments using hypothetical scenarios and an incentivized economic game, we test whether decision-makers driven by image concerns could strategically exploit that asymmetric signal. We found that in the presence of observers, participants were more likely to self-select into defaults that require them to switch to enhance a positive signal and into defaults that require them to stay to attenuate a negative signal. Our results support the framework of choice architecture as an implicit social interaction, and have potential implications for behavioral interventions in real-world settings.

Keywords Choice architecture · Nudges · Default options · Behavioral signaling

Setting default options is a potent and popular behavioral intervention for choice architects (Benartzi et al., 2017; Thaler & Sunstein, 2008). Designating the option you want others to choose as the default—namely, the one that is implemented when decision-makers (DMs) fail to make an active choice—can increase the chances that it is chosen. For example, organ donation consent rates are higher in countries where residents are presumed to be donors by default than countries where they are not (Johnson & Goldstein, 2003). Because defaults do not alter the menu of options, they allow choice architects to nudge DMs toward desired outcomes while preserving their autonomy to choose.

The existing research on default effects has focused almost exclusively on how choices are affected by different randomly assigned defaults. What if DMs can instead choose which default setting to interact with? For instance, imagine that different restaurants have set different side dishes as the default, and you are choosing which restaurant to take your friend out for dinner. You want to impress your friend by showing that you eat healthily so you would obviously order salad over fries. But would you prefer to order that salad at a restaurant where salad is the default or at one where fries is the default? As we explain below, we suspect that many would prefer the latter restaurant, where you would actively order the salad rather than passively receive it. In the experimental paradigm of the present research, we allow DMs to choose between different default settings to obtain the same options.

A key factor that may drive DMs to self-select into different default settings is social image concerns. Although defaults do not change the availability of options, they can affect the social meaning attached to those options (Krijnen, Tannenbaum, & Fox, 2017). That is, the same option can send different signals depending on whether it was obtained by staying with, or switching from, the default. To illustrate, participants in one experiment were asked to rate how altruistic it would be to donate organs under different default policies (Davidai, Gilovich, & Ross, 2012). Participants who read that the country had a "not an organ donor" default judged organ donation as comparable to bequeathing half of one's wealth to charity, but those who read that the country had an "organ donor" default judged organ donation to be akin to letting others go ahead in line. Similarly, participants playing the role of a surviving family member inferred that a decedent on the organ donor register has a stronger underlying preference to donate when the choice was made under an opt-in system than opt-

Preregistration documents, materials, and data can be found at the Open Science Framework page for this article (https://osf.io/d94xn/).

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13423-019-01699-2) contains supplementary material, which is available to authorized users.

✉ Lim M. Leong
lmleong@ucsd.edu

1 Department of Psychology, UC San Diego, La Jolla, CA, USA
2 Rady School of Management, UC San Diego, La Jolla, CA, USA

Published online: 02 January 2020

out (Lin, Osman, Harris, & Read, 2018). Organ donation was perceived as much more meaningful and impactful, and as a stronger indicator of one's true preference, when it required switching from the default.

Given that observers perceive asymmetric signals from choices under different default regimes, we examine whether DMs can strategically exploit this perception. Our main hypothesis is that DMs motivated to send a stronger positive signal will self-select into the default regime that requires them to switch from the default option. Recent findings provide conflicting evidence on whether DMs are capable of such strategic behavior. Zlatev, Daniels, Kim, and Neale (2017) claim that in attempts at social influence, DMs often exhibit "default neglect," or the inability to set optimal defaults. Recent conceptual replications, however, suggest that this poor performance is limited to the original materials and that DMs are instead very good at setting defaults (Jung, Sun, & Nelson, 2018; McKenzie, Leong, & Sher, 2019). Our research question requires even more complicated reasoning about defaults, and if participants indeed do not understand how defaults influence others' choices, then they should similarly fail to self-select into optimal default regimes.

