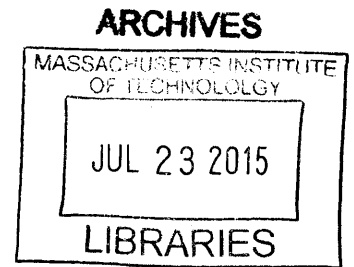


The Effect of Message Framing on Initial Choices, Satisfaction, and Ongoing Engagement

by

Eleanor Putnam-Farr

B.A Economics, Williams College, 2000
M.B.A, Massachusetts Institute of Technology, 2010



Submitted to the Alfred P. Sloan School of Management
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
at the
Massachusetts Institute of Technology

June 2015

© 2015 Massachusetts Institute of Technology.
All rights reserved.

Signature redacted

Signature of Author: _____
Alfred P. Sloan School of Management
April 23, 2015

Signature redacted

Certified by: _____
John R. Hauser
Kirin Professor of Marketing
Thesis Supervisor

Signature redacted

Accepted by: _____
Ezra Zuckerman Sivan
NTU Professor of Strategic Management
Chair, MIT Sloan PhD Program

Table of Contents

ABSTRACT.....	5
ACKNOWLEDGMENTS.....	7
ESSAY 1	8
"Yes/No/Not right now." Yes/No Response Formats Can Increase Response Rates Even in Non-forced-choice Settings	
ESSAY 2	37
The Effect of Benefit Quantification on Goal Setting and Persistence	
ESSAY 3	67
A Picture is Worth a Thousand Words: Photographs as Anchors for Expectations	

The Effect of Message Framing on Initial Choices, Satisfaction, and Ongoing Engagement

by

Eleanor Putnam-Farr

Submitted to the Alfred P. Sloan School of Management
on April 23, 2015 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy
at the Massachusetts Institute of Technology

ABSTRACT

Essay 1: "Yes/No/Not right now." Yes/No response formats can increase response rates even in non-forced-choice settings

Although yes/no response formats have been used to increase enrollment rates in several different types of programs, their use has been largely limited to forced choice settings. Across two field experiments, we demonstrate a substantial advantage in click-through rates for a yes/no response format over traditional opt-in response formats in an email context where choice is not forced. The increase in click-through rate does, under certain conditions, also persist through downstream program enrollment and participation. Finally, though noting that the yes/no format advantage is probably multi-determined, we discuss several potential psychological mechanisms, which are particularly relevant in non-forced choice settings.

Essay 2: The Effect of Benefit Quantification on Goal Setting and Persistence

We look at how language used to describe rewards, specifically the quantification of an expected reward, might lead participants to create specific targets for their own performance based on that language. Through a combination of field and lab experiments, we demonstrate that the use of a high number to describe rewards leads to higher interest and enrollment, but also higher expectations of performance and a higher drop-out rate from the program when the reward is difficult to achieve. Marketers should be aware of this issue when describing benefits to potential customers, particularly if they wish to motivate persistent behavior.

Essay 3: A Picture is Worth a Thousand Words: Photographs as Anchors for Expectations

Marketers often use images to increase the vividness of their communications to customers. These images should make the message easier to process, thereby increasing liking for the product and certainty of expectations. In a series of experiments, I demonstrate that images do indeed increase certainty of preference estimates, both within and between respondents, but may have more mixed effects on the valence of estimates of preference. I also begin to examine how these estimates might impact evaluations of actual products and propose some areas for future exploration.

Thesis Committee:

John R. Hauser (Chair)
Kirin Professor of Marketing

Renée Richardson Gosline
Zenon Zannetos (1955) Career Development Professor
Assistant Professor of Marketing

Drazen Prelec
Digital Equipment Corp. LFM Professor of Management Science
Professor of Economics, Professor of Brain and Cognitive Sciences

Jason Riis
Visiting Assistant Professor of Marketing
The Wharton School, University of Pennsylvania

ACKNOWLEDGMENTS

When I arrived at MIT as an MBA student, I had no idea I would be staying for seven years and finding an entirely new career path. I want to start by thanking Renée Richardson Gosline for introducing me to the world of academic research and taking a chance on an MBA student who wanted to learn to ask and answer interesting questions. John Hauser has adopted me as his student and taught me many lessons, as a researcher, as a member of an academic community, and as a wonderful mentor. Drazen Prelec has been generous with his time and advice and I have been fortunate to receive his feedback.

My research would not be possible without the substantial contributions of collaborators, and Jason Riis has done double duty as collaborator and advisor. He has helped me to navigate the submissions process, to improve my writing, and to ask the right questions. I am very grateful that he said yes when I first asked for his help and has continued to be there ever since. I am also immensely grateful for our collaborators from RedBrick Health, particularly Chelsie Hamsa who made these experiments possible.

The MIT Marketing Group has been a wonderful home for the last five years and all of the faculty have been willing to listen to ideas and provide advice and feedback. Thank you to Birger Wernerfelt, Catherine Tucker, Duncan Simester, Glen Urban, John Little, Josh Ackerman, Juanjuan Zhang, Michael Braun, Sharmila Chatterjee, and Sinan Aral.

Many thanks also go to the PhD students, postdocs, and researchers with whom I have shared the experience and who have offered their endless help, wisdom, and camaraderie: Amal, Artem, Ann, Cristina, Daria, Derek, James, John, Nathan, Nell, Song, Sachin, Xiang, Xinyu, and Yu-Wei. And a very heartfelt thank you to all the participants in B-Lab, who heard countless versions of my research in many different stages and were always willing to listen and provide thoughtful feedback.

My years here have been made much easier and more enjoyable by the very necessary help provided by our group assistants Jason and Tina and by support and encouragement from Hillary and Sarah in the PhD program office.

Finally, while I cannot name them all, I wish to thank the many people who provided encouragement, love, and support during this process. My children have provided inspiration, motivation, and much joy. I have spent their entire lives at MIT and they have been here with me in every way. And last, but certainly not least, my husband Brendan has supported me in every way, both literally and figuratively, and allowed me to pursue work that I love rather than just endure. Without him, my world would be a much smaller place.

ESSAY 1

"Yes/No/Not right now." Yes/No Response Formats Can Increase Response Rates Even in Non-forced-choice Settings

Abstract:

Although yes/no response formats have been used to increase enrollment rates in several different types of programs, their use has been largely limited to forced choice settings. The effects on post choice engagement have not been measured. Across two field experiments, the authors demonstrate a substantial advantage in click-through rates for a yes/no response format over traditional opt-in response formats in an email context where choice is not forced. The increase in click-through rate does, under certain conditions, also persist through downstream program enrollment and participation. Finally, though noting that the yes/no format advantage is probably multi-determined, the authors discuss several potential psychological mechanisms, which are particularly relevant in non-forced choice settings. The authors also discuss how the yes/no response format might operate in other settings, such as the implementation of mandated choice for organ donation.

Keywords: choice architecture, decision making, field experiment, participation

Choice architecture is the art and science of framing options in order to push people towards a desired decision. Among other things, it has been used to increase rates of retirement savings, organ donation, and flu vaccination, often through the assignment of appropriate default options (Choi et al. 2003; Johnson and Goldstein 2003). Because marketers and program designers are often unable or unwilling to assign default options, researchers have started to look for other approaches. One recent example is the use of yes/no response formats, where there is no default option, and consumers are asked to make an explicit choice between “yes, I will enroll” and “no, I will not enroll” (Carroll et al. 2009; Keller et al. 2011).

This yes/no response format has been used in a variety of forced choice settings, and, in some cases, it has led to a marked increase in program enrollment over a more standard opt-in response format (e.g., “click here to enroll”). Forced choice has, however, not always been effective at increasing enrollment (for example, see Kessler and Roth 2014), and, more importantly, forced people to choose a response option is not always feasible. There may also be concerns about the commitment of people who enroll as a result of a mandated choice (as it is often called in practice), particularly if the choice was made quickly, without adequate information (Cotter 2011; Klassen and Klassen 1996). For some important life choices, forcing choice may be ethically inappropriate (Chouhan and Draper 2003).

An examination of the impact of yes/no choice formats outside of forced choice contexts is theoretically and practically important. In this paper, we conduct such an examination through two large email-based field experiments where we measure the impact of choice format on email click-through rates and on downstream behavior (specifically, enrollment and participation in a workplace wellness program). We find

strong positive effects of the yes/no format on initial choices as measured by clicks, and mixed effects on subsequent participation.

This issue of active participation and engagement has also not been studied in previous tests of the yes/no format, which measured both enrollment and subsequent rates of disenrollment as an indicator of participation. We examine the full enrollment funnel, because (1) marketing techniques that attract additional clicks do not always lead to additional conversions or enrollments (Agarwal, Hosanagar, and Smith 2011), and (2) additional enrollments do not always lead to additional usage or participation (Bowman, Heilman, and Seetharaman 2004; Cutler and Everett 2010; Soman and Gourville 2001; Zhu, Billeter, and Inman 2012). Active participation is particularly important in areas such as wellness, where the full benefit of the program comes only with ongoing commitment and engagement.

CONCEPTUAL FRAMEWORK

Choice Architecture and Active Choice

Defaults have enormous impact on enrollment. Extensive research on 401(k) enrollment has shown that more people end up enrolled in retirement savings programs when their employers automatically enroll them (Choi et al. 2002; Madrian and Shea 2001) and more people end up as organ donors when it is the automatic default for a driving license (Johnson and Goldstein 2003). While the opt-out approach (resulting in automatic enrollment) can be an effective tool in policy contexts, marketers are much more restricted in their opportunity to apply an opt-out approach because they risk backlash or even litigation (e.g. Design Less Better 2008). Even when opt-out approaches can be used, they sometimes lead to high levels of waste, as

Just and Price (2014) observed in a school cafeteria study – when vegetables were automatically served with every meal, 74% of the served portions ended up in the trash. This waste highlights the importance of getting consumers to actively participate in a desired program, rather than just being passively “enrolled” in that program.

While program managers and marketers want to avoid the risks and negative side effects of opt-out approaches, they also seek other ways to increase enrollment in programs from the generally low rates that are achieved with simple opt-in approaches. Forcing respondents to make an active choice between “yes” and “no” alternatives has been seen as a viable option. For example, with new employees who are completing a benefits enrollment process, rather than inviting them to tick an optional check box to enroll in a 401(k) plan, they could instead be required to explicitly indicate their choice by ticking either the check box to enroll, or by ticking a different check box if they do not wish to enroll (Carroll et al. 2009).

Forced Choice in Implementation

Active choice response formats that were tested in forced choice settings (where participants must choose an answer before proceeding) have often led to enrollment rates that are higher than those achieved with opt-in response formats. Increased enrollment was not observed, however, in a recent study of real organ donation decisions (Kessler and Roth 2014). Much of this research about the benefits of active choice has taken advantage of settings where participants feel they have to make a choice before continuing to the next step in a particular enrollment process.

Most marketers, however, cannot actually force consumers to choose. And the actual implementation of “mandated choice” for organ donor registries in most states

has avoided forcing people to choose yes or no, in part because they do not want people to make a binding no decision (Thaler and Sunstein 2010). In Virginia, for example, the implementation of “partial mandated choice” uses the options: donor, non-donor, and undecided (Klassen and Klassen 1996). The addition of this third option (undecided) presents the important questions of whether and, if so, why people would choose to enroll when they have the option to defer the decision and thus avoid the perceived negativity of the “no” response without necessarily saying yes.

Potential Mechanism and Impact on Participation

This important addition of the third (undecided) option also raises questions about whether the mechanisms described for forced choice settings would apply when respondents can ignore or defer the choice. Previous explanations for the success of active choice have suggested that “yes” allows respondents to avoid the negative feelings of saying “no” (such as regret or guilt) (Keller et al. 2011). If participants have the additional option to defer the decision and ignore the no, such mechanisms should be less effective. We believe that any advantage of active choice (either with or without an option to defer) is likely to be multiply determined, and that the specific language may prove important in pushing people to choose yes and that their reasons for choosing yes might impact their subsequent participation.

One possibility is that the contrast of positive and negative alternatives in the yes/no format might push people toward the positive response. There is evidence from the survey design literature that people have a general inclination to agree with stated suggestions, known as acquiescence response bias (Paulhus 1991), whether presented as yes/no or agree/disagree (Krosnick and Presser 2010). Additionally, a respondent might believe that the presence and wording of the no option implies that the offer will

no longer be available if he does not respond immediately, thus prompting a feeling of scarcity and urgency which could lead to an increased likelihood of enrollment (Lessne and Notarantonio 1988). Inman, Peter, & Raghurir (1997) argue that the presence of a restriction activates a cognitive response which frequently leads to an inference of good value.

Separately from the dichotomy of the yes/no language, the positive first person language generally used as part of the active choice format (i.e. “I would like to enroll” and “I would not like to enroll” versus “Click here to enroll” which does not use first person language), might lead to an increased likelihood of visualizing oneself doing the activity under consideration and a resulting increase in the behavioral intention that leads to a decision to enroll (Rennie, Harris, and Webb 2014). While we will not disentangle these mechanisms here, we note that they may all work together to promote the positive choice, even when participants are not forced to choose a response, and the various potential mechanisms might have differing effects on the likelihood of ongoing participation.

It is far from obvious that increased clicks will lead to increased participation, and many studies of health communications (similar to our experimental setting) lack sufficient data to be able to make any conclusions about behavior or the link between intentions and health behaviors (Keller and Lehmann 2008). Although studies of active choice have shown higher enrollment, they have not generally measured participation in an active form, as they have focused on programs where participation is automatic after enrollment (e.g. retirement plan enrollment). Family members of organ donors who were enrolled as a result of mandated choice have also questioned whether this was a true choice (Klassen and Klassen 1996). It may be that the yes/no response format leads to a short term preference for agreement that is later forgotten. It may

also encourage enrollment among those who were inherently less interested and therefore less likely to actually participate, even if they do enroll. This paper thus applies the active choice format to the substantially larger domain of non-forced choice settings, while exploring the potential for sustained commitment resulting from that initial choice.

EXPERIMENTS

Overview of Experiments

To test the effects of the yes/no response format in a non-forced choice setting, we conducted two field experiments in partnership with a large wellness services company and several of their employer clients. The company sends regular emails to potential participants aimed at recruiting them to join programs and encouraging their participation once enrolled. All potential program participants had completed a health screening/assessment and were then targeted to one or more specific programs with the goal of increasing their participation in wellness-based activities.

We focused on two particular programs, both of which involved online tracking of health activities, which allowed for measurement of ongoing participation. One program was focused exclusively on activity tracking (Boost), while the other was designed to encourage several different types of health behaviors (Journeys). The company's extensive online enrollment and participation tracking system allowed us to measure whether each respondent clicked a link in the email, whether they completed an enrollment process, and each time they recorded behavior in the online tracking system for 60 days following the message date.

Experiment 1 tested the effectiveness of a reminder message by asking current enrollees to log physical activity into the Boost tracking program. This test provided a useful benchmark for the impact of the yes/no format on click responses and subsequent participation. However, it was not a test of the impact on the overall enrollment rate because the respondents were already enrolled in the program. Experiment 2 tested a recruitment message for a new program (Journeys) and enabled us to measure the response at each stage of the enrollment process, from clicks to enrollment to actual participation.

Experiment 1

Our first experiment tested the basic question of whether a yes/no response format would lead to a higher click-through rate than would opt-in response formats in a non-forced choice setting. This email was sent as a reminder message to people who had enrolled, but had not recently participated in the Boost activity tracking program. Boost rewards participants for recording at least 30 minutes of physical activity each day. All recipients had enrolled in this program at some point in the past, but had not recorded activity in at least 35 days. They were now being encouraged to participate again.

Design. All participants (N=24,863 emails delivered) received emails with the same introductory text, and with one of three randomly assigned response options.

The introductory text was as follows:

You are currently enrolled in Boost Physical Activity Program, but you haven't tracked your activity in over XX days. Boost makes reaching your physical activity goals easier—playing with your kids, dancing, taking the stairs, and even house cleaning counts in the Boost program.

