Why Are Benefits Left on the Table? Assessing the Role of Information, Complexity, and Stigma on Take-Up With an Irs Field Experiment

Saurabh Bhargava, Carnegie Mellon University, USA
Day Manoli, University of California Los Angeles, USA

We address the puzzle of incomplete EITC take-up with a field experiment testing the role of information, complexity, and stigma among 35,050 individuals owed benefits of $26 million. We find that simplicity and display of benefits significantly increases take-up, and our interventions could reduce incomplete program take-up from 25% to 22%.

[to cite]:

[url]:
http://www.acrwebsite.org/volumes/1012975/volumes/v40/NA-40

[copyright notice]:
This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
Designing Effective Choice Architectures  
Chairs: Oleg Urminsky, University of Chicago, USA  
Indranil Goswami, University of Chicago, USA

Paper #1: In Search of Optimally Effective Defaults  
Indranil Goswami, University of Chicago, USA  
Oleg Urminsky, University of Chicago, USA

Paper #2: Redundant Information as a Choice Architecture Tool: How Attribute Decomposition on Displays can be used to Highlight Important Dimensions for Consumers  
Christoph Ungemach, Columbia University, USA  
Adrian Camilleri, Duke University, USA  
Eric Johnson, Columbia University, USA  
Richard Larrick, Duke University, USA  
Elke U. Weber, Columbia University, USA

Paper #3: Product Level and Segment Level Differences in the Effectiveness of a Longitudinal Labeling and Choice Architecture Intervention at a Large Hospital Cafeteria  
Jason Riis, Harvard Business School, USA  
Susan Barraclough, Massachusetts General Hospital, USA  
Doug Levy, Massachusetts General Hospital, USA  
Lillian Sonnenberg, Massachusetts General Hospital, USA  
Anne Thorndike, Massachusetts General Hospital, USA

Paper #4: Why are Benefits Left on the Table? Assessing the Role of Information, Complexity, and Stigma on Take-up with an IRS Field Experiment  
Saurabh Bhargava, Carnegie Mellon University, USA  
Day Manoli, University of California – Los Angeles, USA

SESSION OVERVIEW

A large literature has established that people’s preferences are often malleable and are influenced by context and environmental cues. Much of the research along these lines has provided “existence proofs” – demonstrating factors which can consistently have some effect on decisions. Therefore, a natural question is how to best structure the decision context (e.g., the “choice architecture”, Thaler and Sunstein 2008) such that the impact on outcomes is maximal. Answering this question not only has immense practical implications, but also poses an important theoretical challenge to decision researchers. In most choice environments, there are multiple potential influences and existing theories often provide either no prediction or competing predictions of when contextual factors will be most impactful. Hence, to generate recommendations for designing “optimally effective” choice architecture, we need to pit all the competing forces against one another, identify the influences that yield differential effectiveness and come up with a deeper understanding that can generate simple principles to guide choice architecture decisions. This session is aimed both at providing some answers to such questions, but even more so, to demonstrate the usefulness of this approach for consumer research, both practically and theoretically, spurring related future research.

Pre-set defaults have very robust effects on choices, and have been used effectively in various consequential domains, such as organ donation and retirement savings. Different extant research, however, has implied that defaults that are either too high or too low might have detrimental consequences. Hence it is unclear how to maximize the impact of a default policy, and whether setting a default might sometimes have negative effects relative to no default. Across a range of domains (pricing, charitable donation, savings goals) that differ in the decision maker’s motivations and implications of being amenable to the default, Goswami and Urminsky find that high defaults generally have a greater impact on increasing ‘participation amount’ than low defaults, without any difference in participation rate. Evidence for immediate downstream consequences of high versus low default on goal disengagement and rejection of subsequent defaults seem to be limited. These findings suggest an important principle of choice architecture, that setting low defaults might have negative consequences, but there is little risk of setting too high a default.

