Research Policy 42 (2013) 273–286

Contents lists available at SciVerse ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features

Henry Sauermann[a,*], Michael Roach[b]

[a] Georgia Institute of Technology, College of Management, 800 W. Peachtree St., Atlanta, GA 30308, USA
[b] University of North Carolina, Chapel Hill, Kenan-Flagler Business School, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:
Received 1 May 2011
Received in revised form 1 May 2012
Accepted 16 May 2012
Available online 12 June 2012

Keywords:
Web surveys
Response rates
Experiment
Incentives
Personalization
Survey timing
Response bias

ABSTRACT

Web surveys have become increasingly central to innovation research but often suffer from low response rates. Based on a cost–benefits framework and the explicit consideration of heterogeneity across respondents, we consider the effects of key contact design features such as personalization, incentives, and the exact timing of survey contacts on web survey response rates. We also consider the benefits of a "dynamic strategy", i.e., the approach to change features of survey contacts over the survey life cycle. We explore these effects experimentally using a career survey sent to over 24,000 junior scientists and engineers. The results show that personalization increases the odds of responding by as much as 48%, while lottery incentives with a high payoff and a low chance of winning increase the odds of responding by 30%. Furthermore, changing the wording of reminders over the survey life cycle increases the odds of a response by over 30%, while changes in contact timing (day of the week or hour of the day) did not have significant benefits. Improvements in response rates did not come at the expense of lower data quality. Our results provide novel insights into web survey response behavior and suggest useful tools for innovation researchers seeking to increase survey participation.

1. Introduction

Scholars of science and innovation increasingly employ survey data from individual scientists and engineers as well as from administrators and managers. Although many of the early and most influential surveys were conducted by national agencies such as the National Science Foundation in the United States or various national statistical offices in Europe,[1] there has been a sharp increase in the number of independent survey efforts, especially online surveys. For example, in the past twelve months there have been more than twenty articles published in *Research Policy* that employ survey data, nearly half of which were administered online.[2] Part of the reason behind the growing trend toward online surveys is that they can be conducted at relatively low cost and within a shorter time frame than conventional paper-based or telephone surveys. In addition, it has become quite easy to obtain email contact information for large samples of scientists and engineers by extracting such information from publications, patents, résumés, university websites or similar sources (cf. Bruneel et al., 2010; Fini et al., 2010; Haeussler, 2011).

Despite the important role of surveys in innovation studies, relatively little attention is given to the challenges of achieving high response rates. Survey participation is a particularly acute issue for web surveys, which tend to suffer from lower response rates than other survey modes, especially as low survey costs lead to "oversurveying" (cf. Couper, 2000; Fricker et al., 2005; Kaplowitz et al., 2004; Rogelberg & Stanton, 2007). For example, while short and direct surveys involving phone follow-ups can achieve relatively high response rates of 40–70% (Brostrom, 2010; Van Looy et al., 2011), more detailed online surveys often exhibit lower response rates of around 10–25%. Low response rates, in turn, reduce sample size and statistical power. Moreover, low response rates may also lead to nonresponse bias and affect the validity of survey results irrespective of the sample size. As a consequence, there is a need to better understand web survey response behavior and to develop techniques to increase web survey response rates.

We contribute to the study of innovation by examining how contact design features such as personalization, incentives, and the timing of survey invitations affect response rates among scientists and engineers and by deriving recommendations for survey

* Corresponding author. Tel.: +1 440 385 4883; fax: +1 404 894 6030.
E-mail addresses: henry.sauermann@mgt.gatech.edu (H. Sauermann), michael_roach@kenan-flagler.unc.edu (M. Roach).

[1] For example, the Scientists and Engineers Statistical Data System (SESTAT) is based on individual-level surveys managed by the National Science Foundation, and the Community Innovation Survey (CIS) is an integrated firm-level survey effort of European statistical offices.

[2] We searched each issue of Research Policy from June 2010 through May 2011. We identified a total of twenty-four articles that employed survey data, of which ten were postal mail, seven were online, and the remaining seven did not specify the survey mode. Seven of these twenty-four articles used national surveys such as the Community Innovation Survey, while seventeen were "independent" surveys.