The randomly assigned response options were as follows (underlined text was a clickable link):

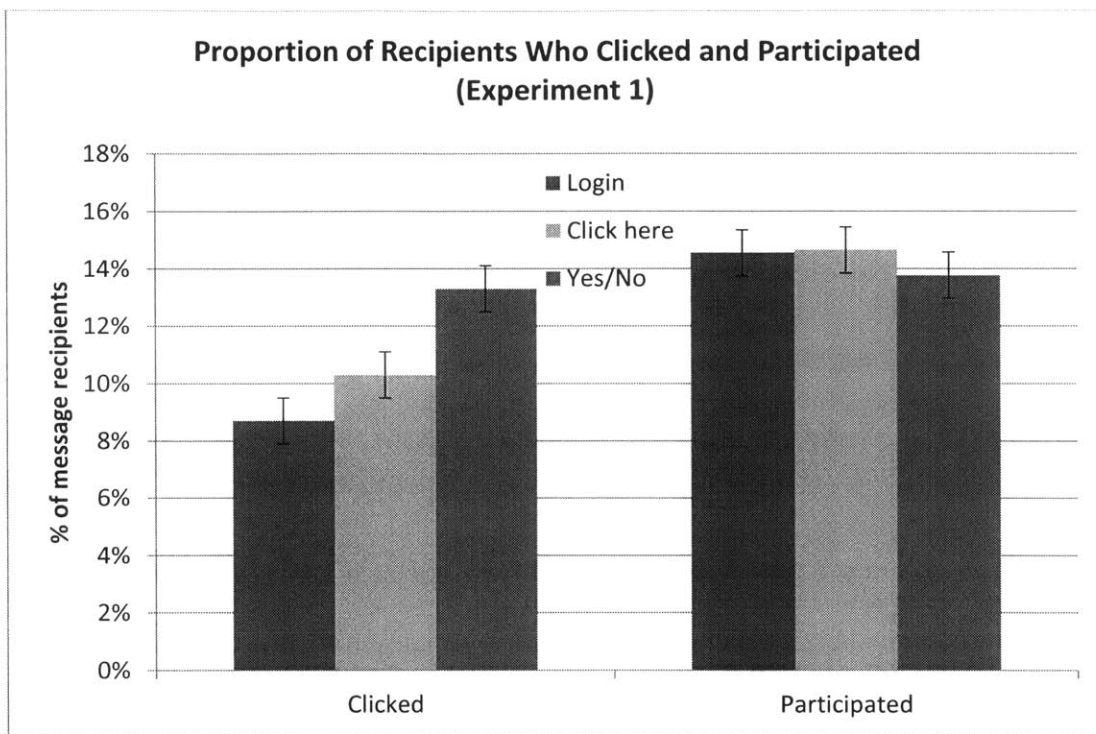
Opt-in 1 (n=8,227):	<u>Login</u> to [website name] to track your activity this week.
Opt-in 2 (n=8,876):	Click <u>here</u> to track your activity this week.
Yes/No (n=7,760):	<u>Yes</u> , I would like to track my activity this week.
	<u>No</u> , I do not want to track my activity at this time.

The first opt-in message was based on the company's standard opt-in message (which asks recipients to "login" to the website as a call to action). The company had typically used the "login" language as a more specific call to action, but we were concerned that the specificity of the call to action might appear to be onerous and deter people from clicking. The second opt-in message, therefore, was a simpler demand ("click here" as the call to action).

If a respondent clicked on the link in one of the opt-in conditions, or on the "yes" option in the yes/no condition, they were re-directed to a page where they could record their physical activity. We also measured whether participants recorded physical activity in the week following delivery of the message. Participants who clicked the no link in the yes/no condition were not directed to this page, but were instead directed to a page that simply thanked them for their response.

Results. There were 24,863 total emails delivered across all three message conditions, with slightly imbalanced final group sizes resulting from variation in the failure rates of the messages delivered. In all three conditions, recipients were 50% female. The mean age was 44 years in both of the opt-in conditions, and 45 years in the yes/no condition (min 18, max 84). The results discussed below focus on response

rates as a percentage of all messages delivered in order to avoid any potential self-selection issues resulting from who chose to open and read the messages.¹



There were significantly more people clicking the yes link in the yes/no condition than there were people clicking the response link in either of the two opt-in conditions (yes/no 13.3% of all message recipients clicked, opt-in combined 9.5%, $z=-8.98$, $p<.001$). There were also significantly more clicks from the low threshold “click here” message compared to the higher threshold “login” message (10.3% versus 8.7%, $z=3.49$, $p=.005$ for click here versus login), although each click rate separately was significantly lower than the yes click rate in the yes/no condition . This strong positive click result for yes/no remained when controlling for the level of previous engagement,

¹ We did not analyze responses as a function of the proportion of people opening the email (open rates), as is sometimes done in email marketing. Open rates, in this case, were not independent of the click-through rate. Recipients who read the message quickly in the viewing pane of their email program without double clicking on the message often do not register as having opened the message. Open rates thus understate the number of people who saw the message. Clicking a link in the message automatically records the recipient as having opened the message. A higher click-through rate thus inflates the open rate, making it a biased measure.

age, and gender. In addition to the enrollment clicks, 2.7% of message recipients in the yes/no condition clicked the “no” link, which took them to a message thanking them for their response. No additional information was collected from these respondents.

In terms of downstream program participation, we found no significant differences between the groups. The yes/no group was marginally less likely to actually track physical activity in the week following the message (14.0% of message recipients versus 14.8% of opt-in recipients, $z=1.71$, $p=.09$). Curiously, the overall proportion of participants tracking a health behavior in the week following the message (14.3%) was actually higher than the proportion of participants clicking a link in the reminder email (10.7%). A substantial number of people therefore tracked activity without ever clicking the link, even though they had not tracked activity in the 30 or more days prior to the message. This suggests that the message may have served as a reminder to log in to the system at a later time. The reminder impact may have been stronger than the actual call for an immediate click through to the website. (Recipients may not have had a physical activity to track at the time they received the email message.)

Discussion. We see a substantially higher click-through rate for the yes/no message compared to the two opt-in messages. However, more than half (62%) of people who click the link in the yes/no format do not actually track their activity, which raises the question of why they chose to click at all. This proportion was substantially higher for yes/no than in either the “click-here” or “login” formats (50% and 49% respectively, combined mean 50%, $z=6.10$, $p<.001$). One possible explanation is that the yes/no format resulted in more low-consideration clicks. This would be consistent with research on acquiescence bias which suggests that people often agree

initially and then reconsider their response during a more effortful later stage of decision making (Knowles and Condon 1999).

Another possibility is that the text of the “click-here” and “login” messages resulted in a more specific expectation of what people would be asked to do following the click. The opt-in messages gave a very specific call to action (“click here to track”) while the yes/no message (e.g., “yes, I would like to track”) may have only implied a general agreement with the statement rather than an immediate intention to act, or even a general intention that the respondents later forgot. This may be compounded by the fact that most people track activity at the end of the day, but emails are sent to work email addresses and are likely read during the day, thus leading to a separation in time between reading (and responding to) the message and actually tracking the activity.

The second experiment addresses these concerns by targeting a new program with which recipients were not already familiar, and in which they were not already enrolled. The new experiment also asks recipients to take an immediate action by enrolling. It allows us to more directly test the effects of the yes/no format on the full funnel sequence – click response, immediate behavior via enrollment, and subsequent participation.

Experiment 2

In Experiment 2, we tested a message inviting people to enroll in a new program. This second study added an immediate action (enrollment) following the click response, and eliminated any potential effects of past experience with the program. This additional action of enrollment should have a positive effect on participation as it

should increase the recipient's commitment to the choice that had just been made by clicking (Cialdini and Trost 1998).

We designed the experiment to also test whether emphasizing (by repeating) information about rewards and other benefits of participation has a positive effect on click-through rates, enrollment, and participation. There are several reasons to expect that repetitive emphasis should help. Research on repetition and familiarity suggests that repetition increases the likelihood of processing a message (Hawkins, Hoch, and Meyers-Levy 2001), and therefore the likelihood that people will respond positively (Skurnik et al. 2005). Additionally, the mere act of making the benefits of enrolling more salient (i.e., more easily visible) should increase the likelihood that the benefits will be noticed and processed (Bettman, Luce, and Payne 1998).

In the domain of active choice, Kessler and Roth (2014) suggest that giving people information about the benefits of their enrollment may actually be more effective than mandating choice. Keller et al. (2011) argue that emphasizing the positive outcomes of the enrollment choice and the negative consequences of the non-enrollment choice should enhance the appeal of the enrollment choice. In Experiment 2, we test this emphasis manipulation both jointly and separately from the yes/no format (where both benefit and foregone benefit are emphasized) in a 2x2 crossed design. The experimental design allows us to compare relative effect sizes as well as potential interactions.

Design. All participants received an email with the following introductory text:

*Try R*** Journeys™, a refreshing, re-energizing alternative to the usual wellness program. Bite-sized, fun ways to get active, eat healthier, lose weight or stress less.*

Starting with as little as a minute a day, your journey grows with you, giving you fun, new things to try at each step. Connect to your journey online from your computer, mobile device or tablet. With journeys you'll not only feel and look better, but you'll also be able to earn up to [X program name points] as you go.

Emails were randomly assigned to have one of the following four calls to action (underlined text was a clickable link):

Basic Opt-in (n=2,129):	<i><u>Log in</u> to R*** Health and try R*** Journeys.</i>
Opt-in with emphasis (n=4,248):	<i><u>Log in</u> to R*** Health and try R*** Journeys. Each activity completed helps to earn more [program name points].</i>
Basic Yes/No (n=4,267):	<i><u>Yes</u>, I would like to try R*** Journeys. <u>No</u>, I do not want to try R*** Journeys.</i>
Yes/No with emphasis (n=4,295):	<i><u>Yes</u>, I would like to try R*** Journeys. Each activity completed helps to earn more [program name points]. <u>No</u>, I do not want to try R*** Journeys because earning [program name points] is not important to me at this time.</i>

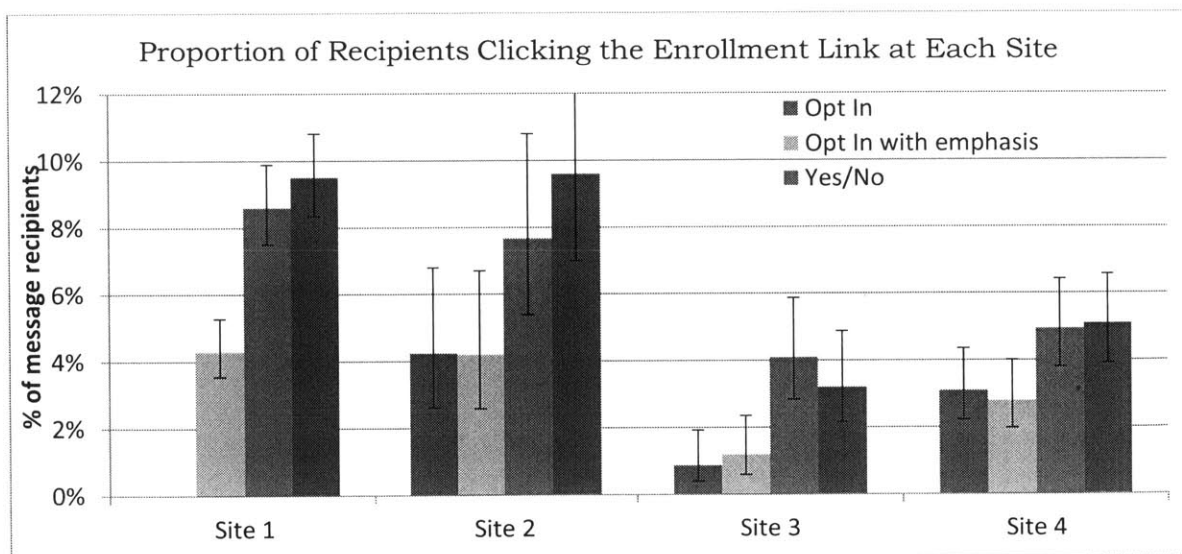
We ran the test at several different client (i.e., employer) sites in order to reach a larger number of respondents. While our initial goal was to randomize all four messages at each client site, we achieve full randomization at three of the four sites. At one site, the research partner wanted to test an unrelated message instead of the basic opt-in message, leaving the four cells imbalanced. Since all yes/no results were in the same in direction and similar in magnitude with the 3 condition site removed, we retain the site in the analyses here to preserve power. All results are given as a proportion of messages delivered to avoid any bias associated with the smaller size of this condition and to avoid any potential self-selection issues resulting from differing open rates (as in Experiment 1).

The messages varied slightly depending on the terminology and specific rewards offered by each client to their employees (designated as “[program name points]” in the message text, above). These differences had no meaningful effect on the outcome of the

yes/no framing. To test robustness, we include tests with mixed effects and with fixed effects to control for site-specific differences.

All participants who clicked an enrollment link (whether from the opt-in or yes/no message) were directed to the same enrollment page, where they were asked for information about their health goals in order to tailor the program to their specific needs. They were also asked to commit to honestly recording their participation. Participation was measured in terms of “steps” completed, with a particular step taking anywhere from 1 hour to several days to complete. Steps could be as simple as watching a video about nutrition, or as complex as adding a vegetable to each meal for a week.

Results. Across all employer sites, there were 14,929 message recipients, of which 50% were female, with a mean age of 44 years (min 19, max 84). Sixteen percent of recipients were spouses or partners rather than employees (74% of the spouses/partners were women). In general, spouses were slightly less likely to respond to the message, and women were more likely to respond overall, but neither group variable interacted with the message format (see Table 1 for details).



We again found a strongly positive effect of the yes/no format on click-through rates. Click-through rates from yes/no were more than double those of the opt-in message (7.2% of message recipients clicked the enrollment link in the yes/no conditions versus 3.2% in the opt-in conditions, $z=-10.67$, $p<.001$) Table 1 summarizes the robustness checks using a logit model with all effects. The effect of emphasizing the benefit was not significant on click rates (5.4% compared to 5.5% of recipients clicked, $z=.036$, $p=.97$), nor was the interaction between the emphasis and the message response format. Across all employer sites and yes/no messages, approximately 1% of people clicked a “no” link (ranging from .4% to 1.8% at different sites).

Table 1: Logit Coefficients Including Robustness Checks for Clicking the Link

<i>Explanatory Variable</i>	<i>Base logit</i>	<i>Logit with controls</i>	<i>Fixed effects logit a</i>	<i>Random effects logit</i>
<i>Yes/ No</i>	1.026*** (.149)	1.011*** (.152)	.692** (.254)	.707*** (.157)
<i>Emphasis</i>	.294+ (.160)	.278+ (.163)	-.074 (.268)	-.030 (.168)
<i>Yes x Emphasis</i>	-.213 (.181)	-.194 (.184)	-.072 (.300)	.108 (.188)
<i>Female</i>		.293*** (.076)		
<i>Spouse</i>		-.252* (.109)		
<i>Age</i>		.023*** (.003)		
<i>Past activity</i>		.023*** (.003)		
<i>Rewards offered</i>		.874*** (.136)		
<i>Constant</i>	-3.630***	-5.612*** (.246)		-3.480*** (.258)
<i>Number of observations</i>	14,939	14,920	5,999	14,939
<i>Number of groups</i>			4	4

+ $p<.01$, * $p<.05$, ** $p<.01$, *** $p<.001$, a random partial sample to avoid Stata numeric overflow

We found a strong but smaller positive effect of the yes/no format on the proportion of people who enrolled in the program as well (3.8% of all messages

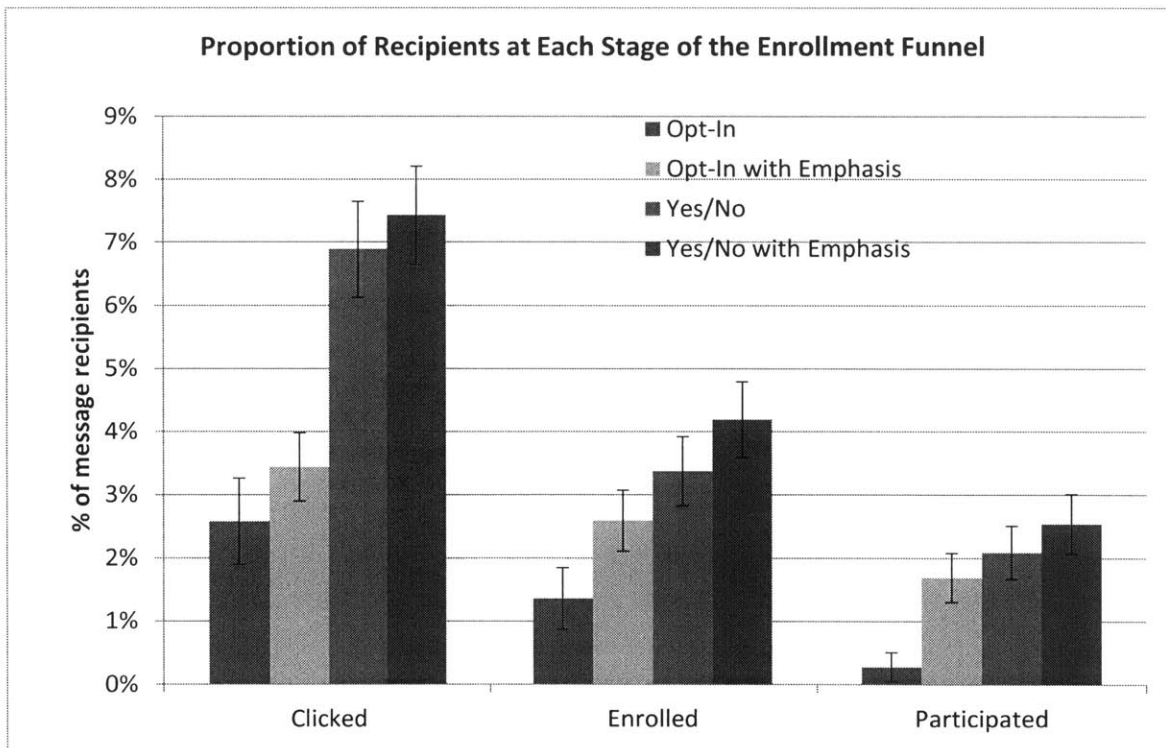
delivered from yes/no versus 2.2% from opt-in, $z=-5.60$, $p<.001$). These results were consistent not just as a basic test but also as a logit regression with control variables. The results were also consistent using mixed effects and fixed effects logit regressions to account for the differences in rewards and program structure at each employer site (see Table 2). While emphasizing the benefit had no effect on click-through rates, it did have a positive effect on enrollment. The effect is greatly reduced in the fixed effects and mixed effects models when controlling for the different reward structures (see Table 2), and goes away entirely when the use of monetary rewards is included as an interaction variable with emphasis. This would suggest that the principal benefit of the emphasis manipulation is in making the presence of the rewards more salient, and that the more highly valued the rewards, the greater the impact of increased salience.