Experiment 1

We first attempted to conceptually replicate the finding that observers perceive a stronger signal from switching than staying with the default in a healthy eating scenario. To this end, we placed participants in the role of observers and asked them to draw inferences about two friends: One who ordered salad at a restaurant where salad is the default, and the other who ordered salad where fries is the default. We predicted that participants would infer that the friend who ordered salad at the fries default restaurant cared more about healthy eating.

Method

We used a rule of thumb and decided in advance to set the target sample size to be 100 participants per between-subjects condition. A total of 225 undergraduates in the UCSD Psychology Department (M_{age} = 19.56 years; 77% female) participated for partial course credit. After participants completed one of our experiments, they were barred from signing up for all subsequent ones. Participants read two scenarios about going out to dinner with a friend, and in each scenario, the friend ended up choosing a salad. In one scenario, the restaurant had salad as the default side dish, but customers could switch to get fries. In the other scenario, the restaurant had fries as the default side dish, but customers could switch to get salad. Thus, the friend in one scenario chose salad by staying with the default, and the friend in the other scenario chose salad by switching from the default. The order of presentation for the two scenarios was manipulated between subjects. For each scenario, participants rated how much they thought that their friend "truly cares about eating healthy" on a scale from 1 (*not at all*) to 9 (*very much*). The participants' rating to this question for the first scenario was displayed to them when they made their ratings for the second scenario. For all experiments, we have reported all conditions, data exclusions, and our main independent and dependent measures.[1]

Results

Figure 1 shows the mean ratings of healthy eating for the friend who chose salad under different defaults and presentation orders. A mixed ANOVA was performed on the ratings, with order as a between-subjects factor and default as a within-subjects factor. This analysis revealed a main effect of default, $F(1, 223) = 159.89$, $p < .001$, $\eta_g^2 = .182$, as well as a significant interaction, $F(1, 223) = 26.36$, $p < .001$, $\eta_g^2 = .035$. Participants rated the friend who ordered salad at the fries default restaurant, $M = 6.92$, $SD = 1.70$, as caring more about healthy eating than the friend who ordered salad at the salad default restaurant, $M = 5.31$, $SD = 1.85$. The magnitude of this difference, however, depended on the presentation order. Additional paired-sample t tests showed that the difference in ratings was larger when participants read the salad default scenario first, $Ms = 7.58$ vs. 5.31, $M_{diff} = 2.27$, $SD_{diff} = 1.89$, $t(112) = 12.74$, $p < .001$, $d = 1.20$, 95% CI [1.91, 2.62], compared with when they read the fries default scenario first, $Ms = 6.27$ vs. 5.31, $M_{diff} = 0.96$, $SD_{diff} = 1.94$, $t(111) = 5.22$, $p < .001$, $d = 0.49$, 95% CI [0.59, 1.32]. To examine whether this pattern holds in a between-subjects design, we compared the participants' ratings of the first scenario they were presented with. An independent t test showed a higher mean rating of healthy eating for the friend who ordered salad at the fries default restaurant compared with the one who did so at the salad default restaurant, $M_{fries} = 6.27$, $SD_{fries} = 1.54$, $M_{salad} = 5.31$, $SD_{salad} = 1.85$, $t(216.3) = 4.23$, $p < .001$, $d = 0.56$, 95% CI [0.51, 1.41]. Together, these results confirm our prediction that ordering salad by switching from the fries default, rather than staying with the salad default, was perceived as a stronger signal of being a healthy eater.

Experiments 2a–b

The previous experiment found that observers perceived switching from the fries default to order salad as a stronger signal of being a healthy eater. In the next two experiments, we placed participants in the opposite role: DMs could choose

[1] See Supplementary Material for our additional dependent variable and its analysis.