0048-7333/$ – see front matter
http://dx.doi.org/10.1016/j.respol.2012.05.003

researchers. We first review prior work on the drivers of response rates and present a generalized cost–benefits framework that incorporates heterogeneity across respondents. We then examine the effectiveness of various contact design features using a sample of over 24,000 scientists and engineers who were invited to respond to a survey on their organizational environment, work activities, and their career choices. To examine causal effects, we randomly assigned potential respondents to 25 experimental conditions that differed systematically with respect to various contact design features.

This study extends prior work on web survey response rates in several ways. First, we consider not only design parameters that were relevant in mail surveys, but also features that reflect new opportunities provided by web surveys such as the exact timing of survey contacts. Second, in addition to design features of survey contacts at any particular point in time ("static design features"), we consider several "dynamic design features" that capture aspects of the survey strategy over time including the number of reminders, the delay between reminders, and changes in design features over the survey life cycle. Finally, much of the prior literature on survey response rates has used household or general population samples. It is not clear whether the resulting insights generalize to scientists and engineers, who may differ from the general population with respect to characteristics such as their interest in research, internet use, or work schedules. Thus, our findings based on a sample of scientists and engineers should be particularly relevant for survey researchers working in the areas of science and innovation.

Our results suggest several design features that significantly increase response rates, but we also show that other features have little to no impact on response rates. As such, our results provide novel insights into web survey response behavior of scientists and engineers and provide survey researchers with guidance regarding where to focus their survey design efforts. In addition, this paper may also be of interest to readers of web survey based studies who seek more background on this important and increasingly utilized methodology.

2. Conceptual framework

2.1. The importance of response rates in survey studies

Survey researchers should seek high response rates for several reasons (Couper and Miller, 2008; Dillman et al., 2009; Simsek and Veiga, 2001). First, for a given initial sample size, a higher response rate will translate into a larger number of responses that can be used for statistical analyses. A higher number of cases, in turn, increases statistical power and the researcher's ability to detect significant relationships among measures of interest (Cohen, 1992). Moreover, a larger number of cases may allow researchers to conduct empirical analyses for different subsets of the population, providing insights into moderating effects and heterogeneity. Examples for such a more nuanced analysis include recent work on industry–academia interactions and academic entrepreneurship (e.g., Ding and Choi, 2011; Haeussler and Colyvas, 2011; Sauermann et al., 2012). Small samples, on the other hand, may not only limit the econometric techniques that can be applied to the data but may also affect the credibility of research results in the eyes of reviewers and readers (Rogelberg and Stanton, 2007). Finally, higher response rates are an important way to increase the representativeness of the sample and to decrease nonresponse bias. As such, survey data with high response rates will typically provide more accurate insights into the underlying population (cf. Rogelberg and Stanton, 2007; Wagner, 2008). At the same time, researchers should be aware that nonresponse bias may not be reduced or may even increase if higher response rates are achieved by using contact design features that attract only particular types of individuals (Groves et al., 2006; Groves and Peytcheva, 2008). Therefore, it is important to understand the effectiveness of design features in increasing response rates while also considering the degree to which they may selectively attract certain types of respondents.

2.2. A generalized cost–benefits approach and heterogeneity across respondents

To discuss the effectiveness of various contact design features, we consider a general cost–benefits framework, where the costs and benefits of survey participation (from the respondent's perspective) include economic as well as non-economic factors (Dillman et al., 2009). Benefits of survey participation may include any financial incentives offered by the researcher, individuals' satisfaction of curiosity regarding the survey topic, the feeling to contribute to research, or a sense of reward from helping others. On the other hand, costs of participation involve factors such as time spent answering the survey questions, discomfort from having to think about difficult questions, and potential risks regarding the disclosure of confidential data (Anderson, 2003; Dillman et al., 2009; Groves et al., 2006; Porter, 2004).[3] The various design features of a survey invitation may affect the perceived costs and benefits of responding and thus recipients' decisions to participate in the survey.