Table 2: Logit Regressions Including Robustness Checks for Enrolling

<i>Explanatory Variable</i>	<i>Base Logit</i>	<i>Logit with Controls</i>	<i>Logit with controls & interaction</i>	<i>Fixed effects logit^a</i>	<i>Random effects logit</i>
<i>Yes/No</i>	.928*** (.205)	.857*** (.207)	.895*** (.208)	.363* (.163)	.441*** (.109)
<i>Emphasis</i>	.655** (.210)	.590** (.213)	.007 (.364)	.083 (.154)	.172+ (.103)
<i>Yes x Emphasis</i>	-.430 (.239)	-.364 (.241)	-.401+ (.242)		
<i>Female</i>		.160 (.098)	.163+ (.098)		
<i>Spouse</i>		-.049 (.134)	-.052 (.138)		
<i>Age</i>		.018*** (.004)	.018*** (.004)		
<i>Past activity</i>		.015*** (.003)	.015*** (.003)		
<i>Rewards offered</i>		.831*** (.173)	.474* (.235)		
<i>Rewards x emphasis</i>			.676* (.343)		
<i>Constant</i>	-4.282*** (.187)	-5.874*** (.320)	-5.593*** (.338)		-4.102*** (.316)
<i>N</i>	14,939	14,920	14,920	5,999	14,939

+ $p<.01$, * $p<.05$, ** $p<.01$, *** $p<.001$, ^a random partial sample to avoid Stata numeric overflow

In contrast to Experiment 1, we saw significantly higher numbers of people participating from the yes/no format. This was true as measured by the percentage of people who committed to start a journey (2.8% of recipients in yes/no versus 1.4% of recipients in opt-in, $z=-5.71$, $p<.001$), and as measured by the percentage who actually completed at least one journey step (2.3% of respondents in yes/no versus 1.2% of people in opt-in, $z=-4.89$, $p<.001$). The effect of emphasizing the benefit was also positive (2.4% committed to a journey and 2.1% completed a journey with benefit emphasis versus 1.9% and 1.5% without, $z=-2.12$, $p=.03$ and $z=-2.84$, $p=.004$ respectively).



While the percentage of people completing at least one step was similar in both groups (54% in opt-in versus 61% in yes/no, $z=-1.3$, $p=.17$), the number of journey steps completed by participants in the opt-in condition was higher than the number of steps completed by those who enrolled from yes/no (16.8 compared to 9.7, $t=1.89$, $p=.06$). With “non-participating enrollees” excluded (those who enrolled but never

participated), the average number of steps taken was 30.9 for opt-in enrollees and 15.8 for yes/no enrollees ($t=2.40$, $p=.01$).

The very high variances in these samples suggest that we should be cautious interpreting these differences in participation intensity. Indeed, the differences are not significant in either case (i.e., with non-participating enrollees excluded or included) if we use the more conservative t-tests for unequal variance ($p>.10$ in both cases). That said, the results are consistent with the possibility that the additional people who enroll through the yes/no link are at least as likely to actually go ahead and try the program (i.e., complete at least one step), but they also may participate with less intensity (complete fewer steps). This interpretation would be consistent with a self-selection argument for participation, whereby the yes/no message is attracting incremental respondents who are lower in average motivation to participate. It would also be consistent with the idea that the language of the yes link (“I would like to try...”) is generating clicks because of the aspirational tone which is not present in the opt-in message (“login ... and try”), and that these clicks might not translate through to actual behavior.

Discussion. We again see strong support for the yes/no format in terms of increasing click-through rates, but this time we also find that the click-through advantage persists, resulting in higher numbers of respondents actually enrolling and participating in programs. While there was a substantial drop off between the number of people clicking and the number of people ultimately enrolling from the yes/no format (47% fewer people enrolling than clicking for yes/no compared to 31% for opt-in), the click advantage for the yes/no format was large enough that it still resulted in higher numbers of people enrolling. This higher level of enrollment came with no incremental drop off between enrollment and participation.

We also see support for the idea that emphasizing the benefit increases enrollment (though not click-through), but the fact that the effect is non-significant when certain other variables (such as the presence of tangible rewards) are included suggests that it is closely related to the actual benefits available, as discussed earlier. Overall, the effect of the yes/no format is larger and more consistent than the effect of emphasizing the benefits of participation.

GENERAL DISCUSSION

Summary and Managerial Implications

The yes/no format gets people to click. Across two experiments, 40-125% more people clicked from the yes/no format compared to opt-in. This click-through advantage also led to 89% more people participating in the program when an enrollment step immediately followed the click (Experiment 2).

We note two key (and related) differences between the experiments which may have implications for why we see a participation effect in Experiment 2 but not Experiment 1. The first key difference is that, in Experiment 2, the desired action (enrolling), immediately followed the choice response (clicking on the link). In Experiment 1, on the other hand, where program enrollment had already occurred weeks or months previously, the desired action was now the tracking of physical activity. This tracking would, in most cases, have been considerably delayed with respect to the initial click response. Tracking was usually done in the evening; a participant might not yet have had any physical activity to track at the time of clicking. The immediacy of the follow-up action to the choice response in Experiment 2 might have increased the effectiveness of the yes/no click-through advantage since

participants were less likely to forget their positive click response between the click and the subsequent action. This suggests that marketers need to be cognizant of the action they are asking respondents to take, as the response advantage is stronger when the action is immediate.

Relatedly, in Experiment 2, the mere act of enrolling immediately after clicking may have served to increase the new enrollee's commitment to participate. Respondents may interpret their yes response as information about their actual underlying preferences and believe that they wish to participate (Amir and Levav 2008). In addition, respondents may perceive an active decision as increasing accountability and thereby feel an increased commitment to the decision (Lerner and Tetlock 1999). Even the mere act of clicking a link may lead to increased commitment, as it establishes a prior action with which respondents will wish to be consistent (Cialdini and Trost 1998), and this effect would be even larger if respondents exert effort in completing an enrollment process immediately following the click. This additional commitment opportunity was not available in Experiment 1, since employees had enrolled in the program weeks or months prior. Organ donation questions might take advantage of this opportunity for commitment by asking people an additional question about their donation preferences after the initial agreement. A strong and specific commitment action might reduce enrollment, but increase the proportion of people actually participating in the program and reduce the concerns of family members about the context of the choice.

Another important consideration that we were not able to fully resolve was whether the response format impacted the quality of the participants and their intensity of participation. Given the higher numbers of participants attracted by the yes/no format, it seems possible, and even probable, that this response format

attracts participants who might have a lower ingoing preference for enrollment. If we think of the population in terms of three groups: those who *would* enroll no matter what, those who *would not* enroll no matter what, and those who are undecided, the goal of a marketer (and our goal with this research) is to attract the undecided. Once enrolled, these previously undecided customers may still be less committed than the “enroll no matter what” customers. This lower ingoing interest level may carry over after enrollment into the participation stage. If the yes/no response format attracts more customers with lower initial interest, then we should expect a lower average level of participation from yes/no recipients. While we cannot make any definitive claims about differences in the populations, we did see suggestive evidence of this participation pattern in the analysis of program enrollees. Given the equivocal nature of even those analyses, however, future research should seek to resolve this issue of participation intensity.

Possible Psychological Mechanisms and Opportunities for Future Research

While the effect is large, it is also likely multiply determined and we can only offer suggestive evidence to implicate or rule out various potential mechanisms. More people in the yes/no condition click the link but then fail to actually track their participation (Experiment 1) or enroll (Experiment 2). This suggests that more respondents are agreeing with the statement than are actually changing their behavior. There are several aspects of the language used in the yes/no framing which could be contributing to this effect. The yes/no framing creates both the dichotomy of the positive and negative responses, but it also involves the first person aspirational language of “I would like to....” This positive/negative contrast could prompt initial agreement, but later reconsideration (acquiescence bias), while the first person

language could generate a general intention which does not have enough force to carry through to actual behavior (intention signaling). Both of these factors could contribute to an increased likelihood of agreement with the yes response option, as well as a drop-off between the agreement and the actualization.

Acquiescence bias and intention signaling would both be enhanced by the aspirational nature of the particular programs we were studying – people are more likely to signal agreement with socially desirable health and wellness objectives that they feel they “should” work toward achieving (Milkman, Rogers, and Bazerman 2008). This aspirational agreement effect could be particularly strong when used with behaviors that conform to positive norms, where the desire to see oneself doing the behavior leads to an aspirational positive response (Kallgren, Reno, and Cialdini 2000; Kivetz and Tyler 2007; Rogers and Bazerman 2008) and the desire to be seen as someone who does these socially desirable behaviors leads to an increased likelihood of agreement (Ross and Mirowsky 1984). The element of aspiration might be further enhanced by the use of first person language which may increase the respondent’s inclination to visualize himself doing the particular activity. Future work could seek to disentangle these effects by testing different response formats in areas with less of an aspirational component (e.g., shopper loyalty programs).

Another area for exploration lies in the nature of the no response option. For example, offering a more positive no, such as “no I would not like to enroll in this program, but I would like more information about your other programs” may reduce the effectiveness of an individual message, but improve the potential for cross-selling other programs. Relatedly, in the current studies, only 1-2% of recipients clicked the no option even though choice was not forced. If future studies were able to learn more about why some people do still click no, they might be able to inform future

manipulations of the no response into something that does allow for cross-selling or other marketing activities.

In conclusion, we believe this work has important implications for both practitioners and researchers. Program managers and marketers wishing to increase response rates can use these techniques to increase the effectiveness of their emails to customers. For researchers, we have expanded the application of choice architecture to an email setting, and have shown that the yes/no format works better than opt-in even here, where attention and consideration are low and where choice is not forced. We hope the magnitude of this effect encourages further research exploring the language of choice formats and their influence on immediate and long term consumer behavior.

REFERENCES

- Agarwal, Ashish, Kartik Hosanagar, and Michael D. Smith (2011), "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets," *Journal of Marketing Research*, 48 (December), 1057–73.
- Amir, On and Jonathan Levav (2008), "Choice Construction Versus Preference Construction: The Instability of Preferences Learned in Context," *Journal of Marketing Research*, 45 (2), 145–58.
- Bettman, James R., Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25 (3), 187–217.
- Bowman, Douglas, Carrie M. Heilman, and P. B. Seetharaman (2004), "Determinants of Product-Use Compliance Behavior," *Journal of Marketing Research*, 41 (August), 324–38.
- Carroll, Gabriel D, James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick (2009), "Optimal Defaults and Active Decisions.," *The Quarterly Journal of Economics*, 124 (4), 1639–74.
- Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick (2003), "Optimal Defaults," *American Economic Review*, 93 (2), 180–85.
- Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2002), "Defined Contribution Pensions: Plan Rules, Participant Choices, and the Path of Least Resistance," in *Tax Policy and the Economy, Volume 16*, J. M. Poterba, ed.
- Chouhan, P and H Draper (2003), "Modified Mandated Choice for Organ Procurement," *Journal of Medical Ethics*, 29, 157–62.
- Cialdini, Robert B. and Melanie R. Trost (1998), "Social Influence: Social Norms, Conformity and Compliance," in *The Handbook of Social Psychology*, D. T. Gilbert, S. T. Fiske, and G. Lindzey, eds., New York: McGraw-Hill, 151–92.
- Cotter, Hayley (2011), "Increasing Consent for Organ Donation: Mandated Choice, Individual Autonomy, and Informed Consent," *Health Matrix*, 177 (1999), 599–626.
- Cutler, David M. and Wendy Everett (2010), "Thinking Outside the Pillbox—Medication Adherence as a Priority for Health Care Reform," *New England Journal of Medicine*, 362 (17), 1553–55.
- Design Less Better (2008), "Down with Opt-Out," [available at <http://www.designlessbetter.com/blogless/posts/down-with-opt-out>].

- Hawkins, Scott A., Stephen J. Hoch, and Joan Meyers-Levy (2001), "Low-Involvement Learning: Repetition and Coherence in Familiarity and Belief," *Journal of Consumer Psychology*, 11 (1), 1–11.
- Inman, J. Jeffrey, Anil C. Peter, and Priya Raghuram (1997), "Framing the Deal: The Role of Restrictions in Accentuating Deal Value," *Journal of Consumer Research*, 24 (1), 68–79.
- Johnson, Eric J. and Daniel G. Goldstein (2003), "Do Defaults Save Lives?," *Science*, 302, 1338–39.
- Just, David and Joseph Price (2014), "Default Options, Incentives and Food Choices," 0–13.
- Kallgren, Carl A, Raymond R Reno, and Robert B Cialdini (2000), "A Focus Theory of Normative Conduct: When Norms Do and Do Not Affect Behavior," *Personality and Social Psychology Bulletin*, 26 (10), 1002–12.
- Keller, Punam Anand, Bari Harlam, George Loewenstein, and Kevin G. Volpp (2011), "Enhanced Active Choice: A New Method to Motivate Behavior Change," *Journal of Consumer Psychology*, 21 (4), 376–83.
- Kessler, Judd B. and Alvin E. Roth (2014), "Don't Take 'No' For An Answer: An Experiment with Actual Organ Donor Registrations."
- Kivetz, Yifat and Tom R. Tyler (2007), "Tomorrow I'll be me: The effect of time perspective on the activation of idealistic versus pragmatic selves," *Organizational Behavior and Human Decision Processes*, 102 (2), 193–211.
- Klassen, Ann C and David K Klassen (1996), "Who Are the Donors in Organ Donation? The Family's Perspective in Mandated Choice," *Annals of Internal Medicine*, 125 (1), 70–73.
- Knowles, Eric S. and Christopher A. Condon (1999), "Why People Say 'Yes': A Dual-Process Theory of Acquiescence," *Journal of Personality and Social Psychology*, 77 (2), 379–86.
- Krosnick, Jon A. and Stanley Presser (2010), "Question and Questionnaire Design," in *Handbook of Survey Research*, P. V. Marsden and J. D. Wright, eds., Emerald Group Publishing, 263–314.
- Lerner, Jennifer S. and Philip E. Tetlock (1999), "Accounting for the Effects of Accountability.," *Psychological Bulletin*, 125 (2), 255–75.
- Lessne, Greg J. and Elaine M. Notarantonio (1988), "The Effect of Limits in Retail Advertisements: A Reactance Theory Perspective," *Psychology and Marketing*, 5 (1), 33–44.

- Madrian, Brigitte C. and Dennis F. Shea (2001), "The Power of Suggestion: Inertia in 401 (k) Participation and Savings Behavior," *The Quarterly Journal of Economics*, 116 (4), 1149–87.
- Milkman, Katherine L., Todd Rogers, and Max H. Bazerman (2008), "Harnessing Our Inner Angels and Demons: What We Have Learned About Want/Should Conflicts and How That Knowledge Can Help Us Reduce Short-Sighted Decision Making," *Perspectives on Psychological Science*, 3, 324–38.
- Paulhus, Delroy L. (1991), "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes*, J. P. Robinson, P. R. Shaver, and L. S. Wrightsman, eds., Academic Press, 17–59.
- Rennie, Laura J., Peter R. Harris, and Thomas L. Webb (2014), "The Impact of Perspective in Visualizing Health-Related Behaviors: First-Person Perspective Increases Motivation to Adopt Health-Related Behaviors," *Journal of Applied Social Psychology*, 44 (12), 806–12.
- Rogers, Todd and Max H. Bazerman (2008), "Future lock-in: Future implementation increases selection of 'should' choices," *Organizational Behavior and Human Decision Processes*, 106 (1), 1–20.
- Ross, C E and J Mirowsky (1984), "Socially-desirable response and acquiescence in a cross-cultural survey of mental health.," *Journal of health and social behavior*, 25 (2), 189–97.
- Skurnik, Ian, Carolyn Yoon, Denise C. Park, and Norbert Schwarz (2005), "How Warnings About False Claims Become Recommendations," *Journal of Consumer Research*, 31 (4), 713–24.
- Soman, Dilip and John T. Gourville (2001), "Transaction Decoupling: How Price Bundling Affects the Decision to Consume," *Journal of Marketing Research*, 38 (February), 30–44.
- Thaler, Richard H and Cass R Sunstein (2010), "How Required Choice for Organ Donation Actually Works in Practice," *The Nudge Blog*, [available at <http://nudges.org/2010/10/10/how-required-choice-for-organ-donation-actually-works-in-practice/>].
- Zhu, Meng, Darron M. Billeter, and J. Jeffrey Inman (2012), "The Double-Edged Sword of Signaling Effectiveness : When Salient Cues Curb Postpurchase Consumption," *Journal of Marketing Research*, 49 (February), 26–38.

APPENDIX

In places where ABC is mentioned, it is a placeholder for the company name.

Messages for Experiment 1:

The company divided participants into three levels of previous engagement based on how long it had been since their last participation (35-90 days, 91-180 days, or 181+ days), and then randomly assigned individuals in each level to receive one of three messages. The three messages varied only slightly in the opening paragraph and did not vary at all in the response options.