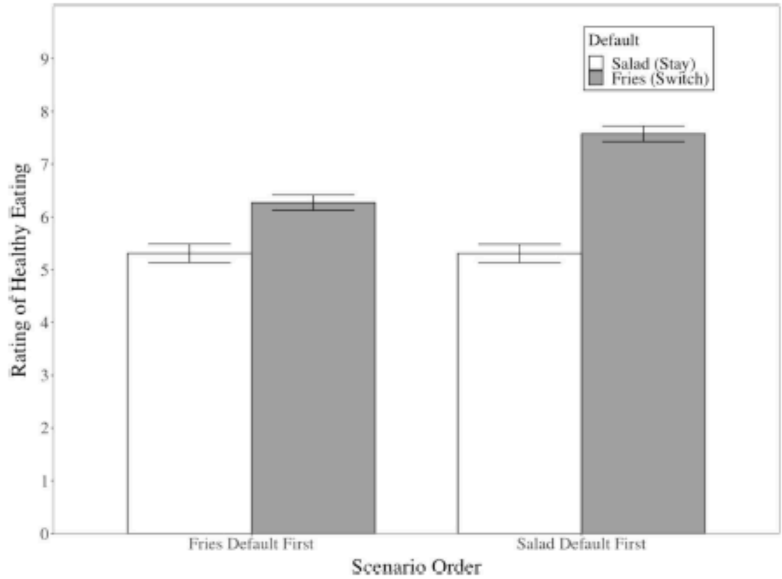

Fig. 1 Mean ratings of healthy eating as a function of side dish default and scenario presentation order in Experiment 1. Error bars represent standard error of the mean

which restaurant to take their friend out for dinner. Experiment 2a provided an initial test, while Experiment 2b replicated and extended it by manipulating the visibility of choice to test whether choices are driven by behavioral signaling. We predicted that DMs would self-select into the fries default restaurant to order salad only in the presence of an observer.

Method

We decided in advance to set our target sample size to be 100 participants per between-subjects condition. A total of 136 undergraduates participated in Experiment 2a (M_{age} = 19.88; 70% female), and a total of 202 undergraduates participated in Experiment 2b (M_{age} = 20.00; 63% female, 1 reported "Other"). All participants received partial credit in psychology courses.

Experiment 2a had only one condition. Participants imagined going out to dinner with a friend and could choose one of two restaurants to order a salad. The restaurants were said to be comparable in every way and offered the same choice set, except fries come on the side by default at one and salad comes on the side by default at the other. Furthermore, the friend has never heard of these restaurants, and would only know about the one that participants decide on. Finally, participants were told that they want to impress their friend by showing that they eat healthily and were asked to indicate which restaurant they would choose to take their friend.

In Experiment 2b, participants were randomly assigned to either the "public" or "private" condition. The "public" condition is a direct replication of Experiment 2a, and participants again imagined that they were going out to dinner with a friend whom they want to impress. In the "private" condition, participants chose between the same two restaurants, but imagined that they were going out to dinner alone and that they were conscious about their health.

Results

Figure 2 shows the percentage of participants who chose the restaurant with the fries default for Experiments 2a and 2b. When dining with a friend, 71.3% (97/136) of participants in Experiment 2a chose the fries default restaurant, which was significantly higher than chance, $\chi^2(1, N = 136) = 23.89, p < .001$, 95% CI [62.8%, 78.6%]. In Experiment 2b, 76% (77/101) of participants in the "public" condition similarly chose the fries default restaurant, but only 12% (12/101) of participants in the "private" condition chose the fries default restaurant, $\chi^2(1, N = 202) = 82.27, p < .001, \varphi = .64$. These results suggest that motivation to signal is a key factor that drives selection into different default regimes. When there was an opportunity to signal, participants strategically chose the default setting with the intention of switching away from the default. By doing so, their combination of default and choice matched the one that observers perceived as a stronger signal of being a healthy eater. When there was no opportunity to