Prior research has focused on how contact design characteristics affect average response rates and has not typically considered heterogeneity across individuals. However, it is likely that recipients differ with respect to the costs and benefits implied by a particular survey attribute. For example, some individuals may face high opportunity cost of responding when approached on Mondays while others may face particularly high costs on Tuesdays. Survey researchers may be able to exploit such heterogeneity through a "dynamic strategy" that varies design features over the survey life cycle, e.g., between the initial contact and subsequent reminders. Such a dynamic strategy essentially attempts to appeal to different segments of the survey population in each round and exploits the fact that only one response is needed from each person.

In the following conceptual part of this article, we draw on considerations of costs and benefits as well as heterogeneity across individuals to discuss potential effects of static as well as dynamic survey contact features on response rates. We include in our discussion features that have received considerable attention in the context of mail surveys as well as factors that may represent new opportunities in the particular context of online surveys.

2.3. Static design features

2.3.1. Personalization

Survey researchers can approach potential respondents using some general salutation such as "Dear colleague" but can also

[3] We focus on the costs of responding from the respondent's perspective and do not provide an explicit discussion of the costs of conducting the survey from the researcher's perspective. The overall costs of conducting web surveys tend to be quite low and, except for incentives, none of the design features discussed in this paper should significantly affect those costs. The survey literature provides extensive discussions of survey costs, especially in the context of person-to-person interviews and of mail surveys (Cobanoglu et al., 2001; Dillman et al., 2009; Shannon and Bradshaw, 2002). An interesting recent development is the idea to minimize costs by conducting surveys in multiple phases and to modify the survey strategy in response to observed response patterns over time ("responsive design") (Groves and Heeringa, 2006; Wagner, 2008).

personalize invitations using first or last names. Personalization may establish a connection between the researcher and the recipient, likely increasing the psychological benefits the recipient derives from responding (Dillman et al., 2009). A personalized email may also convey to the recipient that she was selected specifically and that her response is important for the success of the study. Personalization has generally been found to increase response rates in mail surveys as well as Internet surveys (Cook et al., 2000; Dillman et al., 2009). However, a recent study did not find a significant impact of personalization on web survey response rates and the authors suggested that personalization may have lost credibility in today's world of "mass customization" (Porter and Whitcomb, 2003a).

We extend the existing work by treating personalization as varying in degree. More specifically, it is conceivable that recipients' evaluations of the costs and benefits of participating in the survey depend on the specific way in which personalization is implemented. Among young scientists and engineers (our sample), the use of the first name alone may be considered as more personal than the use of the full name and build a stronger relationship between the researcher and the recipient. Moreover, the use of the full name could increase the psychological costs of responding by raising confidentiality concerns, i.e., that the researcher can identify the respondent in the "offline" world. Overall, we expect that personalization increases response rates but that the use of the first name is more effective than the use of the full name.

2.3.2. *Financial incentives*

Many survey researchers seek to increase response rates by using financial incentives. This practice is consistent with standard economic theory, which suggests that participation and effort increase with the utility individuals expect to derive from engaging in an activity. A large body of literature has shown the power of financial incentives generally, and a growing body of research also highlights the importance of financial motives in the science and innovation context (Azoulay et al., 2011; Camerer and Hogarth, 1999; Lazear, 2000; Rynes et al., 2005; Sauermann and Cohen, 2010). At the same time, some psychologists and economist have expressed the concern that contingent pay may undermine actors' intrinsic or social motivations to engage in a task, potentially resulting in a negative net effect. Such "motivation crowding-out" may occur if individuals feel that incentives are controlling, if pay is interpreted as a sign that the task cannot be "fun", or if pay leads actors to focus their cost–benefits analysis narrowly on financial aspects (Deci et al., 1999; Frey and Oberholzer-Gee, 1997; Lacetera and Macis, 2010). Motivation crowding-out may be particularly relevant in the survey context since intrinsic and social motivations tend to play an important role in survey participation (Dillman et al., 2009). Moreover, individuals typically have little information on the interestingness of the survey when they receive an invitation to participate, and the financial incentives that are offered by survey researchers are rarely enough to adequately compensate respondents for their time.