Here are the variations:

Introductory paragraph (For recipients whose most recent previous engagement with the program was 35-90 days ago):

You are currently enrolled in Boost Physical Activity Program, but you haven't tracked your activity in the last XX days. Boost makes reaching your physical activity goals easier—playing with your kids, dancing, taking the stairs, and even house cleaning counts in the Boost program.

Introductory paragraph (For recipients whose most recent previous engagement with the program was 91-180 days ago):

You are currently enrolled in Boost Physical Activity Program, but you haven't tracked your activity in over 90 days. Boost makes reaching your physical activity goals easier—playing with your kids, dancing, taking the stairs, and even house cleaning counts in the Boost program.

Introductory paragraph (For recipients whose most recent previous engagement with the program was 180+ days ago):

You are currently enrolled in Boost Physical Activity Program, but you haven't tracked your activity recently. Boost makes reaching your physical activity goals easier—playing with your kids, dancing, taking the stairs, and even house cleaning counts in the Boost program.

Response Format for Each Experimental Condition:

The response options did not vary by level of previous engagement but were unique to each experimental condition.

Login Condition:

Login to ABC to track your activity this week.

Click here Condition:

Click here to track your activity this week.

Yes/No Condition:

Yes, I would like to track my activity this week.

No, I do not want to track my activity at this time.

Rewards structure for each site (Experiment 2)

Site 1: (N=6,439) Participants can earn \$125 for each quarter that they earn at least 60 points. Points are earned through completing a health assessment and participating in various types of health improvement activities. Each journey could earn at least 10 points (bonuses for certain activities), with a maximum of 35 points per quarter from journeys.

Note: This was the site with only 3 experimental conditions.

Site 2: (N=1,512) Participation in programs earns wellness bricks (raffle tickets) for Visa gift card prizes worth up to \$100. Participants can earn a maximum of 25 bricks per quarter, and a must earn a minimum of 5 bricks per quarter to be eligible. Participants received 2 bricks for enrolling in a journey and 10 bricks for completing a journey.

Site 3: (N=2,731) This site offered no rewards for participation. Messaging emphasized the health benefits of participation.

Site 4: (N=4,527) Participants can earn up to \$25 per quarter for participating in programs. One point equals \$1, and each journey step was worth approximately 1 point, with bonus points for completing stages and the total journey.

ESSAY 2

The Effect of Benefit Quantification on Goal Setting and Persistence

Abstract:

We look at how language used to describe rewards, specifically the quantification of an expected reward, might lead participants to create specific targets for their own performance based on that language. Through a combination of field and lab experiments, we demonstrate that the use of a high number to describe rewards leads to higher interest and enrollment, but also higher expectations of performance and a higher drop-out rate from the program when the reward is difficult to achieve. Marketers should be aware of this issue when describing benefits to potential customers, particularly if they wish to motivate persistent behavior.

Keywords: anchoring, expectations, goal-setting, participation, goal failure

Targets abound in advertising messages. “Lose up to 4 dress sizes” or “I lost 39 pounds” or even, “Get 5000 points by signing up today.” These messages share a common theme of specific and high numbers, designed to entice people who want to attain those kind of results themselves. Rarely does one see a message that says, “I lost about 5 pounds” or “earn an average of 2 bonus points.” Headline numbers are designed to attract attention and make the benefits of participation look as appealing as possible.

We propose that people attach themselves to these headline numbers, even when the target is arbitrary or unreasonably difficult to achieve. Marketers mention big potential rewards (lose 39 pounds, earn 5,000 bonus points, etc.) in order to tempt users with specific, appealing offers of benefit. The specificity of these offers increases credibility and persuasion (Nisbett and Ross 1980). But the specificity of the offer might also increase the specificity of the expectation it raises in potential customers. In particular, the use of a number, which both increases concreteness and credibility, might also lead to anchoring and specific product or performance expectations.

If people do anchor on potential outcomes, there may be negative consequences to having such a high target in the minds of new customers as these general targets might become specific goals in the minds of participants. While high, but achievable goals should motivate performance (Locke and Latham 1990), goals which are too high might lead to dissatisfaction, poorer performance, and potentially even abandonment of the goal particularly when consumers do not accept that goal as their own (Erez and Zidon 1984). When goals are set appropriately, they take into account “such factors as their beliefs about what [the goal setter] can achieve, their recollections of past performance, their beliefs about consequences, and their judgments of what is appropriate to the situation.” (Locke et al. 1981 p. 10) Instead of this careful goal

setting process, we would argue that people do not always know what is appropriate to the situation, and may therefore look to extraneous information in establishing personal targets.

Through a combination of field and lab studies, we not only show a real world example of how numeric anchoring can occur in practice, we also contribute to the understanding of how people set goals through environmental cues, by demonstrating that people use the presence of numbers in marketing messages as anchors for their own performance expectations. We also build on the existing literature about goal re-engagement by showing how ongoing participation can be impacted by the goal setting process. Respondents who use situational cues may set their goals unreasonably high and be dissatisfied with an average level of performance.

CONCEPTUAL BACKGROUND

High Specific Targets

There is a large body of research which would justify advertising the maximum potential benefit in order to attract potential customers. Not surprisingly, higher incentives are more attractive than lower incentives for motivating participation (Locke, Latham, and Erez 1988), and higher targets may actually be more motivating and generate more satisfaction than lower targets (Jain 2009). A program goal that is too low might lead to procrastination and lack of effort, whereas a high goal will be motivating as an achievement for some participants and an aspiration to others thus appealing to a broader audience than a low goal which is too easy for many.

For products with a performance component, the use of a high target can be particularly beneficial because it sets a positive initial perception. People tend to

anchor on ingoing expectations of quality and those expectations can be sticky when it comes to actual perceptions of the product's performance (Anderson 1973) leading to an ongoing positive perception. In fact, marketers have an incentive to err on the side of maximum potential performance, as overstatement of benefits can lead to more favorable ratings of product quality than understatements (Olshavsky and Miller 1972).

The specificity of these targets may matter as well. "The weaker the data available upon which to base one's conclusion, the greater the precision which should be quoted in order to give the data authenticity." (Augustine 1997, p.231) While this is a somewhat cynical view, concretely worded messages receive more attention and are more influential than abstractly worded messages, potentially because they are easier to understand but also because they may appear "stronger" or more convincing in their arguments to the viewer. (MacKenzie 1986). Other research, which demonstrates that quantification increases persuasion when the source is credible (Yalch and Elmore-Yalch 1984), would also support the idea that concrete quantified arguments are perceived as more persuasive, as long as the message is not overly technical or delivered by a source perceived to be biased. Overall this suggests that quantification should generally increase the effectiveness of marketing messages and that higher numbers should lead to greater interest.

Anchoring and Expectations

Beyond generating interest, quantification may also cause respondents to anchor on the particular number that is given. The mere presence of a number, regardless of its validity as a target, can often act as a starting point for expectations from which people insufficiently adjust (Epley and Gilovich 2006; Tversky and

Kahneman 1974). The more reliable and relevant that the initial number appears to be, the more strongly it will weigh on expectations and the less people will adjust.

Moreover, consumers may draw inferences from how numbers are placed in context (Prelec, Wernerfelt, and Zettelmeyer 1997) and may be uncertain about how much they should adjust away (Simmons, LeBoeuf, and Nelson 2010). Research has shown that even when disclaimers are placed with numeric indicators, respondents tend to ignore the disclaimer and assume that the number is present for a reason and therefore is a reasonable source of information, particularly when the source is considered trustworthy (Frederick, Mochon, and Savary 2014).

This same credibility which increases persuasion might also create challenges for the message designer. Concretely worded messages which give specific performance targets might logically be seen as more definite than more abstract messages as they lead to accountability, thus creating stronger expectations in the minds of customers. Products which deviate from an expected level of quality are often penalized for the deviation more than products with a lower or less specific expectation (Anderson 1973; Cardozo 1965). While a strong argument should lead to more interest in the product and higher expectations of quality, a deviation from that expected level of quality may be perceived as a larger violation.

There is also evidence that people anchor in their performance expectations for actions, not just products, particularly in the areas of gambling and lotteries. Lottery players are often driven by the maximum potential payout, rather than a calculation of expected return (Forrest, Simmons, and Chesters 2002) and gamblers generally use the amount they could win as an initial anchor, from which they adjust downward (perhaps insufficiently) to account for other aspects of the bet (Lichtenstein and Slovic 1971, 1973). While this has not been specifically applied to the goal setting context, it

also seems possible that people could anchor in setting expectations for their own performance.

Goal Performance and Reengagement

As previously mentioned, goals should be set based on information about abilities and what is appropriate to the situation, and indeed goal performance is better when goals take ability differences into account (Locke et al. 1981). But many goals are set by circumstances beyond the control of the goal setter. In particular, program targets in marketing communications, which are often designed to reach a large number of people, may not be able to take individual differences into account, so that something which is a specific challenging (and therefore motivating) target for one participant might appear daunting for another. For example, current wellness targets often emphasize 150 minutes of exercise per week (at least 30 minutes per day, five days per week), which may seem unattainable to some (<http://www.medicinenet.com/script/main/art.asp?articlekey=186436>). In these cases, the program designer must decide whether to set an easily achievable goal, which is not overly intimidating to those with a lower level of fitness, or one which will be motivational for all potential participants.

When these goals are set, they may serve as reference points, from which future performance can be judged as either success or failure. Given that marketers have an incentive to set high goals, this can lead to situations where a reasonable level of performance might feel like underperformance if the goal is set too high. Indeed, a particular level of performance may actually feel worse for someone with a goal than someone without a goal if it leads to feelings of loss relative to a particular expected outcome (Heath, Larrick, and Wu 1999). Mossholder (1980) notes that subjects with a

goal evaluated their performance less positively than those with no goal, despite outperforming on the assigned task. And those whose commitment to the goal is more uncertain may lose motivation by focusing on the distance to the goal rather than the amount accomplished to date (Koo and Fishbach 2008).

This perceived underperformance may lead to disengagement with the particular target area. While progress towards a goal is often seen as commitment to that goal (Zhang and Huang 2010), performance that is perceived as a personal failure can result in decreased commitment to a goal (Soman and Cheema 2004). This abandonment may result from lower self-efficacy (Bandura 1977) such that people then perceive the goal as being beyond their ability to achieve, or from a desire to protect the ego by making the goal unappealing rather than unachievable (Baumeister, Heatherton, and Tice 1993) and thus making it a problem of the program rather than a lack of ability. Thus it seems possible that high goals might lead to lower relative performance and increased abandonment, particularly if those goals are not tailored to the individual's situation.

In the studies that follow, we demonstrate the positive effect of quantification on initial interest, but also note neutral to potentially negative long term effects. We demonstrate these effects in both a field experiment using actual performance measures over time and heterogeneous respondents, as well as in a more controlled hypothetical setting.

EXPERIMENTS

Study 1

Study 1 is a large field experiment designed to measure responses to an email recruitment message which highlights either a quantified (\$125) or a non-quantified

reward. We believe that the specificity of a quantified potential reward should increase attention to the message and interest in the program, thus increasing the likelihood that respondents will click the enrollment link.

Mentioning the \$125 also serves as a specific challenging target for potential customers, which might increase their motivation to participate if it is set appropriately for their abilities. This, however, is a single target for all potential participants, which does not take individual differences into account. Individuals should (but may not) adjust from the anchor once they read the actual program details at enrollment. This anchor and insufficient adjustment could subsequently cause participants to evaluate the program and their own performance more negatively as their expectations will be too high. We believe that this anchoring effect may dominate when the potential reward is high and participants lack information about their own capabilities, leading them to infer goals from the contextual information.

All participants had completed a health assessment and were given a recommendation to enroll in an online activity tracking program. This program, which gives all participants one point for each day that they tracked at least thirty minutes of physical activity, was part of a larger wellness campaign which was sponsored by their employer. As part of this overall campaign, participants could earn up to \$125 per quarter (as a rebate on their health insurance) by participating in various health and wellness programs. This particular program offered one point (convertible to one dollar) for each day that they tracked at least thirty minutes of physical activity. Thus participants could earn ninety points toward their wellness rewards from this program, but could not earn their entire wellness reward just from this program. Participants could also earn points from participation in other wellness activities

offered by the same company and were given specific information about how many points they could earn at the time they enrolled.

Method. All respondents received the same information in the body of the email and were randomly assigned to one of two response options which followed the descriptive text.² (The underlined word was a hyperlink.)

Quantified (N=4,489): Yes, I would like to participate in the Boost tracking program. I want to earn points toward rewards of up to \$125 per quarter and I understand that the more often I track my activities, the more points I will earn.

Non-Quantified (N=4,429): Yes, I would like to participate in the Boost tracking program. The more often I track my activities, the more points I will earn towards the incentive rewards program.

When a respondent clicked the “yes” hyperlink (in either message), he or she was taken to a landing page with enrollment information for the program and details about how to earn points.

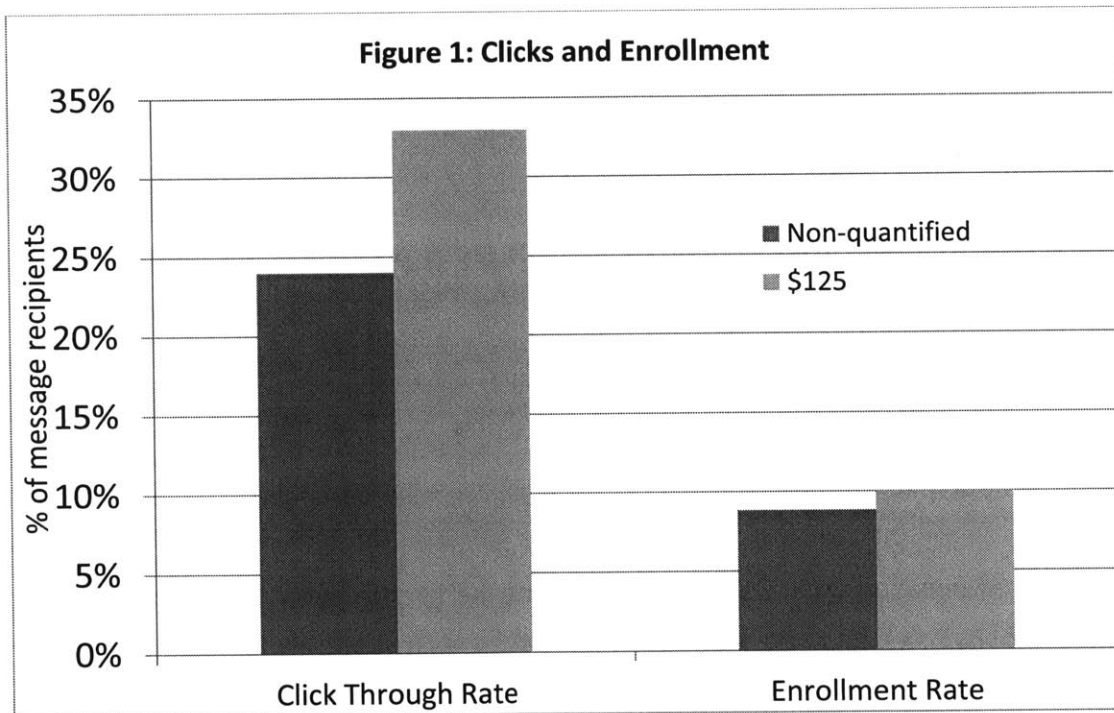
With this design, we were able to measure initial responses via the click-through rate, conversions via enrollment, and ongoing participation via the number of days that they tracked at least thirty minutes of activity. We also had information about whether they had previously earned points in other programs (including in the current quarter), and basic demographic information.

Overall, 8,918 subjects received the message. Thirty-four percent of recipients were female and the average age was thirty-eight years. Sixty-one percent of them had, at some point, participated in one of the company’s online wellness programs, and 31% were currently enrolled in at least one other online program at the time the message was sent. All participants had also been assigned a health score before being

² There was also a contrasting no response which followed the yes, for which there were 2 slightly different versions. These are detailed in the appendix and there was no difference in response rates based on the different no responses.

recommended for the program, with higher health scores being associated with more healthy behaviors.

Enrollment and Initial Participation. Substantially more people clicked the link from the quantified message (33%) compared to the non-quantified message (24%, $z=-9.28$, $p<.001$), suggesting that quantification did indeed increase attention and interest in the program. After clicking, all respondents were directed to the same enrollment page where they were given more information about the program and could complete the enrollment process. Substantially more people from the quantified group dropped out between clicking and enrolling, resulting in a much smaller difference between groups for actual enrollments (10% for quantified versus 9% for non-quantified, $z=-1.93$, $p=.05$, see Figure 1 and Table 1 for logit with controls) compared to clicks.



One potential explanation for the smaller difference in enrollment compared to clicks is that the quantified message attracted different participants than the non-

quantified message. Featuring the dollar amount prominently highlights the extrinsic benefits of participation over the intrinsic health benefits, and therefore might attract people who are more driven by the reward and deter people who care less about the reward. We attempted to control for this by prominently mentioning the existence of rewards in both condition, varying only the specificity of quantification (see Appendix for exact language). While we cannot eliminate the possibility that the quantified message attracted different customers, we see no direct evidence supporting the idea that it did, and the relatively small amount of incremental enrollees suggest that it is more likely to be a substitution effect than an additive effect.