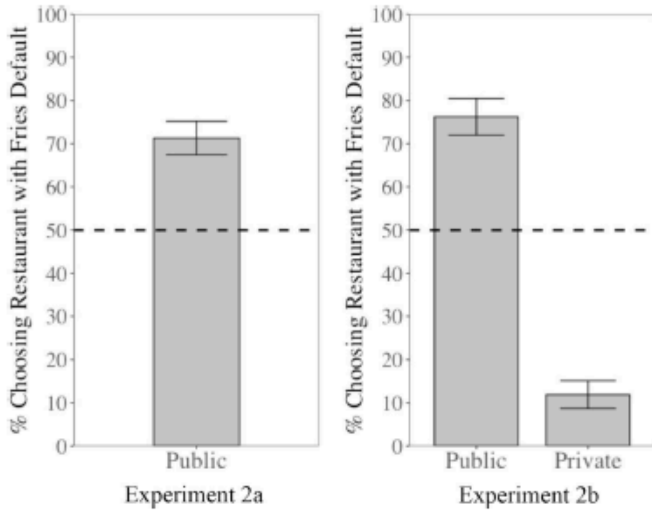

Fig. 2 The percentage of participants who chose the restaurant with the fries default to order a salad in Experiments 2a and 2b. Experiment 2a only has the public condition, while Experiment 2b manipulated the visibility of choice by either having dinner with a friend or alone. Error bars represent standard error of the proportion

signal, however, participants simply chose the default setting that matched their choice.

Experiment 3

Thus far, we have tested our research question using only hypothetical scenarios. In our next experiment, we adapted the paradigm of Pleasant and Barclay (2018) and used incentivized economics games through which participants compete to be chosen by observers for a subsequent task. Specifically, participants played a modified Dictator Game and could strategically self-select into a fair split default or a selfish split default. As before, we manipulated the visibility of choice to test whether signaling drives their self-selection. We predicted that participants would select into the selfish split default with the intention of switching to the fair split to send a stronger positive signal to observers.

Method

We anticipated that this task would be more difficult for participants to understand, and decided in advance to increase our target sample size to 250 participants per condition. A total of 506 undergraduates at UCSD's Rady School of Management (M_{age} = 20.84 years, one did not report age; 52% female, one reported "Other") participated for partial course credit. All participants played a modified Dictator Game as the dictator and were told that one randomly selected participant would have her choices played out for real money at the end of the experiment. In the modified Dictator Game, participants were endowed with $10 and chose between a selfish split ($10, $0) and a fair split ($5, $5). In addition, participants chose between playing two versions of the game: In one version, the selfish split is the default, and in the other version, the fair split is the default. Participants watched four short clips that depict the four possible combinations of default and split, and selected one clip to represent their choices (see here for the four clips: https://osf.io/7qpm5/).

Participants were randomly assigned to one of two conditions. In the "public" condition, participants were told that their choices in the Dictator Game would be shown to a separate group of third-party observers, and these observers would determine whether participants have the chance to play a subsequent Trust Game.[2] Thus, through their choices in the Dictator Game, participants competed against each other to be selected by observers for the opportunity to earn additional money. The "private" condition serves as the control condition, and is identical to the "public" condition, except there were no observers and no opportunity to earn additional money in a subsequent Trust Game.

[2] In the Trust Game, Player 1 (the third-party observer) is endowed with $10 and can send any of that amount to Player 2 (her chosen partner). Any amount that Player 1 sends to her partner is tripled. Player 2 can then return any of the amount she received back to Player 1. At the end of the experiment, one randomly selected pair played the Trust Game for real money.

Results

We first examined the participants' choice of split. While 45.1% (114/253) of participants in the private condition chose the fair split, 68.4% (173/253) of participants in the public condition did so, $\chi^2(1, N = 506) = 27.08, p < .001, \varphi = .23$. This suggests that participants believed that the fair split sends a more positive signal to observers than the selfish split. More importantly, we predicted that to amplify this positive signal, participants would pair this fair split with the selfish split default. Indeed, of those who chose the fair split, 67.1% (116/173) of participants in the public condition chose the selfish split default, compared with only 52.6% (60/114) in the private condition, $\chi^2(1, N = 287) = 5.43, p = .02, \varphi = .14$ (see Fig. 3). Moreover, of those who chose the fair split, the percentage of participants who chose the selfish split default is significantly higher than chance (i.e., 50%) in the public condition, $\chi^2(1, N = 173) = 19.45, p < .001$, 95% CI [59.4%, 73.9%], but not in the private condition, $\chi^2(1, N = 114) = 0.22, p = .64$, 95% CI [43.1%, 62.0%].