Another common approach to choice architecture is to provide (or mandate) detailed information for consumers. Research has indicated that consumers tend to undervalue the importance of long-term savings factors such as fuel economy, even when provided with detailed information about it in their car purchasing decisions. Ungemach, Camilleri, Johnson, Larrick and Weber show that a simple attribute decomposition intervention, communicating the fuel economy information multiple times and in different formats (e.g. miles per gallon, annual fuel cost etc.), can increase the weight decision makers assign to this attribute irrespective of the purchase motivations. The efficacy is not driven by changing inferences about superiority on the most salient dimension, since even redundant information on the attribute influenced assigned weights. Rather, these findings suggest that optimizing choice architecture involves taking into account the weight and emphasis placed on information by considering the distribution of attributes across the key dimensions.

Along similar lines, recent attempts to improve consumers’ food choices by providing calorie information have yielded largely disappointing results. In a large scale field study in a hospital cafeteria, Riis, Barraclough, Levy, Sonnenberg, and Thorndike demonstrate that labeling foods as red, yellow, or green, coupled with interventions to make the green food items more visible and accessible, had a dramatic impact. Specifically, the sale of bottled water increased 26%, seemingly driven by consumers’ desire to make a healthy choice on a complimentary item in their meal without making sacrifices on the taste of focal component of the meal. These findings demonstrate that optimizing choice architecture may occur when information is made as easy to process and as accessible as possible, even at the expense of detail, and when it is offered on complementary items rather than focal items.

In another large scale field study, Bhargava and Manoli show how Earned Income Tax Credit (EITC) claims can be increased among eligible taxpayers (the working poor). They find that simplifying the program information meant to address issues like lack of awareness, misunderstanding of EITC rules including eligibility, and underestimation of benefit size substantially increases claims. More importantly, they test these interventions against competing interventions, including attempts to reduce perceived transaction cost and social stigma and risk of audit, which did not have any effect. Overall, the successful interventions reduced incomplete take-up from 25% to 22%, with a cumulative effect measured in millions of dollars. Their field study (and accompanying survey) findings further demonstrate that optimizing choice architecture involves maximizing comprehension and confidence with the information provided, rather than maximizing the quantity or detail of information.

Taken together, the four papers suggest that equally plausible interventions in the decision context can often have profoundly different effects, across many consequential domains. We expect that this
session would be of interest to a broad audience including consumer and decision researchers, economists, and policy practitioners. This session will also contribute to the diversity of types of research at ACR, by presenting data from two large scale field studies. We believe that our understanding of how to use choice architecture most effectively is under-developed, and that attempting to identify the conditions that yield optimal effectiveness is of crucial importance for both theoretical and practical reasons. We hope that this session will stimulate ideas and promote future research in the optimal effectiveness of choice architecture design.

In Search Of Optimally Effective Defaults

EXTENDED ABSTRACT

Pre-set default options are a key element of choice architecture. However, recommendations and defaults might yield a behavioral backlash if they are inconsistent with initial impressions (Fitzsimons and Lehmann 2004), invoke negative assessment of motives (Brown and Krishna 2004), or if the individual or group setting the default option is distrusted (Liersch and McKenzie 2011). Companies often set a fairly low default 401(k) contribution rate relative to chosen rates of saving, partly out of a concern that a higher default rate could reduce enrollment (WSJ, 7/2011). Thus, a critical unanswered question is how to set optimally effective defaults and whether setting defaults too low or too high is likely to have a greater impact.

In six completed studies, we set out to empirically test the potential for negative backlash effects of high defaults and anchoring effects of low defaults on active choices (i.e. not passive enrollment). To answer this question, we investigated the impact of default size on total amounts, as well as average ‘amounts’ among participants, participation rate, and immediate downstream consequences on goal disengagement and response to subsequent defaults. In order to generalize our findings, we use a range of domains, from savings to pricing, where the decision makers’ motivations are very different and hence defaults are likely to be operating via different routes. In addition, we specifically investigate the impact of psychological reactance (Brehm 1966; Hong and Faedda 1996) and trust on different levels of default.