Table 1: Clicks, Enrollment, Participation in Field Experiment (logit coefficients)

	Clicked	Enrolled	Enrolled	Participated beyond 1 day	Participated beyond 2 weeks if enrolled	Participated beyond 2 weeks if participate beyond day1	Participated beyond 2 weeks if participate beyond day1
Constant (base group is non-quantified)	-1.14	-2.33	-3.69	-2.67	.79	2.70	1.55
Quantified (\$125)	.44***	.14*	.15*	.06	-.33*	-.55^	-.52^
Female			.26***				
Health Score			.02***				.01
Participated over 8 times in first 2 weeks							1.66**
N	8,918	8,918		8,918	841	595	580

^ p≤.1, *p≤.05, **≤p.01, ***p≤.001

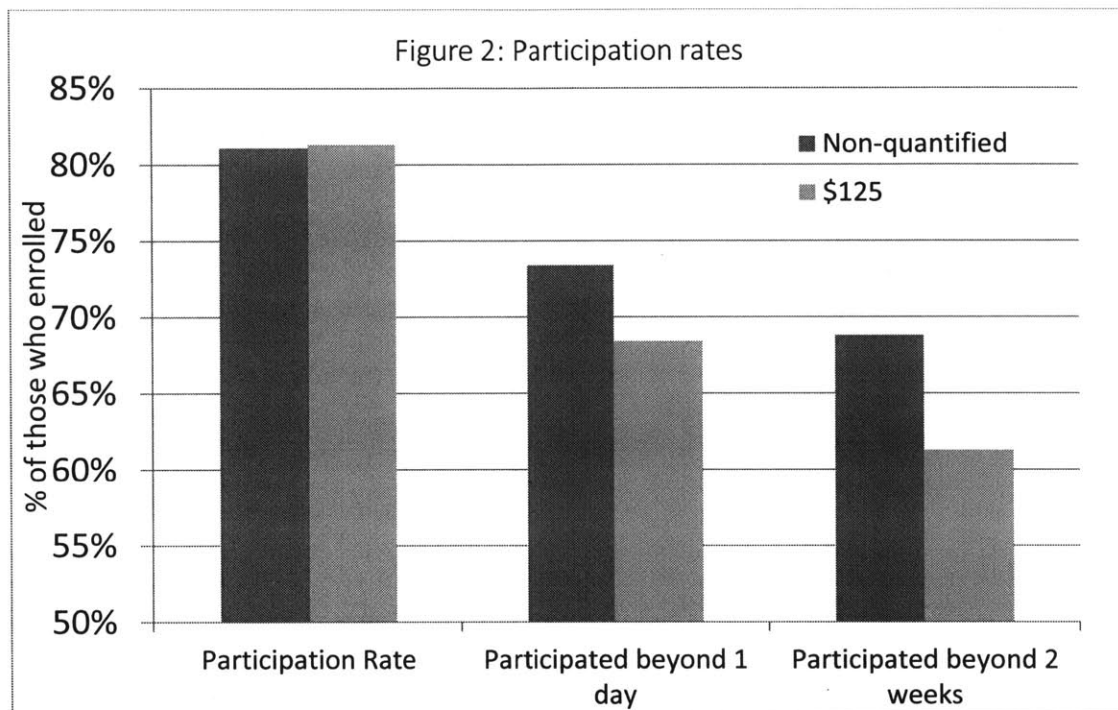
Overall, the number of people participating was slightly higher in the quantified group (M=8% versus M=7%, z=-1.77, p=.08), driven by the larger number of people enrolling, however this difference was only marginally significant and disappears entirely when looking at people who participated for more than one day. In fact, of

those who enrolled from the quantified message, 13% participated for only one day, versus 8% of those who enrolled from the non-quantified message ($z=-2.47$, $p=.01$). In total, only fifty-nine more people enrolled from the quantified message, out of a total population of 8,918, and only twenty-one more people participated for two or more days. Overall, this suggests a very small (if any) incremental gain in actual participants from the quantified message compared to the non-quantified message.

Ongoing Participation. When considering the overall population of message recipients, there is no difference in overall participation between those who received the quantified message and those who received the non-quantified message, whether this is measured by general participating (logging behavior at least once), participating beyond the first day, or participating beyond the first two weeks. This implies that, at a minimum, if the quantified message is attracting additional interest, this incremental interest is not enduring beyond the first day.

In order to estimate the effect of quantification on actual participation, we will look only at the participation rates among the population of enrolled subjects (given the relatively small differences in actual numbers of people enrolling). While we cannot control for selection issues in this field setting, we think it is important to examine the potentially negative effects of quantification among those who did enroll, whether those negative effects are due to population differences or because of message effects. In general, among those who enrolled, the effect of quantification on long term participation (more than two weeks) is significant and negative, with only 61% of participants who enrolled from the quantified message continuing beyond two weeks while 69% of those who enrolled the non-quantified message continued ($z=2.26$, $p=.02$, Figure 2). Overall, the average length of participation in our sample period for those

who enrolled from the quantified message is slightly shorter (37 days) than the length of participation for those who enrolled from the non-quantified message (40 days).



While part of this duration difference is caused by the greater proportion of people in the quantified group who participate for only one day, it does not explain the entire difference. Even when we exclude those who participated for only one day, we still see a moderate difference between the two groups, with fewer of those who received the quantified message participating beyond two weeks (90% of those who participated for more than 1 day) compared to those who received the non-quantified message (94%, $z=1.81$, $p=.07$). In addition, we see this moderate difference persist, even when including controls for health score and the level of participation (see Table 1). While this is a relatively modest effect (Cohen's $d=.14$), it is larger than the positive effect of quantification on enrollment (Cohen's $d=.04$). Thus while we saw the expected positive effect on initial interest from quantification, it was arguably smaller than the

neutral to negative effects on actual participation in the program resulting from the same message.

Overall, these results are consistent with both the idea that the quantified message is attracting more people with a relatively superficial interest in the program (those who drop out after just one day), as well as the idea that it may be decreasing the length of participation from those with a more significant interest (who participate for longer than one day). We also note that this difference in participation is not driven by initial performance. We see no difference in the average frequency of participation in first two weeks between enrollees who received the quantified message or not (5.2 points earned in first 2 weeks versus 4.7, $p=.10$), and no significant difference in the proportion of enrollees participating more than eight times in the first two weeks (16% versus 18%, $p=.4$), indicating that similar proportions of each group are participating at a high level.

Discussion. The \$125 message attracts substantially more participants to click, but most of these additional clicks do not translate through to actual enrollments and participation. This may be a result of superficial interest which wanes once participants realize the underlying program details (failure to enroll), or once participants realize the difficulty in actual participation (dropping out after one day). However, even when we control for information (all participants who enroll see the same details about earning points), and actual participation rates, participants who received the \$125 message are more likely to drop out of the program in the first two weeks.

We believe that this difference in participation is, at least in part, a result of anchoring – that participants who see the \$125 message fail to sufficiently adjust away from the target, and are therefore disappointed with an equivalent level of

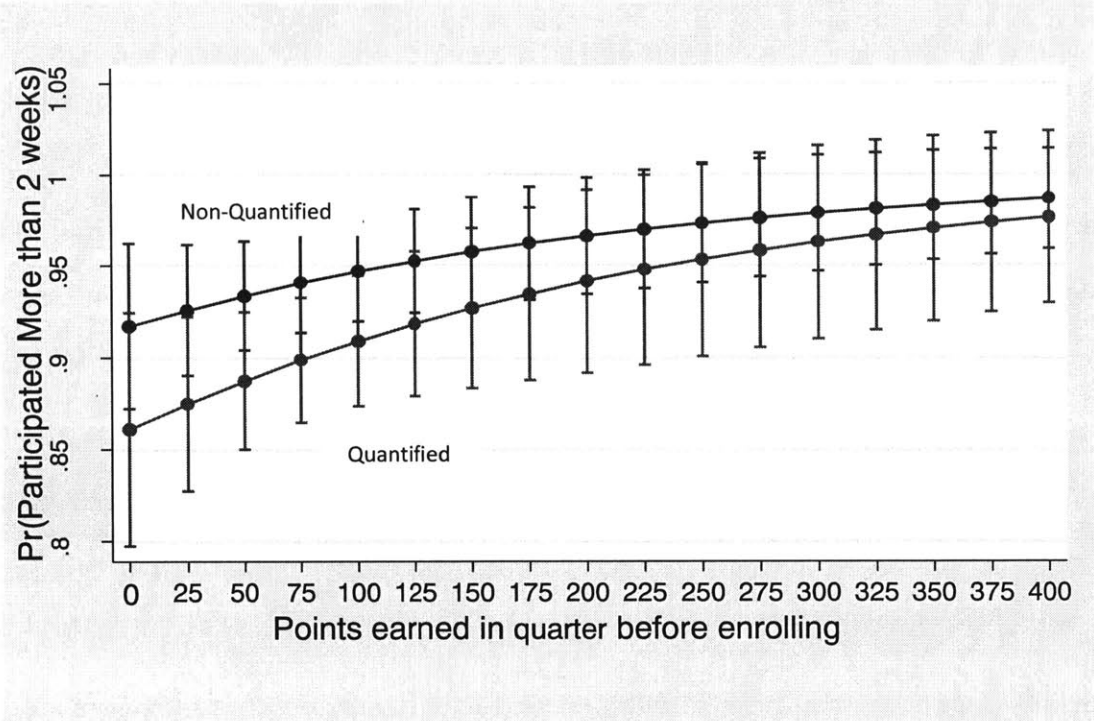
performance in the program. If indeed we are seeing effects that result from anchoring, then the impact of the quantified message should be mitigated by having more information at the time of the message. Knowledge about the actual likelihood of achieving that reward, whether provided by the marketer or in the form of additional experience should mean that participants would be less subject to anchoring.

Our sample did not permit a comparison between experienced and unexperienced customers as almost all the respondents had participated in some other similar program in the recent past (so all had at least some information), but none had participated in this particular program (so no direct experience) and all received the same information in the message other than the quantification manipulation. However, we do see suggestive evidence that if a participant earned points in other programs, the effect of the quantified message is diminished. While the interaction is marginally significant ($p=.11$), the graph in Figure 3 demonstrates that the more points a participant had earned in the current quarter, the narrower the difference in likelihood of participation between the quantified and non-quantified conditions. This could occur because the \$125 goal is either more achievable or less important (for those who have already earned the maximum amount of points), or because more points is an indicator of more knowledge and therefore less dependence on the quantification target as a signal.

Study 1 provides strong evidence that quantification increases interest in a marketing message, but much more mixed evidence about the effectiveness of that message on actual enrollment and participation. While the field experiment setting allowed us to collect data on actual respondent behaviors, it made it impossible to measure the psychological mechanism underlying the behavior, or to control for self-selection. We also could not address whether they abandoned the program because of

dissatisfaction with their performance or whether they were just less interested in the program benefits once the initial excitement had worn off. In order to control for these factors, we ran a more controlled simulation in Study 2.

Figure 3: Plotted Probability of Participating Beyond Two Weeks Based on Prior Points Earned



Study 2

In order to better isolate the possibly psychological mechanism of anchoring and to eliminate the selection issues of the field experiment, we designed a more controlled study. Study 2 was conducted as a survey on Amazon’s Mechanical Turk, allowing us to measure again whether numbers in a program description affected the interest in enrolling, but also whether they led to anchoring on the specific number, and how they impacted people’s satisfaction with a given level of performance. All

participants completed all questions, so there was no selection bias affecting the results, and while we randomized the dollar amount mentioned, all were told to imagine that they had performed at the same level to control for differences in individual performance.

While we tried to closely mimic the field experiment setting, the survey design added key questions during the consideration process in order to understand how quantification impacted people's expectations. We included both high and low potential rewards to assess the importance of magnitude of the potential reward in attracting people's interest and their subsequent expectations. We also varied the language so that the potential reward mentioned was either a minimum ("or more") or a maximum ("up to"), in order to see how people adjust from the number given. In theory, a minimum amount should be more attractive than a maximum amount, as people would not be limited in the potential upside with a lower bound as they would be with an upper bound.

Method. We recruited 440 subjects via Amazon's Mechanical Turk to answer a brief survey. Of these 41 were eliminated from the results for giving responses that indicated they had not read the questions, leaving 399 for analysis. Participants were told to imagine they had received the following message about a wellness plan sponsored by their employer:

Join our new activity tracking program. Each day that you track at least 30 minutes of physical activity you can earn \$2 for participation.

Just below this text was show one of the following four quantification messages (randomly assigned):

Non-quantified: *Earn rewards for participation.*

Quantified 1 (high maximum): *Earn up to \$50 per month for participation.*

Quantified 2 (low maximum): *Earn up to \$20 per month for participation.*

Quantified 3 (low minimum): *Earn \$20 or more per month for participation.*

Participants were asked how likely they would be to join such a program (six point scale, very unlikely to very likely) and, assuming they enrolled, how much they would expect to earn for participating in the first month (free response). After answering those two questions, they were then told to imagine they received the following message: “Great Job! You have earned \$15 this month.” We chose \$15 to represent a participation rate of roughly 2 days per week, which was a low but very achievable level of participation based on what we had seen in the field experiment where almost half of participants (48%) had earned an average of three or fewer points per week.

After giving them the information about performance, we then asked questions to assess their satisfaction with their performance and reward:

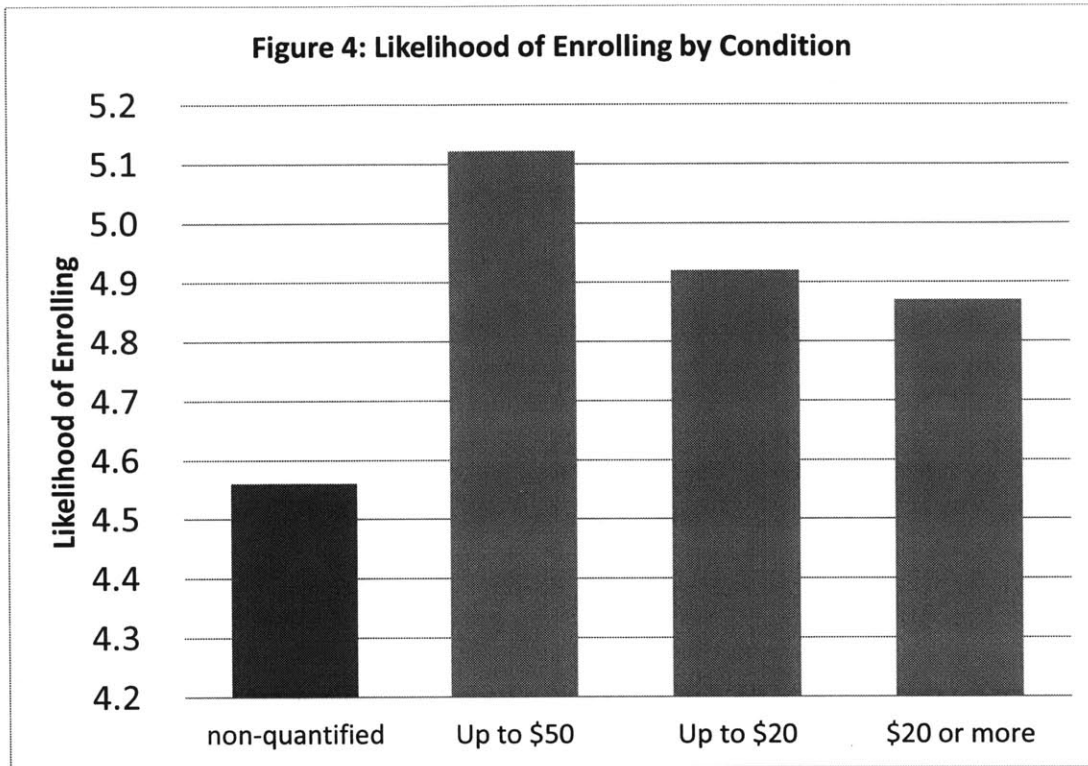
How satisfied are you with this performance?
(4 point scale, very dissatisfied to very satisfied)
How likely are you to continue with this program next month?
(5 point scale, very unlikely to very likely)

We expect that those with higher expectations will be more disappointed with this level of performance than those with lower expectations. And if, indeed, dissatisfaction causes people to be more likely to drop out of the program, as we think may have happened in the field experiment, those with higher expectations should indicate that they are less likely to continue with the program.