We also explored whether those who chose the selfish split would be more likely to choose the selfish split default to avoid sending a stronger negative signal. On the contrary, of those who chose the selfish split, 70.0% (56/80) of participants in the public condition chose the selfish split default, compared with 84.9% (118/139) in the private condition, $\chi^2(1, N = 219) = 6.02, p = .014, \varphi = .17$. We speculate that these participants may not have cared about their image, and therefore did not have the goal to avoid sending a stronger negative signal.

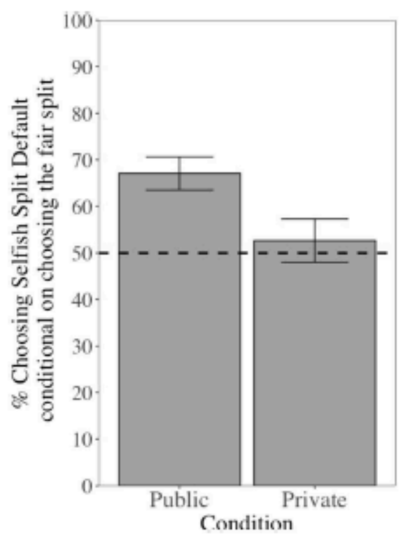

Fig. 3 Of those who chose the fair split, the percentage of participants in the private and public conditions who chose the selfish split default (rather than the fair split default) in Experiment 3. Error bars represent standard error of the proportion

To ensure that the fair split paired with the selfish split default in fact sends a stronger positive signal, we ran another study that placed participants in the role of third-party observers ($N = 157$). When presented with the same four clips that depicted the possible combinations of default and split, 60.5% (95/157) of participants selected the dictator who implemented the fair split under the selfish split default as their partner in the Trust Game, $\chi^2(1, N = 157) = 103.7, p < .001$, 95% CI [52.4%, 68.1%] (against chance of 25%). Together, these results provide further evidence that participants can anticipate how their choices will be perceived and can strategically select into default regimes with the intention of switching from the default to send a stronger signal.

Experiment 4

So far, we have focused on choices that convey positive signals, but choices can also convey negative signals (e.g., Young, Monin, & Owens, 2009). When choices communicate negative signals, DMs should avoid sending stronger negative signals by choosing the default regime where they can simply stay with the default. In our final experiment, we further explored this hypothesis by manipulating both the signal and the visibility of choice. In a hypothetical grocery bag scenario, participants imagined shopping while using either their own reusable bags or store-provided plastic bags and either with a friend or alone. We predicted that participants would be more likely to choose a store where plastic bags were the default when shopping with a friend than when shopping alone, both to enhance a positive signal when using reusable bags (by switching from the store's default) and to attenuate a negative signal when using plastic bags (by staying with the store's default).

Method

Using a rule of thumb, we decided in advance to set our target sample size to be 150 participants per condition. A total of 603 Amazon Mechanical Turk workers participated in exchange for $0.20, and were randomly assigned to one of four conditions. After excluding 35 participants who failed the attention check, we were left with a final sample of 568 participants (M_{age} = 36.30 years; 48% female, one reported "Other"). Participants imagined that they were going grocery shopping and could choose one of two stores. The two stores were said to be comparable in every way, except customers are provided with plastic bags by default at one store and are asked to make an active choice between plastic and reusable bags at the other (no default). In the "positive signal" conditions, participants were told that they are using their own reusable bags, whereas in the "negative signal" conditions they were told that they are using store-provided plastic bags. In addition, participants in the "public" conditions are shopping with a friend who is environmentally conscious, and this friend would only know about the store where they shop. Participants in the "private" conditions are instead shopping alone.