In Study 1 (n=502), we tested default policies on consumers’ choices of savings goals (defaults of 7%, 5%, 3%, 1% or no default; b/w/s). The default was calculated from data attributed either to the US government (a well-known source) or U.S. bureau of Economic Analysis (a lesser known source). Overall, the average savings goal increased monotonically with increasing default. When government was the source, the savings rate was significantly increased by a high (7%) default and significantly decreased by a low (1%) default, relative to no default. Furthermore, high default did not increase drop-out rates (not setting any savings goal at all) relative to other conditions. Reactance reduced the overall efficacy of defaults, but did not differentially impact high defaults. In a follow-up study, Study 2 (n=443), we replicated our findings with unrealistically high default savings goals (defaults of 40%, 20%, 5%, 1% or no default; b/w/s). Even for these extreme rates, we found no evidence of backlash for high defaults (relative to no default), even among high reactance participants. In addition, non-participation rates, self-reported confidence and decision satisfaction did not differ.

In the next two studies, Study 3 (n=241) and Study 4 (n=453) we tested these results with participants facing real stakes. In study 3 we used a setting where ‘higher’ choice is not intrinsically desirable for the participants. Participants picked their own price to buy academic planners from a menu of prices with a default of $40, $5, $1, or no defaults b/w/s. Inexperienced participants, who did not use planners in the previous year, valued the product more when the default was higher. The effect could not be explained by increased attractiveness of the planners for inexperienced buyers, or default acting as a signal of information. Also, there was no difference in the participation rates across conditions. The high defaults ($40, $5) were effective in increasing willingness to pay, even though no one selected the defaults.

In Study 4, participants decided on how much they would be willing to donate from a real lottery award which every participant had an equal chance to win. High defaults (but now low defaults) significantly increased intended donations when participants had low information (resulting in low trust) about the charitable organization. Defaults did not have an effect on participants with high information. The pattern was true for participation rate as well. Therefore, in the absence of credibility bolstering information, high defaults helped to increase both average amounts as well as participation rates.

In two other hypothetical studies we tested people’s response to defaults in a choice of how many promotional emails to allow a company to send them. In both Study 4 (n=192) and Study 5 (n=71), the average number of emails permitted did not vary significantly based on the default, and there were no negative consequences of high defaults. Also, across all studies we find very limited evidence of immediate downstream consequences of encountering high versus low defaults in a prior decision task.

Our findings suggest minimal potential for backlash against high defaults, particularly in multi-option choices. Even when a high default is rejected, it can influence choices by anchoring the participant on a high value. In contrast, we find that deviation from maximal defaults occurs primarily when defaults are too low, rather than too high. While psychological reactance reduces the impact of defaults, we find little evidence that high reactance leads to a backlash against high defaults. Trust is an important factor in the absence of defaults, but defaults can improve outcomes when trust is low.

Redundant Information as a Choice Architecture Tool: How Attribute Decomposition on Displays can be used to Highlight Important Dimensions for Consumers

EXTENDED ABSTRACT

Among the scientific community there is no doubt that climate change is a reality and that we must act immediately to reduce CO2 emissions (Karl, Melillo, & Peterson, 2009). While purely economic solutions have already been leveraged, it has been increasingly recognised that insights from psychology and behavioural economics are crucial to initiate and maintain the changes required to combat the threat (Johnson et al., 2012; Weber & Stern, 2011). One of the most promising psychology-based approaches to producing behavioural change is through the development of “choice architecture”.

The choice architecture approach to behavioural change rests on the established finding that individuals often construct their preferences in the immediate choice context (Slovic, 1995; Lichtenstein & Slovic, 2006) and that decision context often influences the choice that is made (Thaler & Sunstein, 2008). A key principle of this approach is that there is no neutral choice context and any framing can and will affect what is chosen. Examples of nudges that have been shown to be effective include the selection of defaults (Johnson, Bellman, & Lohse, 2002), the number of alternatives presented (Payne, Bettman, & Johnson, 1993), the partitioning of options and attributes (Fox, Ratner, & Lieb, 2005), and rescaling (Burson, Larrick, & Lynch, 2009). A related observation with potential choice architecture implications is “attribute splitting”, describing the phenomenon that decomposing an attribute into several sub-attributes...
can increase the psychological weight assigned to that dimension (Jacobi & Hobbs, 2007; Weber, Eisenfurth & von Winterfeldt, 1988). Attribute splitting has been demonstrated within weighting tasks but it remains unclear how the distribution of attribute values affects preferences.