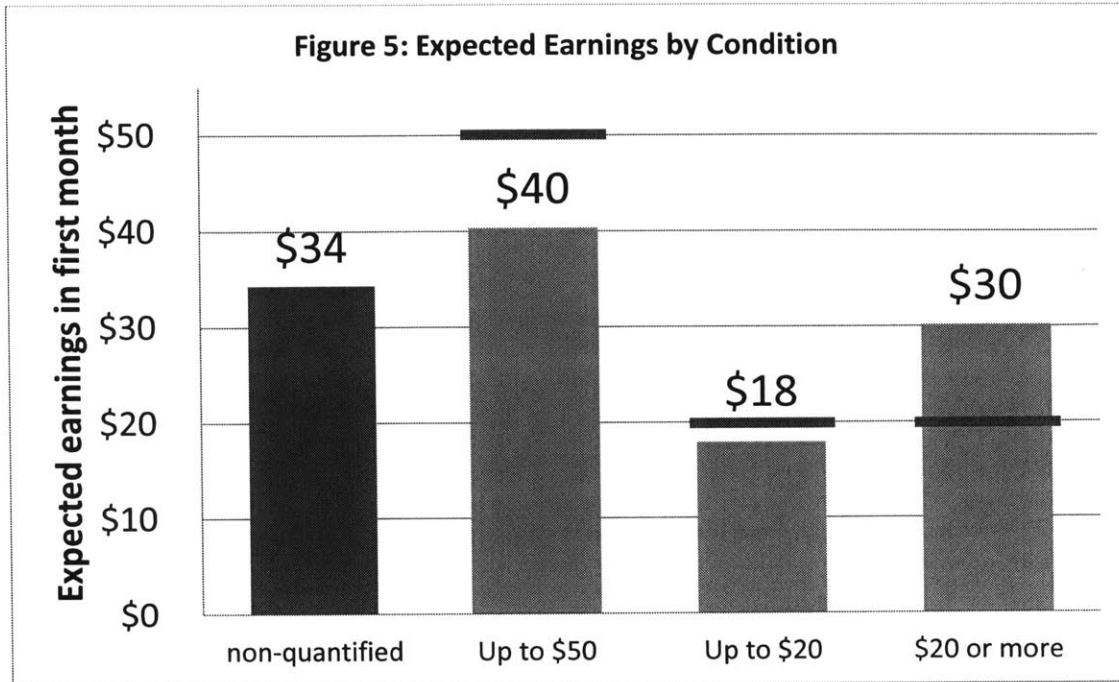
Participants were also asked if they had previously participated in a program like this. 11% of participants indicated they had, but they were roughly evenly distributed across conditions, and their responses were not significantly different from those who had not.

Results. People indicated they were more likely to enroll in all three quantified conditions compared to the non-quantified condition ($M=4.97$ vs $M=4.56$, $t=-2.65$,

p=.008), with higher potential rewards associated higher likelihood of enrollment (see Figure 4).



Respondents also appeared to anchor on the given number for their expectations of their own performance. Despite receiving the same information about how many points could be earned for each day of participation, the groups were significantly different from each other (ANOVA $R^2=.21$, $F=35.15$, $p<.001$) and no two groups significantly overlapped (see Figure 5 and Table 2 for means and confidence intervals). The mean expected earnings for the non-quantified group was higher than we expected. This appeared to result from a large peak of responses at \$60, which we believe resulted from a quick monthly potential earnings calculation (\$2 per day times ~30 days per month).



We also saw significant differences in satisfaction with \$15 of first month earnings, depending on how much people expected to earn (Correlation $r=-.46$, $p<.001$), indicating that the expected earnings did serve as an indicator of personal performance. This satisfaction also different by message assignment (ANOVA $F=10.7$, $p<.001$), with those who received the “earn up to \$50” message expressing the highest average expected earnings (\$40) and the lowest average satisfaction ($M=2.32$) and those who received the “earn up to \$20” message expressing the lowest average expected earnings (\$18) greatest satisfaction ($M=2.96$). (See Table 2 for regression results)

	Likelihood of enrolling	Expected earnings	Satisfaction with \$15	Likelihood of continuing
Up to \$50	.56**	5.97**	-.35**	-.02
Up to \$20	.37^	-16.48***	.29**	.26^
\$20 or more	.31^	-.422*	.01	-.09
Constant (base group is non-quantified)	4.56	34.38	2.66	3.98

^ p≤.1, *p≤.05, **≤p.01, ***p≤.001

Most participants said they would continue to participate in the following month (M=4.01 out of 5), making inference more difficult given the relatively small range of responses. There was a significant and positive correlation between satisfaction and the likelihood of continuing (correlation $r=.44$, $p<.001$), but no measurable direct relationship between expected earnings and desire to continue. There was, however, a positive relationship between expected earnings and desire to continue when satisfaction was also included as a control. This would seem to suggest that, for a given level of satisfaction, people are more likely to want to continue when their expected earnings are higher – a logical outcome. However, people were also somewhat more likely to say that they would continue in the “earn up to \$20” condition, which was the least economically appealing of all the conditions, thus suggesting that satisfaction may have played a greater role than potential remuneration (see Table 2).

When splitting the group between those who expected to earn more or less than \$20 (chosen for its proximity to the given level of \$15, and for the fact that it splits the population into roughly equal groups), there is a significant difference in the likelihood of continuing – those who expected to earn less than \$20 in the first month (M=4.13) were more likely to continue than those who expected to earn more than \$20 (M=3.90, $t=2.18$, $p=.03$).

Discussion. The MTurk results support the outcome seen in the field experiment, namely that higher quantified targets are more likely to motivate enrollment, but also more likely to cause dissatisfaction with a particular level of performance. All three quantified messages resulted in higher potential likelihood of enrollment, suggesting that the quantification of the potential outcome is more persuasive than just a general rewards message.

People also appear to anchor on the specific amount mentioned in the message, adjusting downward slightly for “up to” messages and upward slightly for “or more” messages. People also appear to look for any numerical clues they can find for setting goals, often taking the maximum potential outcome (\$2 per day times 30 days as the numbers provided in the text) rather than a more realistic expectation of performance when no other benchmark was provided.

These higher expectations for performance resulted in lower satisfaction with a moderate level of achievement. In contrast to economically rational behavior, which would predict that people should most want to continue in the group with the highest potential earnings (up to \$50 or \$20 or more), we saw that people in the group with the lowest potential earnings (capped at \$20) were most likely to say that they would participate in the following month, thus suggesting that satisfaction and a sense of achievement plays a substantial role in persistence.

GENERAL DISCUSSION

Summary and Managerial Implications

Numbers can indeed be helpful in attracting customer attention. In both our field experiment and MTurk experiment, the presence of a quantified outcome attracts

substantially more initial interest than the non-quantified version of the message. The MTurk experiment also demonstrates that, not surprisingly, a larger number generates more hypothetical interest than a smaller one does. However, while the large number may attract interest, this interest may be more short term in nature, as we saw a substantially smaller, indeed almost neutral, effect on actual enrollments in the field experiment.

One explanation for this neutral effect on enrollment is that the quantified message is attracting a different set of customers in place of those who might have joined from a non-quantified or less extrinsically focused message. The quantification not only sets up a target, which might deter potential participants who view it as unachievable, but it also sets a target (in this case) in dollar form, which puts greater emphasis on the monetary reward. This substitution effect could be costly if it is replacing intrinsically motivated customers with more extrinsically focused ones who are joining only for the potential reward.

We cannot rule out substitution, but we would make two comments about it. One is that it is necessarily occurring at the moment of click (or earlier) since all information and treatment from that point forward is the same. This emphasizes the importance of studying post choice behavior, since most tests would only see the positive impact at the point of choice not the more neutral or possibly negative effect which follows. Second, if customer substitution is occurring, this is also potentially problematic for marketers, as they are attracting a less intrinsically motivated customer who will require more costly reinforcement to continue to perform at the same level.

Whether because of the different audience or because of the shifted focus resulting from the quantified message, we see a negative impact on participation

beyond the initial enrollment on those who do join. Respondents who enrolled after seeing the quantified message were more likely to drop out after just one day of participation and were less likely to make it beyond two weeks of participation. This was true even though they participated at approximately the same frequency during those first two weeks. And even when there was no substitution, in Study 2, we still saw the same effect, thus suggesting that the message does have an impact on evaluations of performance. Participants thus appear to be differentially reacting to the same outcome based on which message they received.

One potential moderator for this effect might be their ingoing knowledge and expectations. When people lack ingoing information (or a strong signal), they should be more likely to look to infer information from what is given. We saw that respondents in the MTurk study who were not given a potential target were more likely to calculate from the numbers in the message than to calculate from their own experience. This is logical given that most of them had not participated in this type of wellness program before and so had minimal information about reasonable expectations and therefore looked to clues in the message itself.

In contrast, we found suggestive evidence from the field experiment that people with more information were less likely to respond to the marketing message. Participants in the quantified condition were somewhat less likely to drop out of the wellness program when they had earned more points elsewhere. This could be because their progress in this particular program seemed like less of a failure since they were adding it to points they had already earned elsewhere. Or it could be because the number in the marketing message was less salient for them – they had already earned points and therefore had a stronger ingoing signal.

While participant knowledge might weaken the signal from quantification, a highly credible sender might enhance it. Respondents will likely discount messages that seem overly grand in their promises, or which come from non-credible sources. In this case, both in the field experiment and in the hypothetical setting, the message comes from a sender who is affiliated with their employer and therefore brings substantial credibility.

This enhanced credibility is a double edged sword – while people pay more attention to the message, there is also more risk to getting the message wrong. If people are going to anchor on targets in the recruitment message, these targets must be set appropriately. For a goal of attracting attention, high numbers might be best, but when the goal is to attract ongoing participation, numbers should be carefully chosen. Broad targets, which affect a large number of heterogeneous people, could initially be set high in order to attract as many people as possible, or could be lower in order to maximize their likelihood of achievement and satisfaction.

Theoretical Contributions and Next Steps

We extend the literature on anchoring to demonstrate that people do indeed use externally provided anchors and that the effect of these anchors may endure beyond an initial choice decision. While we could not demonstrate in the field experiment that people were actually anchoring as a result of the number, we use the hypothetical setting to demonstrate that respondents do actually anchor on the number provided, and that this anchor is strong enough to impact their perceptions of satisfaction with the program and thus their likelihood of continuing to participate.

We also examine some factors which might affect how people actually set goals and expectations. Rather than a rational, carefully calculated process, actual goals are

often influenced by environmental clues. We expect a similar process may operate for product expectations, such that a printer cartridge which could print “up to 5,000 pages” would be expected to actually print something near 5,000 pages, or a hotel loyalty program offering “up to 10 free nights per year” will yield an expectation of earning something close to ten free nights, regardless of the difficulty of actually performing at that level.

This opens up broader questions of what factors might impact how people set goals and targets in real-world settings. We demonstrate here that the presence of numbers may be important, but there are other potential cues which might have similar effects. The presence and salience of disclaimers might change the effectiveness of the message, but also the strength of the anchor. Pictures might create more concrete expectations as well and might interact with numbers to create stronger expectations. A picture of a hotel room with a beautiful view of the ocean could lead visitors to be disappointed when their actual room faces the parking lot.

While we did not define boundary conditions in this particular paper, we do have some hypotheses about when they might occur. In addition to disclaimers, providing additional (salient) information might reduce the salience of the anchor and/or increase the likelihood that people adjust away from that initial target. As we saw in the field experiment, personal experience and outside information could also reduce the importance of the target mentioned by the marketer.

We also believe that knowledge about progress is an important element of this particular situation. The attractiveness of the \$125 target should be diminished by the realization that it is difficult, or potentially impossible, to achieve, at which point a small amount of progress starts to seem more like a failure relative to that high target. The realization of difficulty only comes with knowledge about one’s own progress to the

goal and could be affected by the way in which that progress is presented. If one sees the total goal when progress is presented (e.g. a bar filling up to \$125), it could seem more like a loss than if one simply sees the amount earned to date, in which case the total goal might be less salient.

In addition to how progress is presented, one could also change how the goal is initially presented. Our manipulations set out a maximum potential target for a month or quarter, but the effect might change if the marketer set attractive subgoals. Rather than saying “earn up to \$125 per quarter,” one could instead say “earn up to \$1 for each 30 minutes of activity.” While these are both descriptions of the same program, the latter emphasizes the individual unit of activity whereas the former emphasizes the total potential outcome. Focusing attention on the individual unit of activity might attract lower initial interest because the potential earnings appear smaller, but it might create a greater feeling of accomplishment as people build up earnings.

We believe that this paper sets up an interesting effect that brings together the research on anchoring with the research on goal setting and accomplishment. The knowledge that people do use numbers from the environment to set their goals also opens up more lines of inquiry around what other types of information could impact goal setting. We also demonstrate that information from a marketing message has enduring effects on participation well beyond the initial click response and believe that the issue of differing short term and long term effects is an interesting one with many more avenues of exploration.

REFERENCES

- Anderson, Rolph E (1973), "Consumer Dissatisfaction: The Effect of Disconfirmed Expectancy on Perceived Product Performance," *Journal of Marketing Research*, 10 (1), 38–44.
- Augustine, Norman Ralph (1997), *Augustine's Laws*, AIAA.
- Bandura, Albert (1977), "Self-efficacy: toward a unifying theory of behavioral change.," *Psychological review*, 84 (2), 191–215.
- Baumeister, R F, T F Heatherton, and D M Tice (1993), "When ego threats lead to self-regulation failure: negative consequences of high self-esteem.," *Journal of personality and social psychology*, 64 (1), 141–56.
- Cardozo, Richard N (1965), "An Experimental Study of Customer Effort , Expectation , and Satisfaction," II, 244–49.
- Epley, Nicholas and Thomas Gilovich (2006), "The Anchoring-and-Adjustment Heuristic: Why the Adjustments Are Insufficient," *Psychological science*, 17 (4), 311–18.
- Erez, Miriam and Isaac Zidon (1984), "Effect of Goal Acceptance on the Relationship of Goal Difficulty to Performance," *Journal of Applied Psychology*, 69 (1), 69–78.
- Forrest, David, Robert Simmons, and Neil Chesters (2002), "Buying a Dream: Alternative Models of Demand for Lotto," *Economic Inquiry*, 40 (3), 485–96.
- Frederick, Shane, Daniel Mochon, and Jennifer Savary (2014), "The Role of Inference in Anchoring Effects."
- Heath, Chip, R P Larrick, and G Wu (1999), "Goals as reference points.," *Cognitive psychology*, 38 (1), 79–109.
- Jain, Sanjay (2009), "Self-Control and Optimal Goals: A Theoretical Analysis," *Marketing Science*, 28 (6), 1027–45.
- Koo, Minjung and Ayelet Fishbach (2008), "Dynamics of self-regulation: How (un)accomplished goal actions affect motivation.," *Journal of personality and social psychology*, 94 (2), 183–95.
- Lichtenstein, Sarah and Paul Slovic (1971), "Reversals of Preference Between Bids and Choices in Gambling Decisions.," *Journal of Experimental Psychology*, 89 (1), 46–55.
- and ——— (1973), "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas.," *Journal of Experimental Psychology*, 101 (1), 16–20.

- Locke, Edwin A and Gary P Latham (1990), *A theory of goal setting & task performance.*, Prentice-Hall, Inc.
- , ———, and Miriam Erez (1988), “The Determinants of Goal Commitment,” *The Academy of Management Review*, 13 (1), 23.
- , Karyll N Shaw, Lise M Saari, and Gary P Latham (1981), “Goal Setting and Task Performance: 1969–1980,” *Psychological Bulletin*, 90 (1), 125–52.
- MacKenzie, Scott B (1986), “The Role of Attention in Mediating the Effect of Advertising on Attribute Importance,” *Journal of Consumer Research*, 13, 174–95.
- Mossholder, Kevin W. (1980), “Effects of externally mediated goal setting on intrinsic motivation: A laboratory experiment.,” *Journal of Applied Psychology*, 65 (2), 202–10.
- Nisbett, Richard E and Lee Ross (1980), *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Olshavsky, Richard W and John A Miller (1972), “Consumer Expectations, Product Performance, and Perceived Product Quality,” *Journal of Marketing Research*, 9 (February), 19–21.
- Prelec, Drazen, Birger Wernerfelt, and Florian Zettelmeyer (1997), “The Role of Inference in Context Effects: Inferring What You Want from What Is Available,” *Journal of Consumer Research*, 24 (1), 118–26.
- Simmons, Joseph P, Robyn a LeBoeuf, and Leif D Nelson (2010), “The effect of accuracy motivation on anchoring and adjustment: do people adjust from provided anchors?,” *Journal of personality and social psychology*, 99 (6), 917–32.
- Soman, Dilip and Amar Cheema (2004), “When Goals Are Counterproductive: The Effects of Violation of a Behavioral Goal on Subsequent Performance,” *Journal of Consumer Research*, 31 (1), 52–62.
- Tversky, Amos and Daniel Kahneman (1974), “Judgment under Uncertainty: Heuristics and Biases,” *Science (New York, N.Y.)*, 185 (4157), 1124–31.
- Yalch, Richard F and Rebecca Elmore-Yalch (1984), “The Effect of Numbers on the Route to Persuasion,” *Journal of Consumer Research*, 11, 522–27.
- Zhang, Ying and Szu-Chi Huang (2010), “How Endowed versus Earned Progress Affects Consumer Goal Commitment and Motivation,” *Journal of Consumer Research*, 37 (4), 641–54.

APPENDIX

Text common to all participants:

The Boost® Physical Activity Program encourages participants to make healthy behaviors part of daily life. By tracking your physical activities—from conditioning exercises to dance, and even home activities—you'll see first-hand just how easy it is to maintain a healthy lifestyle.

As a Boost participant, you'll challenge yourself to achieve a minimum of 30 minutes of daily physical activity. Once enrolled, you can start tracking up to 200 different physical activities, while assessing your progress on the path to better health! Plus, you can earn points towards Quarterly Health Incentive Credits*, which reduces your medical plan contributions.