Results

Figure 4 shows the percentage of participants who chose the store with the plastic bag default for each of the four conditions. The left bars correspond to staying with the default to avoid sending a stronger negative signal, and the right bars correspond to switching from the default to send a stronger positive signal. We performed a general linear regression with signal and visibility as the independent variables, and the participants' binary choice of store as the dependent variable (1 = plastic bag default, 0 = no default). As predicted, participants were more likely to choose the plastic bag default store in the public condition than in the private condition, 52.7% (146/277) versus 39.5% (115/291), $\chi^2(1) = 4.18$, $p = .041$. Furthermore, participants were more likely to choose the plastic bag default store in the negative signal condition than in the positive signal condition, 67.8% (192/283) versus 24.2% (69/285), $\chi^2(1) = 63.37$, $p < .001$. That is, they were more likely to choose the plastic bag default setting when it matched their use of plastic bags. More importantly, the signal-by-visibility interaction was not significant, $\chi^2(1) = 0.39$, $p = .53$. In other words, we did not find evidence that the effect of visibility depends on the valence of the signal. These results suggest that those who care about sending a positive signal and those who care about avoiding a negative signal both prefer the same default setting, the former to enhance their signal by switching and the latter to attenuate their signal by staying.

General discussion

Unlike past studies that randomly assigned participants to different defaults, our paradigm allowed participants to choose their own default setting. Using healthy eating (Experiments 2a and 2b) and grocery bag (Experiment 4) scenarios as well as an incentivized economic game (Experiment 3), we found that participants were more likely to self-select into different default settings in the presence of observers. Specifically, they were more likely to choose default settings that require them to switch when sending a positive signal, and to choose default settings that require them to stay when sending a negative signal. Furthermore, the combination of default and choice that participants selected matches the one that observers actually perceive as indicating a stronger positive signal (Experiments 1 and 3). Taken together, these results suggest that DMs understand how their choices under different defaults are construed by others and can choose the optimal combination to enhance a positive signal or attenuate a negative signal.

Our results challenge the claims of default neglect by Zlatev et al. (2017). Rather than failing to exploit the influence that defaults have on others' choices, DMs appear to have a sophisticated understanding of what their choices signal under different defaults and can use this to their advantage in impression management. This provides convergent evidence that

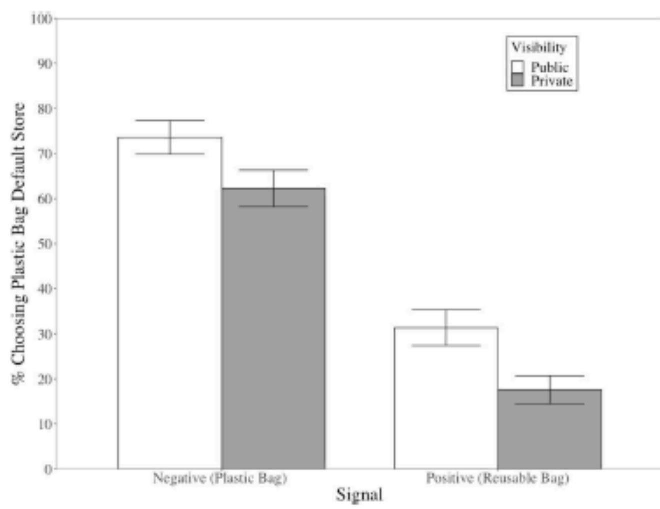

Fig. 4 The percentage of participants who chose the store with the plastic bag default as a function of the valence and visibility of choice in Experiment 4. Error bars represent standard error of proportion

default neglect may not be a general phenomenon, but rather is restricted to Zlatev et al.'s original materials (see also McKenzie et al., 2019).

Although choice architecture is conventionally discussed as a way to leverage cognitive biases for good, our findings support the framework of choice architecture as an implicit social interaction (Krijnen et al., 2017; McKenzie, Sher, Leong, & Müller-Trede, 2018; Sher & McKenzie, 2006). In this framework, choice architecture can influence DMs in two ways: (1) by changing the signal that their choice conveys and (2) through the inferences that they draw from the decision context. First, the same choice that requires switching rather than staying with the default is perceived by observers as more meaningful and impactful (Davidai et al., 2012). Our experiments complement this work by reversing the roles and allowing DMs to select the default setting. When given this opportunity, DMs were able to strategically exploit the asymmetric signal to influence others' perceptions. Second, previous studies have demonstrated that default effects occur in part because DMs perceive defaults as implicit recommendations (McKenzie, Liersch, & Finkelstein, 2006). The choice of default can "leak" the default setter's personal attitudes and intentions, and DMs are sensitive to this cue. Thus, DMs appear to consider the social aspect of the choice context, both in terms of what their choices in that context may reveal to observers and the information that the context may implicitly convey.