The inherent goal of the presented series of experiments is to further understand the impact of and mechanisms behind attribute decomposition, and develop a new choice architecture tool within the domain of environmental decisions.

Experiment 1. A primary source of human-induced CO2 emissions is transportation, contributing approximately 27% of total U.S. greenhouse gas emissions in 2008 (http://www.epa.gov/otaq/climate/basicinfo.htm). However, prior research seems to indicate that consumers presented with relevant information undervalue fuel economy information in their car-purchasing decisions (e.g. German, 2002; Maples, 2003; Patterson, 2002). An interesting feature of the new fuel economy labels launched by the Environmental Protection Agency is that the fuel economy information is presented in a number of different, but highly correlated, metrics (e.g. MPG, GPM, annual fuel cost). In order to test whether repetition of the same attribute potentially reduces undervaluation of fuel cost information, as implied by the findings on attribute decomposition, we conducted a discrete choice experiment. 200 online respondents were asked to choose between pairs of cars for which EPA fuel economy labels (7 correlated metrics of fuel economy) and the prices of the car (one single price) were provided. The choice pairs were created using three MPG levels (High, Medium, Low) and three Price levels (High, Medium, Low). Each participant had to make 4 choices in 2 different scenarios: In the ‘Capability’ scenario, participants helped a friend select the car with the lowest overall cost (i.e. price and operating cost) given a fixed yearly mileage of 15,000 miles and 5 years before replacing the car; In the ‘Preference’ scenario, participants could choose the car they personally preferred assuming the same mileage and driving period. The label presentation format was also manipulated such that each pair of labels was either presented simultaneously or sequentially.

Discrete choice analysis showed that when using the label with multiple fuel efficiency metrics, participants tended to slightly overweight the fuel economy information relative to the car price. This tendency was observed in both scenarios and was not affected by the order the scenarios were presented in.

Experiment 2. In order to investigate whether people actually make use of all the highly correlated information on the label, we conducted a second experiment in the laboratory (n=10) where we tracked participant’s eye movements while going through the same choice task as in Experiment 1. Although the calculation of the overall cost on the ‘Capacity’ task only required the combination of two numbers (one fuel economy metric and the price) the process tracking data revealed that most participants did indeed allocate attention to all of the available metrics, including the redundant information. Furthermore, the gaze paths show that redundant information was not merely scanned initially as part of the search process but participants repeatedly transitioned between similar attributes across labels, indicating that this information was used as part of the choice process itself.

Experiment 3. The observations made in Experiment 1 and 2 suggest that the presentation of multiple, highly correlated attributes may cause decision-makers to psychologically weigh this attribute more heavily. Our investigation is also related to the “majority rule” heuristic, which asserts that decision-makers tend to choose the option superior on most of the available cues (Zhang et al., 2005). However, this heuristic is applicable only to alternatives with more than two attributes and it is unclear whether the heuristic can apply to multiple redundant attributes.

Our aim in Experiment 3 was to systematically manipulate the proportion of redundant attributes for a given dimension and observe the impact on the psychological weight assigned to that dimension. To fulfill this aim, we carried out a 2 (Upfront Cost Information: 1 metric vs. 4 metrics) x 2 (Running Cost Information: 1 metric vs. 4 metrics) x 3 (Upfront Cost: High, Medium, Low) x 3 (Running Cost: High, Medium, Low) between-subjects discrete choice experiment. 400 online respondents were presented with nine decision problems that each required a choice between two vehicles that varied in Upfront and Running costs, as in the ‘Capability’ scenario in Experiment 1 and 2. We found that, in general, presentation of multiple redundant attributes shifted the psychological weight assigned to that dimension.

Together, the results from these studies show that redundant information is utilized in the formation of preferences and that it might be overweighted. This observation offers choice architects a new tool to help decision-makers identify the important dimensions in a choice problem and nudge people towards buying more energy efficient cars.