Quantified A:

Yes, I would like to participate in the Boost tracking program. I want to earn points toward rewards of up to \$125 per quarter and I understand that the more often I track my activities, the more points I will earn.

No, I do not want to participate in the Boost tracking program.

Quantified B:

Yes, I would like to participate in the Boost tracking program. I want to earn points toward rewards of up to \$125 per quarter and I understand that the more often I track my activities, the more points I will earn.

No, I do not want to participate in the Boost tracking program because earning points toward rewards of up to \$125 per quarter is not important to me at this time

Non-quantified A:

Yes, I would like to participate in the Boost tracking program. The more often I track my activities, the more points I will earn towards the incentive rewards program.

No, I do not want to participate in the Boost tracking program.

Non-quantified B:

Yes, I would like to participate in the Boost tracking program. The more often I track my activities, the more points I will earn towards the incentive rewards program.

No, I do not want to participate in the Boost tracking program because earning points toward the incentive rewards program is not important to me at this time.

ESSAY 3

A Picture is Worth a Thousand Words: Photographs as Anchors for Expectations

Abstract:

Marketers often use images to increase the vividness of their communications to customers. These images should make the message easier to process, thereby increasing liking for the product and certainty of expectations. In a series of experiments, I demonstrate that images do indeed increase certainty of preference estimates, both within and between respondents, but may have more mixed effects on the valence estimates of preference. I also begin to explore how these estimates might impact evaluations of actual products and propose some factors which might impact this evaluation for future exploration.

Keywords: images, expectations, preferences, certainty, satisfaction

“A picture is worth a thousand words.”

“Results not typical.”

Marketers regularly use photographs to convey a thousand words. Imagery is more vivid, easier to process, and, in many cases, more concrete than textual descriptions. One could try to describe a red dress using only words, but a Google image search for “red dress” yields hundreds of different waistlines, lengths, fabrics, necklines, not to mention actual shades of red. In this case, the picture is indeed worth a thousand words, and the consumer is going to be able to predict how much she likes the dress much more accurately from an image than from a text description.

But what happens when you try to get people to put together a mental representation of a product based on descriptions of different attributes. Potential customers often read a product description or look through a series of pictures describing different attributes of the product, whether a gallery of hotel images or a virtual walkthrough of a home for sale, or even a description of a piece of jewelry. Customers put these images and descriptions together in unique ways to form individual mental representations of the product. One could expect that the customer’s mental representation of the product would be more precise if he saw images, but that it might also be more strongly dependent on the actual images used to form that impression.

There is also a risk, when describing a product with a wide variety of potential versions, that using an image narrows expectations to one particular version of that product. For example, when viewing pictures of a potential home online, one might view carpet in the photograph and not realize there are hardwood floors underneath because the photo represented such a strong anchor. For someone with a strong

preference for hardwood floors, even the nicest looking carpet is going to negatively impact his estimate of preference for the home. A large photograph of a chocolate muffin from a bakery might cause one to discount the idea that there are healthy options available, even when they mention it in the text. The strong mental image created by the photograph might make other potential options less salient, which is helpful when the photograph matches the user's preference, but potentially problematic when preferences are very heterogeneous.

In this paper, I investigate whether images generate more precise expectations when evaluating products, both through a survey based online experiments and through a choice based conjoint evaluation. I also begin to investigate whether the images can cause people opinions to be biased in their initial impressions. I test this effect of images in a category (smartwatches) where the appearance of the product is an important element of its overall valuation and where the relative novelty of the product category might lead users to be more uncertain of their preferences before measurement.

BACKGROUND

Images and Information Processing

There is a substantial amount of evidence that people process images differently than text. Images are often easier to understand, and are therefore processed more fluently, which can increase liking and positive judgments (Reber, Winkielman, and Schwarz 1998). An array of images can increase perceived variety, but also perceived complexity (Townsend and Kahn 2014), perhaps because images are processed more holistically and with broader functionality (Sharps and Nunes 2002). But perceived

complexity could be reduced given that images are processed more quickly and more automatically with a closer connection to meaning (Townsend and Kahn 2014). Thus it seems likely that while images are processed differently, the specifics of the context and motivation for processing may lead to different outcomes.

An additional important distinction for images versus text is in how people deal with missing information. Consumers often make judgments based on limited information because marketers omit to include unfavorable information, which consumers may try to correct if they recognize it is missing (Kamins et al. 2009). The more “gestalt” nature of image processing may make missing information less salient and lead to more intransitive preferences as a result of customers focusing only on what is observed (Kivetz and Simonson 2000). In addition, abstract representations and more vague textual descriptions often omit irrelevant or inconsistent information which can lead to simpler and more schematic representations of a potential outcome (Trope and Liberman 2010) and more heterogeneity of assumed information.

Images might also be perceived as more representative than text, which can also lead to weighting biases (Kahneman and Frederick 2002). By offering a photo (even if not exact), the marketer is offering an ostensibly similar rendering of the attribute which can be mapped to expectations. Text, in contrast, requires the additional step of finding a representative attribute in memory, introducing more variability into the judgment. Images generate a stronger impression of resemblance to the attribute under consideration, thus increasing the likelihood that it will be considered representative (Tversky and Kahneman 1974), but which may introduce errors if the image is meant to be illustrative rather than exact.

Finally, images are likely to be processed more emotionally than text, which can result in greater preference consistency and less cognitive noise (Lee, Amir, and Ariely

2009), but this assumes that the images convey additional information compared to text. One possibility is that the images convey additional emotion, but that this emotion may not actually be informative. Photographs may also be seen as more concrete, which can lead people to be more critical when evaluating. A high level (abstract) construal tends to involve the value of end state whereas feasibility dominates for low level (concrete) construal. They might be more inclined to picture themselves actually using the product being depicted, whereas a more abstract textual description might lead to a more general consideration of the product potential.

Images, Precision, and Accuracy

Within images, it may be important to distinguish between photographs versus more abstract drawings or artistic renderings of an image. Photographic images have traditionally been taken as proof that the scene did occur as pictured (Messaris 1997) – even now when we see an image that seems like it must have been photoshopped, we start from the assumption that it is true and often pay little attention to disclaimers (Ata, Thompson, and Small 2013). In fact, given that disclosures are often used to avoid censure even when marketers are deliberately blurring the lines of truth and staging, they are often in small print or designed in other ways to minimize the potential impact. Drawn or artistic versions of an image do not convey the same assumption of reality, so while the norms of image processing may apply, both the concreteness and the reality of the image are more flexible.

In general one would expect greater precision to convey benefits. It is seen as a more accurate representation, and enhances the ability of the customer to both predict his liking for the product and imagine himself using it. However, this makes the very important assumption of image accuracy. Marketers often have an incentive

to use the best possible version of their product in communications to potential customers, leading to a blurring of what might be considered accurate. Since expected utility and purchasing decisions are made based on expectations about product quality (Goering 1985) so inaccurate information may lead to inefficient purchasing decisions.

In addition, while evaluations of performance tend to assimilate toward manipulated expectations of quality when the evaluation is close to the expectation (Olshavsky and Miller 1972), large deviations from expected quality can lead to contrast effects and magnify the distance relative to ingoing expectations (Cardozo 1965). So the more vivid and more positive the initial image is, the more risk that a customer sets a strong expectation and could be dissatisfied with the final product, particularly if this image is of an exemplar rather than the average product, or of one particular version of a product when many are available.

This idea of multiple versions of a product also introduces the idea of matching images to heterogeneous customer preferences. While advertisers will generally try to target their content towards a particular segment, good targeting can be challenging, particularly for newer products or new customers. The same concreteness of a specific image which raises its persuasive appeal might also reduce its persuasion if not well matched to initial preferences. Text leads people to generate mental images, thereby increasing the heterogeneity of responses based on those different mental representations. While this can be problematic when using responses for estimation of preferences, it might be an effective tool for a marketer who wants to appeal as broadly as possible. There appears to be a potential tradeoff between precision and accuracy when either the product and/or the audience preferences are heterogeneous.

In order to analyze this tradeoff, I will measure several different text versus image comparisons in survey based experiments and then use a conjoint analysis to estimate whether images help to discriminate between alternatives at the moment of choice, as measured by the magnitude of the partworth. This has been used as a technique for measuring for the effects of brand credibility (Swait and Erdem 2007) and I will use it here to measure the effects of images on choice certainty. In addition to measuring the estimated certainty of the preference, I will also compare those estimates to later product evaluations, in order to test whether the more precise estimates of preference are, in fact, also more accurate and whether this might depend on the type of image used (actual product or generic image or drawing).

SURVEY EXPERIMENTS

As a preliminary test of the effect of images of estimates of preference, I ran a series of survey based experiments on Amazon's Mechanical Turk asking participants to choose and value various product attributes and then asking them how much they liked the product that matched their selections. Both experiments used the same products (smartwatches) and the same series of attributes (face shape, bezel color, strap material plus sync capability which was not part of the test but used to lend authenticity to the smartwatch evaluation and was text for all participants and evaluated last), but varied the images used as stimuli. Both experiments then also presented respondents with an image of their chosen product and asked them to rate willingness to pay, liking for the product and how well the product matched their expectations.

Based on the different processing styles for images compared to text, I expected two main differences in how people process images of attributes compared to text descriptions of those attributes. The first is that people who see images will have an easier time processing those images and rely on more emotional processes, thus leading to increased certainty in their estimates of preference. The second is that people who see photographs will view those images as representative examples of the product attribute, and are likely to pay little attention to disclaimers of potential dissimilarity in their evaluation. This latter effect implies that those who see generic photographs (with disclaimers) should feel the same degree of certainty as those who see photographs of the actual attributes of the final product (and thus have no disclaimer), but that any photograph should lead to more certainty in estimates than a drawing. People who see generic images might, however, be less satisfied with the actual product given that their estimates of preference were based on less accurate images.

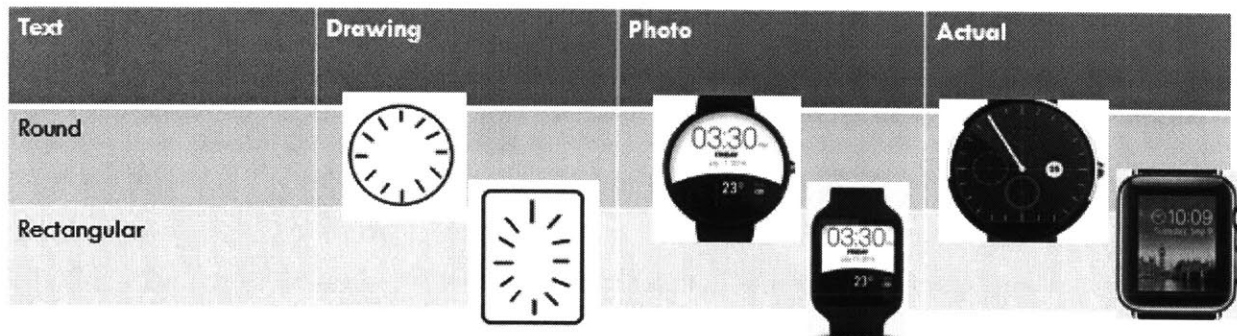
Method

366 participants were recruited on Amazon's Mechanical Turk. All participants were given a description of a smartwatch. Then they were presented with a series of attributes and asked to choose between two different options, estimate how much extra (above the base price) they would be willing to pay for their selection and whether they would accept the alternative if their choice was not available. All participants were randomly assigned at the first attribute to either the text condition or one of the image conditions and had the same format for other attributes (between subjects manipulation).

After making their choices, they were then asked to confirm their choices and presented with a watch that matched their selection. They were asked how much they would be willing to pay for this watch, how much they liked it (1-10 scale) and how closely they felt it matched their expectations (1-10 scale). Finally participants were asked their age and gender, whether they currently owned a smartwatch and how interested they were in smartwatches before the survey.

Experiment 1

Experiment 1 (N=181, 58% male, mean age 33) tested a text version in comparison to a drawing and a generic photograph of the attribute, both of which included a disclaimer that this was not an image of the actual product being evaluated. One would expect that the drawing should increase respondents' ability to visualize the product compared to text, although the more general nature of the drawing should also lead to less specific expectations. Examples of the manipulation are included in Figure 1.



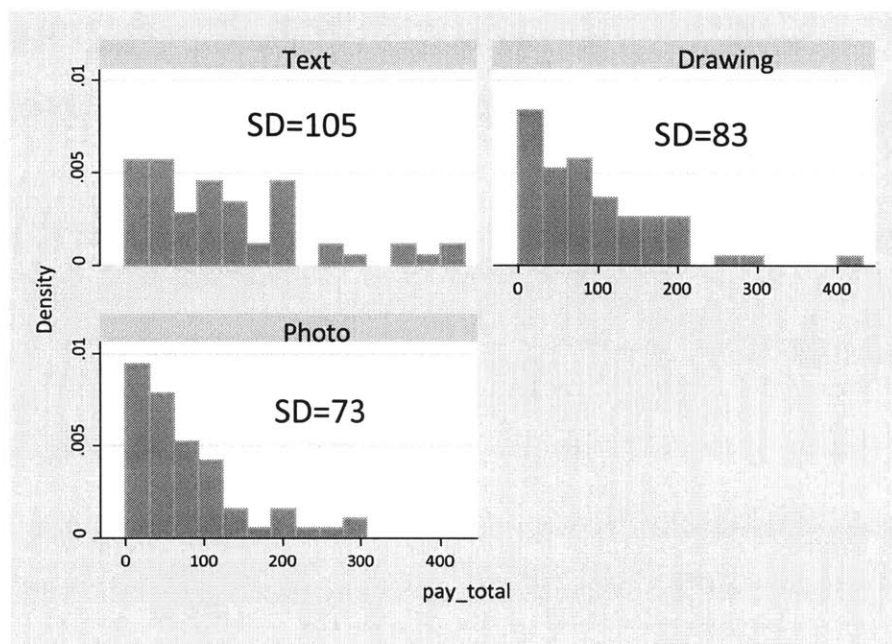
Willingness to pay for each attribute was summed together to create a total willingness to pay value. In general participants who saw the text manipulation were willing to pay more (\$129) than participants who saw either of the image manipulations (\$92 for the drawing and \$78 for the generic photo, combined mean

\$85, $t=-3.1$, $p=.002$, see Table 1 for means and standard deviations). In addition, while the drawing and photograph manipulations did not differ in their willingness to pay, both separately were significantly smaller than the text manipulation.

Perhaps more importantly than the absolute difference in the willingness to pay was that the variance of the willingness to pay was significantly smaller in the image manipulations than in the text manipulation (Variance ratio test, $F=1.7$, $p=.02$). As with the absolute willingness to pay, the variances for both image manipulations separately ($SD=83$ for drawing and $SD=73$ for the photo) were smaller than the variance for the text manipulation ($SD=129$), although this effect was greater for the photograph ($p=.006$) than for the drawing ($p=.07$) suggesting that the photograph did lead to more certainty than the drawing.

In addition, the Kolmogorov-Smirnov test for equality of distributions shows a significant difference in the distributions for willingness to pay in the text condition compared to the others (Combined K-S $D=.23$, corrected $p=.02$). As one can see in Figure 2, the distribution is substantially smaller and right skewed in the image manipulations compared to the text manipulation.

Figure 2: Distribution of Total Willingness to Pay for Attributes in Experiment 1



This difference in certainty of estimates did not translate into a difference in evaluations of the actual product, with no significant differences between estimates of liking for the actual product (“how much did you like the watch you were shown?”). The attribute based estimates of willingness to pay were substantially higher than the final product estimates in all conditions, but declined more in the text condition than in the images conditions. This suggests that the lower willingness to pay from images might have been more accurate as well as more precise, although this difference is difficult to evaluate since both conditions substantially overestimated willingness to pay from the attribute estimates. There was no significant difference in how closely they felt the actual watch matched their expectations or how much they liked the actual watch.

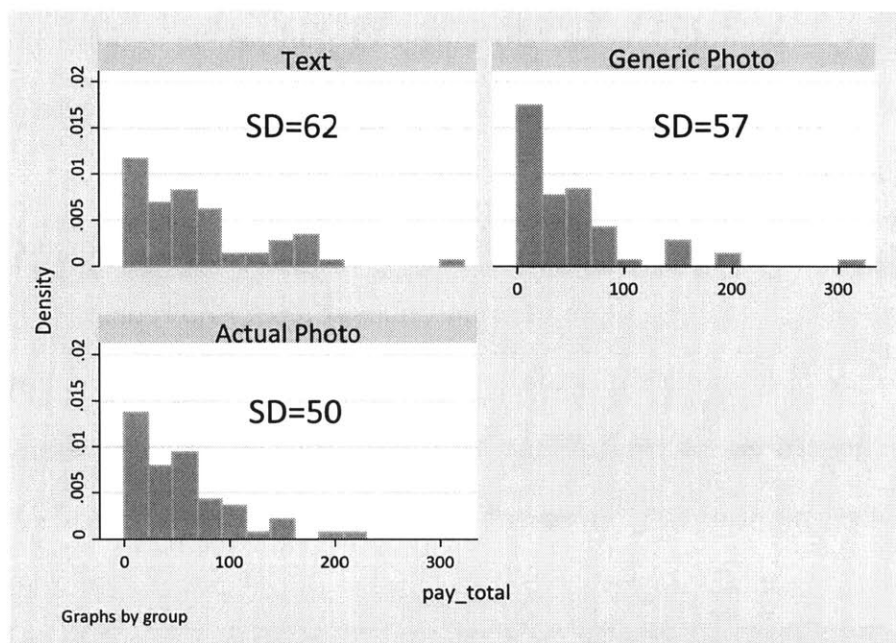
Table 1	Experiment 1 WTP for attributes		Experiment 2 WTP for attributes	
	Mean	Standard Deviation	Mean	Standard Deviation
Text	128.7	105.3	67.8	62.4
Drawing	92.3	83.1		
Generic Photo	77.7	73.1	51.2	56.7
Actual Photo			52.8	50.2

Experiment 2

Experiment 2 (N=185, 61% male, average age 34) was almost identical to Experiment 1, except that a photograph of the actual attributes from the final product was used instead of a drawing (the text, generic photo, and actual photo conditions from Figure 1). Given that people generally don't pay much attention to disclaimers, one would expect the generic photograph to perform similarly to the actual product photograph in estimates of preference, but that they might be more disappointed in the actual product shown since it is less likely to match their expectations.