Our findings have potentially important implications. Much evidence has shown that an effective way to promote a certain option is to set that option as the default. Our results caution that such defaults may sometimes backfire by driving away DMs who want to choose that option, but whose positive signal is diluted by staying with the default. These DMs will prefer default settings where they can switch from the default to send a stronger signal. But before generalizing our findings to real-world settings, a few limitations are in order. First, in our experiments, participants were explicitly given the goal to signal, and we do not know how often DMs have a strong motivation to signal outside the laboratory. Second, the DM's choice between different default settings could itself send a signal if observers were aware of the choice. In our experiments, participants were explicitly told that the observer would not know about their choice between different default settings, but only the one they end up choosing. The conditions under which these assumptions are met and whether this backfiring would occur in typical real-world default settings can be explored in future field studies.

Future research can also investigate the mechanisms for why switching from the default is perceived as a stronger signal than staying. Promising starting points are the theoretical explanations for why default effects occur: loss aversion, effort, and implicit recommendation (Jachimowicz, Duncan, Weber, & Johnson, 2019). If observers believe that people tend to stick with defaults because doing so is less effortful, then switching could indicate a stronger preference because it is costly. Or if observers believe that people tend to stick with defaults because it is the recommended option, then switching could indicate that DMs have good reasons to reject this recommendation. Another interesting question that arises then is whether the factor that drives observer perception matches the one that DMs consider when choosing.

In sum, people care about what their choices communicate to others. Under different default settings, choices that lead to identical outcomes may nevertheless convey different social meanings. When people can select their choice environments, they do so strategically to exploit the asymmetric signals that observers perceive from choices under different defaults.

References

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., … Galing, S. (2017). Should governments invest more in nudging? *Psychological Science, 28*, 1041–1055. doi:https://doi.org/10.1177/0956797617702501

Davidai, S., Gilovich, T., & Ross, L. D. (2012). The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 15201–15205. doi:https://doi.org/10.1073/pnas.1211695109

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 1–28. doi:https://doi.org/10.1017/bpp.2018.43

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science, 302*, 1338–1339. doi:https://doi.org/10.1126/science.1091721

Jung, M. H., Sun, C., & Nelson, L. D. (2018). People can recognize, learn, and apply default effects in social influence. *Proceedings of the National Academy of Sciences of the United States of America, 115*(35), E8105–E8106. doi:https://doi.org/10.1073/pnas.1810986115

Krijnen, J. M. T., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy, 3*(2), 1–18. doi:https://doi.org/10.1353/bsp.2017.0010

Lin, Y., Osman, M., Harris, A. J., & Read, D. (2018). Underlying wishes and nudged choices. *Journal of Experimental Psychology: Applied, 24*(4), 459–475. doi:https://doi.org/10.1037/xap0000183

McKenzie, C. R. M., Liersch, M. K., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science, 17*, 414–420. doi:https://doi.org/10.1111/j.1467-9280.2006.01721.x

McKenzie, C. R. M., Leong, L. M., & Sher, S. (2019). *Default sensitivity in attempts at social influence*. Manuscript submitted for publication.

McKenzie, C. R. M., Sher, S., Leong, L. M., & Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics, 5*, 337–360. doi:https://doi.org/10.1561/105.00000091

Pleasant, A., & Barclay, P. (2018). Why hate the good guy? Antisocial punishment of high cooperators is greater when people compete to be chosen. *Psychological Science, 29*(6), 868–876. doi:https://doi.org/10.1177/0956797617752642

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101*, 467–494. doi:https://doi.org/10.1016/j.cognition.2005.11.001

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. London, England: Penguin Books.