**Product Level and Segment Level Differences in the Effectiveness of a Longitudinal Labeling and Choice Architecture Intervention at a Large Hospital Cafeteria**

**EXTENDED ABSTRACT**

Although many food retailers are trying to encourage their customers to make healthier choices, few of their interventions have been studied formally, and few are able to claim much success. The current study incorporated a two-phase retail intervention in a large hospital cafeteria. In the first phase, a simplified menu labeling system was applied throughout the cafeteria. Since calorie labeling interventions have shown limited success in changing customer purchase behavior, this study implements a simpler and more meaningful labeling scheme whereby all foods in the cafeteria were labeled as red, yellow, or green (with green being the healthiest). The second phase of the study altered the choice architecture of the food and beverages in the cafeteria, making it easier for customers to see and find the healthier food items. We compared the purchase patterns made during these two 3-month intervention phases with purchase patterns during a 3-month baseline period.

The setting for the study was the main cafeteria at large hospital in Massachusetts. More than 5000 individuals visit the cafeteria during a typical weekday and daily revenues exceed $30,000. Just over a quarter of the revenues come from employees who use a “meal card” which allows purchases to be directly deducted from their paychecks. The purchases of these individuals can hence be tracked over time.

The first phase of the intervention involved a point-of-purchase labeling intervention designed to educate the employees, patients, and visitors about the nutritional value of the foods and beverages of the cafeteria. All foods and beverages in the cafeteria were labeled as red, yellow, or green based on criteria that reflect USDA healthy eating recommendations. Displays around the cafeteria informed customers that Red items should be consumed rarely, as they are high in calories or unhealthy fat while Green items, which feature fresh fruits and vegetables, whole grains, and healthy proteins, should be consumed often. Yellow items are intermediate. The red-yellow-green categories were developed by a team of nutritionists at the hospital.
Following the three months of Phase 1, the second phase of the cafeteria intervention began. The labeling scheme did not change from Phase 1; food and beverages continued to be labeled red, yellow, and green; but choice architecture changes were made.

The intervention at Phase 2 involved a series of changes to the layout of the cafeteria and the standard servings of some items. These changes were designed to increase the purchase rates of green items by making them more salient and more convenient to select. Examples include the following: a) Bottled water and diet soda (both green items) appeared at more locations throughout the cafeteria, including several prominent, easy-access baskets near each food service station; b) Pre-packaged side salads became available beside the pizza and grill stations, where previously customers who wanted a side salad must make a separate trip to the salad bar, and c) Healthier snack items and sandwiches were moved to the most visible and accessible shelves, while less healthy items in these categories were moved to less visible and less accessible shelves.

Cash register data recorded all purchases for the 9 month study period. In addition to analyzing cafeteria-wide purchase trends, we followed a cohort of 5000 meal card users. For each user in the cohort, we knew their age, sex, race/ethnicity, and job type, in addition to their purchase history.

The labeling intervention led to large reductions in the consumption of unhealthy foods and beverages, but this paper will focus on the choice architecture intervention. Some analyses of category effects and segment interactions are pending, but some results are clear. By far the largest single product category change due to choice architecture was for bottled water which showed a 26% increase in sales during the choice architecture period. Changes in other categories were less than half that. We believe that this is because beverages are complements to meals, and taste sacrifices here will seem relatively modest compared to taste sacrifices on a meal entrée or on a food favorite.

In terms of segment differences, there were very large differences in purchase patterns at baseline: For example, Black and Latino employees purchased 31% red items while, white employees consumed just 18%. Reductions of red item purchases and increases in green item purchases were however largely uniform across ethnic and income groups. All groups showed similarly substantial improvements in healthy purchasing. For example, the average reduction in calories per beverage was about 15% for each group, from baseline to Phase 2.

Finally, cafeteria revenues were consistent through the three periods, suggesting that such interventions may be sustainable. The interventions are ongoing at the hospital and we are currently looking at the second year of register and panel data.