We again saw the exact same pattern of lower willingness to pay for the combined attributes in the images conditions, but also a lower standard deviation of the estimates for images compared to text, but in this particular test the differences were only marginally significant (although the same direction). There were no significant or noticeable differences between the two image conditions in terms of willingness to pay or the distribution of willingness to pay. Respondents were willing to pay \$68 in the text condition versus \$52 in the image condition ($t=-1.8$, $p=.07$). Similarly, the standard deviation of the estimate was 62 in the text conditions versus 53 in the image condition ($f=.73$, $p=.15$). The Kolmogorov-Smirnov test of equality of distributions again suggests that the distributions of the groups are different (combined K-S, $D=.18$, corrected $p=.09$), but again marginally significant in this experiment (Figure 3).

Figure 3: Distribution of Total Willingness to Pay for Attributes in Experiment 2



We also see generally no difference in respondents' ratings of the actual watch with the exception of the liking measure. Respondents who saw the actual photograph rated their liking for the final product (mean liking 6.6) as significantly lower than those who saw text (mean liking 7.5, $t=-2.16$, $p=.03$). This effect remains marginally significant ($t=-1.63$, $p=.10$) when controlling for how well they felt the final product matched their expectations. While this one result is not enough to use as the basis for any conclusions for their preferences, it may be worthy of further exploration as it is the only time when respondents see attribute photos that are of equal or higher quality to the final product.

Discussion

Across two experiments, I demonstrate that photographs generate decreased heterogeneity in estimates of preference, but I see no significant effect on accuracy of the estimates as a predictor for their evaluations of actual products. Showing photographs results in a lower variance in estimates of willingness to pay compared to

text or drawings, but there is no difference in the variance between actual product images and generic product images. This suggests that people may be paying little attention to the disclaimer and instead relying on the generic image the same way they would consider an image of the actual product.

While the wider range of estimates of willingness to pay in the text condition could imply a greater heterogeneity of preferences, there is no reason to think that that people in the text condition actually had a wider range of preferences (across both experiments) given random assignment. The differences in the preference estimates should thus be a result of the manipulations rather than any difference in actual preferences. It is possible, however, that the lower variance in the photograph condition was a result of estimates being truncated by the low end at zero and that the lower variance was a result of the lower willingness to pay in the photo condition.

A lower willingness to pay could be because concrete photos were evaluated more critically than abstract (text) descriptions. But it could also be because people's preferences narrow on a particular version of the product rather than the potentially more heterogeneous mental representations in the text condition. In situations where only one version of the product can be shown, respondents who saw images might have a difficult time mentally representing other versions, or combining different attribute photographs together into a final product.

These differences in initial preference estimates also did not result in any consistent and significant differences in evaluation of the final product. The only difference in liking of the final product was when I used actual product images for the initial ratings, which was also the highest quality images of the attributes depicted in all conditions. The more modest generic versions of attributes used in the other conditions might have made the final products look more positive in contrast.

I did not test the use of an exemplar (high quality) image compared to a more modest actual product in this particular manipulation. Given the relatively high quality of the actual watches chosen as products (the Moto 360 and Apple Watch were used as final product images), it is not surprising that people would express a relatively high level of liking for the actual product displayed. An interesting potential follow up test would be to look at high quality images with more modest final products. This could cause users who book based on the description from a hotel website (and see a beautiful photograph of a spacious hotel room with a skyline view) to have substantially higher expectations and lower satisfaction for the same room compared to ones that look at TripAdvisor user pictures.

CONJOINT EXPERIMENT

In order to more accurately measure whether images help customers to discriminate between alternatives, I also used a choice based conjoint design to elicit estimates of relative utilities for attributes which were scaled by their consistency in choices. I then combined these attribute utility scores into a predicted estimate for a particular product and compared that predicted estimate to participants rank ordered preferences for different versions of the product (different combinations of attribute levels) and their ratings for those same product versions. The advantage of the conjoint is that it allows me to estimate certainty rather than just liking.

The images for attributes were almost identical to those used in the previous experiments, with the exception of the bezel color, which was changed to an even more generic version of the product in order to avoid overly specific attributes.

Method

I used Sawtooth Software's web based Discovery tool, again with a population drawn from Amazon's Mechanical Turk. Respondents were randomly assigned to either a text or a generic photograph condition. I chose not to use the actual product images to avoid using products which were clearly recognizable as either Apple watches or a different branded product, although I did use these images for the final product evaluations.

After being asked to rate each of the attributes (shape, bezel finish, strap material) and levels, respondents were given price level information (\$=250, \$\$=300, \$\$\$=350)³ before being asked to choose which product they preferred among three choices given. They repeated this choice seven times with different attribute combinations, from which Discovery calculates both a partworth for each attribute level and a rescaled utility score for each attribute level which is scaled to show relative importance between attributes. These seven choices, generated fourteen inequality constraints which were used to measure six independent partworths. These estimated partworths are based on the empirical Bayes utilities which assigns a lower utility for higher error respondents. While seven choices for six partworths may lead to relatively high standard errors for individual estimates, it was enough to allow us to test differences between the estimates of errors (certainty) across conditions.

After completing the choice based conjoint, respondents were shown photographs of four actual watches, mixing attributes so that all attributes and levels appeared at least once in a level balanced design. They were asked to rank these watches in order of preference. They were then asked to rate how much they liked

³ I chose the dollar signs as a commonly accepted indicator of relative price and to avoid introducing text into the images condition. Respondents in the images condition saw a clipart version of the dollar sign compared to a normal typeface version in the text condition.

each of the watches individually in order to evaluate whether their estimates of preference for the attributes corresponded to their actual preferences for the product. Finally they were asked which operating system they currently had and whether they had a prior interest in smartwatches.

Results and Discussion

I administered the survey to an initial sample of 205 respondents. Of these I eliminated all of those who completed the survey in less than two minutes, which represented approximately 10% of the sample (final N=183). Of these only three indicated that they already owned a smartwatch, and another twenty indicated that they were very interested in the category. The rest (88% of the sample) were moderately or not at all interested before the survey and therefore had a fairly low ingoing level of knowledge about their preferences and the attributes.

Discover uses a maximum likelihood model via individual level logit to estimate the individual utility partworths for each attribute level. For each attribute, the estimates are scaled by the error term of the estimates, resulting in larger values for the estimates of the partworths when choices are more consistent (lower error term). The attributes are rescaled such that the levels of each attribute sum to zero, such that the distance between the maximum and minimum value within the attribute provides an estimate of the estimated utility scaled by certainty, with a larger distance representing more certainty in the estimate. For each user and attribute, I calculated the maximum distance within an attribute and then added these together to form an importance score for the attributes in total.

In general, the images result in higher importance scores (more certainty) for the attributes, with a mean importance score of 5.3 for images versus 4.8 for the text

condition ($t=-2.16$, $p=.03$). This importance score difference was significant for the visual attributes (3.8 versus 3.3, $t=-2.11$, $p=.03$), but there was no difference in the scores for price (1.58 for images versus 1.50 for text, $t=-.56$, $p=.57$).

For each watch that users were asked to rate, I calculated an estimated utility (excluding price, which was not part of the actual watch evaluation as it was not a visual attribute and should not have been impacted from the manner of depiction). Within each user, a higher estimated utility should represent a stronger preference for that particular watch, but differences in ratings cannot easily be compared across users. I therefore use that estimated utility to calculate the intra-user correlations for rating and rank to measure whether the higher (more precise) estimates of preference can translate to more accurate estimates of their preference for the actual products that were shown. Overall, I find that the higher utilities from the images condition do not translate to higher ratings for the actual watches and the utility estimates from the images condition did no better at predicting ratings than those from text, but I cannot rule out the possibility that brand or operating system information in the final product evaluations did not confound their evaluations.

Table 2		Text	Images	T-test
Intra User correlation	Actual Ratings to Actual Rank (opposite scales)	-.05	.01	$t=-1.9$, $p=.06$
	Predicted Ratings to Actual Ratings	.42	.44	$t=-.52$, $p=.61$
	Predicted Rank to Actual Rank	.19	.10	$t=2.3$, $p=.02$
	Hit Rate, #1 ranking	34%	32%	$z=.31$, $p=.75$

I also looked at the hit rate for predicting the top ranked watch. In general, the predicted utilities did slightly worse at predicting the top ranked watch than a naïve model of always predicting the watch with the highest share. The degree of misprediction for the top ranked watch was slightly higher in the images condition than in the text condition, but this difference was not significant. It seems possible that this could be because the images condition leads people to overstate the predicted utility, but this could also be confounded by brand or operating system preferences. The predicted utility for the images condition suggested that 43% of people would choose Watch #3 as the top pick whereas only 33% of people actually did so.

GENERAL DISCUSSION

Summary & Managerial Implications

Overall, the presence of images led to more precise estimates of preference compared to the text descriptions alone. Across two survey based experiments and a choice based conjoint design, the use of images resulted in greater consistency of estimates within users (conjoint) and smaller range of estimates across users (survey experiments). In the survey experiments, the images conditions also resulted in lower estimates of willingness to pay compared to the text condition.

One potential explanation for both the wider range of estimates and the higher willingness to pay is that text descriptions resulted in more heterogeneous (and likely less accurate) mental representations of the product compared to the more concrete images. These mental images might call to mind previous products that the user has seen or related products, but without capturing these mental representations as part of the measurement process, it is impossible to know what product people had in

mind. It is also possible that participants in the text condition paid somewhat less attention because the words required more effort in order to process. However lower attention would generally result in lower effort and indicating a higher willingness to pay required moving the slider bar more than the participants in the images condition – arguably more effort rather than less.

This higher willingness to pay is worthy of further exploration. The images are consistently leading to more accuracy, but they may also be related to lower initial positive feelings for the product instead of reduced heterogeneity of estimates of preference. If a marketer is interested in casting a wide net, the images might actually lower initial interest if they do not match the initial expectations of potential customers. Given that first impressions can be sticky, showing the wrong picture might deter a potential customer from being interested in the actual product.

If images increase prediction certainty, but not prediction accuracy, then these images can lead to biases, both on the part of consumers in terms of their choices, but also on the part of marketers as they try to determine customer preferences. Across all three studies in this paper, I demonstrate that images lead to smaller variation in estimates of preference, both across subjects in the first experiments, and within subjects as show in the conjoint study. However there were no statistically significant differences in accuracy so further work is required in this area before any interpretation can be made.

Limitations & Next Steps

While the experimental designs here showed potentially more precision in the estimates of preference, there was not enough information to make conclusions about accuracy. The lack of predictive accuracy is likely due to limitations in the design and

measurement. The images condition in general resulted in lower willingness to pay estimates (first two experiments), but not lower evaluations of the actual products. This could be because there is no difference in their actual willingness to pay, but both conditions expressed a lower willingness to pay for the final product than they did for the sum of the attributes, which limits the potential conclusions. One cannot really determine whether the images condition underestimated willingness to pay or the text condition overestimated since both conditions were inaccurate.

Another potential reason for this null result is that I generally used attribute images which were arguably less attractive than the final products, which should lead people to feel more positively about the final product than they do about the attributes. Marketers often do the opposite in practice – using images that are more attractive than the actual product being consumed. If images cause people to be more precise in their estimates, but there are estimating preferences based on the wrong product, it seems possible that they will be more disappointed than someone with a more general estimate.

It is also interesting, that the images condition over-predicted the number of people choosing the highest rated watch. This suggests that perhaps the increased certainty in the images estimates may have been a sign of overconfidence. It is worth considering further tests to determine whether the use of images consistently over-predicts share of the most popular option. It might also be interesting to ask people directly how certain they are of their choices or whether they feel additional information might be more useful.

In general, more work is needed regarding the accuracy and external validity of the predicted choices. Among the many factors which could be influencing people's choices are concreteness of the images, use of exemplars versus average depictions,

prior product knowledge, heterogeneity of the choice set, and importance of the appearance of the product to overall evaluation. Each of these factors might have different effects on the quality of predictions. Precision is important and desirable, but only if accompanied by accuracy.

REFERENCES

- Ata, Rheanna N., J. Kevin Thompson, and Brent J. Small (2013), "Effects of exposure to thin-ideal media images on body dissatisfaction: Testing the inclusion of a disclaimer versus warning label," *Body Image*, 10 (4), 472–80.
- Cardozo, Richard N (1965), "An Experimental Study of Customer Effort , Expectation , and Satisfaction," II, 244–49.
- Frey, Kurt P and Alice H Eagly (1993), "Vividness can undermine the persuasiveness of messages.," *Journal of Personality and Social Psychology*, 65 (1), 32–44.
- Goering, Patricia A (1985), "Effects of Product Trial on Consumer Expectations, Demand, and Prices," *Journal of Consumer Research*, 12 (1), 74–82.
- Kahneman, Daniel and Shane Frederick (2002), "Representativeness revisited: Attribute substitution in intuitive judgment," ... *of Intuitive Judgment*, intuitivejud, 49–81.
- Kamins, Michael, Ingrid M Martin, Long Beach, Track Chairs, Todd Arnold, Chris Blocker, Wendy a Boland, Deanne Brocato, Rosa Chun, Paul Connell, Mary Curren, Gary Davies, Kristin Diehl, Anthony Foley, Mark Houston, Wes Johnston, Lynn Kahle, Stephen Kim, Yong Liu, Lawrence J Marks, Shashi Matta, Michael a Mayo, Sam Min, Joseph Nunes, Scott Radford, Lisa Scheer, Jeff Tanner, Clay Voorhees, Kristen Walker, Susan Whelan, and Eric Yorkston (2009), "Enhancing Knowledge Development in Marketing," 20.
- Kisielius, Jolita and Brian Sternthal (1984), "Detecting and Explaining Vividness Effects in Attitudinal Judgments," *Journal of Marketing Research*, XXI (February), 54–64.
- Kivetz, Ran and Itamar Simonson (2000), "Incomplete Information on Choice Consumer," *Journal of Marketing Research*, 37 (4), 427–48.
- Lee, Leonard, On Amir, and Dan Ariely (2009), "In Search of Homo Economicus: Cognitive Noise and the Role of Emotion in Preference Consistency," *Journal of Consumer Research*, 36 (2), 173–87.
- Messaris, Paul (1997), "Visual Truth, Visual Lies," in *Visual Persuasion: the Role of Images in Advertising*, 129–61.
- Nisbett, Richard E and Lee Ross (1980), *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Olshavsky, Richard W and John A Miller (1972), "Consumer Expectations, Product Performance, and Perceived Product Quality," *Journal of Marketing Research*, 9 (February), 19–21.

- Paivio, Allan (1971), *Imagery and Verbal Processes*, Holt, Rinehart and Winston.
- , John C Yuille, and Stephen A Madigan (1968), “Concreteness, Imagery, and Meaningfulness Values for 925 Nouns.,” *Journal of experimental psychology*, 76 (1), Suppl:1–25.
- Reber, Rolf, Norbert Schwarz, and Piotr Winkielman (2004), “Processing fluency and aesthetic pleasure: is beauty in the perceiver’s processing experience?,” *Personality and social psychology review: an official journal of the Society for Personality and Social Psychology, Inc*, 8 (4), 364–82.
- , Piotr Winkielman, and Norbert Schwarz (1998), “Effects of Perceptual Fluency on Affective Judgments,” *Psychological Science*, 9 (1), 45–48.
- Sharps, Matthew J and Michael A Nunes (2002), “Gestalt and Feature-Intensive Processing: Toward a Unified Model of Human Information Processing,” *Current Psychology*, 21 (1), 68–84.
- Swait, J. and T. Erdem (2007), “Brand Effects on Choice and Choice Set Formation Under Uncertainty,” *Marketing Science*, 26 (5), 679–97.
- Townsend, Claudia and Barbara E. Kahn (2014), “The ‘Visual Preference Heuristic’: The Influence of Visual versus Verbal Depiction on Assortment Processing, Perceived Variety, and Choice Overload,” *Journal of Consumer Research*, 40 (5), 993–1015.
- Trope, Yaacov and Nira Liberman (2010), “Construal-level theory of psychological distance.,” *Psychological review*, 117 (2), 440–63.
- Tversky, Amos and Daniel Kahneman (1974), “Judgment under Uncertainty: Heuristics and Biases,” *Science (New York, N.Y.)*, 185 (4157), 1124–31.