Young, S. D., Monin, B., & Owens, D. (2009). Opt-out testing for stigmatized diseases: A social psychological approach to understanding the potential effect of recommendations for routine HIV testing. *Health Psychology, 28*(6), 675–681. doi:https://doi.org/10.1037/a0016395

Zlatev, J. J., Daniels, D. P., Kim, H., & Neale, M. A. (2017). Default neglect in attempts at social influence. *Proceedings of the National Academy of Sciences of the United States of America, 114*(52), 13643–13648. doi:https://doi.org/10.1073/pnas.1712757114

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 3, in full, is a reprint of the material as it appears in Psychonomic Bulletin & Review, 2020, Leong, Lim M.; Yin, Yidan; McKenzie, Craig R. M. The dissertation author was the primary investigator and author of this paper.

CONCLUSION

In three chapters, this dissertation illustrates the important role that inferences play in how people respond to two common choice architectures. Different frames and different defaults, despite their logical equivalence, can convey additional choice-relevant information to decision makers and observers. Building on the information leakage framework, we demonstrate that reference point inferences are sufficient to generate attribute framing effects (Chapter 1), and that these inferences are adaptable to changes to the incentive structure between speakers and listeners (Chapter 2). In addition, we found that decision makers have a sophisticated understanding of how observers would perceive their choices under different defaults, and can strategically exploit this signal by self-selecting into optimal default settings (Chapter 3).

The conventional nudge approach has advocated a variety of different behavioral interventions to improve people's welfare. But before widely implementing these interventions in the real world, we should first build a more comprehensive understanding of the underlying mechanisms for why people behave the way they do. If the target behaviors are not the result of faulty cognitive processes as commonly believed, then these interventions may prove ineffective, or worse yet, backfire. That is, whether people are in fact predictably irrational – or instead, *predictably rational* – matters when designing interventions. We believe our perspective of choice architecture as an implicit social interaction offers key insights on factors that choice architects should consider.

One factor is how much prior knowledge decision makers have. While reference point inferences are sufficient to generate attribute framing effects (Chapter 1), we have also identified boundary conditions for when these inferences matter less. The less relevant knowledge decision makers have, the more they should look to and rely on subtle cues in their choice environment.

Conversely, the more relevant knowledge decision makers have, the less they should rely on these inferences. Indeed, we found that those who are more knowledgeable about NBA basketball, and thus need not rely on the frame to interpret how valuable a player is, show a diminished framing effect (Experiment 2 in Chapter 1). Likewise, the implicit recommendation in defaults should exert a smaller influence on those who are knowledgeable about the options.

Another factor that should affect framing and default effects is speaker intentions. In the typical cooperative context, speakers are expected to make informative utterances. When this assumption is violated, decision makers should cease to respond systematically to these cues. Contrary to this prediction, we found that people continued to exhibit framing effects despite being informed that a computer has randomly selected the frame (Study 1 in Chapter 2), and when a constraint made it harder for the speaker to use one frame over another (Study 2a and 2b in Chapter 2). However, when the speakers' and listeners' interests were misaligned in a communication game, we found that listeners no longer reacted systematically to frames. These results suggest that framing effects may be more adaptable to certain manipulations than others, and future research can investigate other speaker characteristics such as trust and conflict-of-interest. For instance, future studies can manipulate the speaker's role and compare how people respond to frames selected by marketers and politicians versus those selected by friends. Another interesting avenue for future research is whether a design that presents frames and defaults via in-person communication would lead to larger effects than using hypothetical vignettes.

In summary, this dissertation has further elucidated the role that inferences play in the judgments and choices made under different framing and default options. Judgment and decision making do not occur in a vacuum, but are instead situated within a social setting. Because information cannot simply be implanted into the minds of decision makers, *how* that information is communicated can itself be informative. By examining the implicit interaction between choice architects, decision makers, and observers, we can better understand why people behave the way they do and design more effective behavioral interventions.