Why are Benefits Left on the Table? Assessing the Role of Information, Complexity, and Stigma on Take-up with an IRS Field Experiment

EXTENDED ABSTRACT

A well-recognized, and perhaps surprising, feature of transfers to the economically and socially disadvantaged is that many targeted individuals fail to take-up their benefits (Currie 2006). The Earned Income Tax Credit (EITC), the nation’s largest means-tested cash transfer program, is a prime example with an estimated 25 percent rate of incomplete take-up that amounts to 6.7 million non-claimants each year (Plueger 2009).

The policy consequences of incomplete take-up are significant. A typical EITC non-claimant forgoes credits equivalent to 33 days of income. Moreover, non-claimants forfeit other advantages, such as those related to family health, education, or consumption, that may be linked to transfers (Hoynes, Miller, and Simon 2011; Dahl and Lochner 2011; Smeeding, Phillips, and O’Connor 2001). The problem may be even more severe for other means-tested programs.

Several explanations are generally cited for incomplete take-up: lack of information, stigma, transaction costs, and complexity (Currie 2006). In this paper, we test the effect a set of novel interventions on take-up with a unique field experiment administered in collaboration with the Internal Revenue Service (IRS). In addition to identifying strategies to improve take-up, our experiment allows us to evaluate leading theories as to its causes. Specifically, we test the role of information (regarding program benefits, costs, and rules), the complexity of such information, and program stigma on the take-up of the EITC.

We implement the experiment by modifying the informational content and complexity of IRS tax mailings and distributing these to the universe of over 35,000 tax filers from California who failed to claim their TY 2009 EITC despite presumed eligibility and the receipt of an initial reminder notice. Each mailing communicates likely eligibility for the program, and includes a worksheet which a recipient can complete and return to claim a credit. We use the differential response across mailings to adjudicate among the tested mechanisms. To maximize statistical efficiency, and to permit tests of treatment interactions, we assign and distribute the mailing elements—that is, the reminder notice, claiming worksheet, and an experimental envelope—with three independent randomizations to blocks defined by zip code and dependent status. The packets were published, assembled and mailed by the IRS in a single batch in mid-November of 2010. All told, we inform individuals of $26 million in unclaimed government benefits, of which about $4 million is ultimately claimed.

The experiment offers five main findings. First, we observe that the mere receipt of the experimental “control,” just months after the receipt of a first, similar, IRS reminder mailing, prompts 0.14 of the residual non-respondents to take-up (this compares to an initial notice response of 0.41). Second, the experiment suggests that informational complexity influences response. Relative to the control notice and worksheet (0.14), a notice with a simplified layout and less repetition improves take-up by 0.06 (p < .01), while a shorter worksheet, without the inclusion of criteria that does not substantively screen for eligibility, increases response by 0.04 (p < .01). Importantly, the basic information conveyed by the control notice and worksheet, and the complexity treatments, is equivalent. Third, providing benefit information also raises take-up. Displaying the upper bound of a potential benefit improves take-up by 0.08 relative to the 0.23 response of the baseline notice in which no benefit is displayed (p < .01). Fourth, we find that attempts to clarify the time and penalty costs associated with completing and returning the worksheet do not improve response. Finally, our attempts to reduce program stigma do not improve take-up.

By integrating the results from the experimental findings and randomized surveys of tax-filer beliefs, we can arbitrate between competing explanations for incomplete take-up in the literature. Our findings suggest that incomplete take-up in the specific context of EITC filers is due primarily to low program awareness, incomplete information regarding benefits and eligibility, as well as the high complexity of information. We do not find evidence to implicate direct transaction costs, the likelihood of an audit, or program stigma. A set of psychometric surveys offers additional insight into why take-up is so sensitive to modest changes in informational content and presentation. The evidence suggests that information, and informational complexity, may shape behavior by prompting both
direct and indirect inferences regarding program parameters, as well by, possibly, changing the degree to which readers attend to the information.

Overall, the potential policy impact of the tested interventions is large. We calculate that the most effective experimental treatments, if applied to the entire population of filing non-claimants, could reduce incomplete take-up among filers from 10% to 7%, and overall incomplete take-up from an estimated 25% to 22%. The increase in response due to our context-based interventions is equivalent to that which would be produced by expanding benefits by up to 101% for this population.