What is the evidence regarding the effects of financial incentives on response rates? Many studies have shown positive effects of pre-paid "token" incentives such as two-dollar bills that are mailed with the survey instrument (Anseel et al., 2010; Baruch and Holtom, 2008; Roth and BeVier, 1998; Yammarino et al., 1991). The small size of these pre-paid incentives and the fact that they are not contingent upon performance suggest that the standard economic model does not explain their effectiveness. Rather, scholars argue that token incentives work because they convey the researcher's trust and respect for the recipient and invoke norms of reciprocity (Church, 1993; Porter, 2004; Simsek and Veiga, 2001). Unfortunately, the lack of a universal infrastructure for small online payments makes the use of pre-paid incentives in web surveys difficult. Moreover, even small pre-paid incentives lead to high total survey costs if samples are large.

An alternative to pre-paid token incentives is post-paid incentives such as lotteries where respondents are entered into the drawing of gift certificates. However, the expected payoff from such lotteries is typically very small and post-paid incentives do not invoke the same norms of reciprocity as pre-paid incentives. As such, some survey researchers suggest that post-paid incentives are not effective (cf. Dillman et al., 2009; Porter, 2004), while others find positive effects (Cobanoglu and Cobanoglu, 2003; Goeritz, 2006; Porter and Whitcomb, 2003b). One of the few studies conducted in the online context (Bosnjak and Tuten, 2003) found that lottery incentives had positive effects on response rates while pre-paid token incentives had no effect, adding to the ambiguity regarding the use of financial incentives in the online context.

We suggest that a deeper understanding of the effectiveness of lottery incentives requires the distinction between two key dimensions: the probability of winning and the size of the prize. From a respondent's perspective, the expected payoff increases with the probability of winning and with the size of the prize, suggesting that offering many large prizes would maximize response rates. Given budgetary constraints of researchers, of course, the more relevant question is whether a given incentive budget should be used for a large number of small prizes or a small number of large prizes. Even though the expected payoff from the respondent's perspective is the same, response rates may differ depending on how the lottery is structured. First, prospect theory suggests that individuals facing the chance to win a prize overweigh small probabilities (Fox and See, 2003; Kahneman and Tversky, 1979). Thus, individuals may subjectively perceive a higher expected payoff when offered a small chance to win a large prize compared to being offered a high chance to win a small prize. In addition, we expect large prizes to be more effective to the extent that recipients incur certain fixed transaction costs when using gift certificates (e.g., creating an account with the retailer, searching for useful products, etc.); these transaction costs may offset the value of a small prize but are small in comparison to a large prize. Overall, we predict that financial incentives increase response rates, but that a given incentive budget is more effective if used for a small number of large prizes than for a large number of small prizes.

2.3.3. *Contact timing: day of the week and hour of the day*

The timing of survey contacts has been considered extensively in the context of personal interviews and of telephone interviews (e.g., Piazza, 1993; Weeks et al., 1980). However, timing has received little attention in the context of mail surveys since regular postal services do not typically provide the option to specify particular delivery times. Online surveys offer new opportunities regarding the timing of contacts because they allow researchers to time survey contacts quite precisely with respect to the day of the week and even the time of the day.

It seems generally advisable to time survey contacts such that respondents are not too busy when they receive a survey invitation. However, it is difficult to predict which days of the week or hours of the day are more convenient for the average person. It is conceivable that individuals who work are less likely to respond if the invitation is received during regular work hours because the opportunity cost of responding are high (falling behind in work). However, while evening hours may be less busy in terms of work, individuals may routinely engage in other activities (sports, TV, family activities, etc.) and thus face different kinds of opportunity costs of responding. Similarly, we have no strong predictions regarding days of the week. While weekdays may be busier in terms of work, weekends may be occupied by non-work activities and time spent with the family. It is also conceivable that time of the day and day of the week interact, e.g., that certain times of the day are more effective on some days than others. Even though it is difficult to make any specific predictions, we will examine such interactions in the empirical part of the paper.

While our discussion up to this point assumed that contact timing matters, it is also possible that the timing of email invitations makes little difference. First, to the extent that scientists and engineers work non-conventional work hours and are constantly connected to the Internet, different contact times may be similarly (in-)convenient for them. Second, unlike telephone or in-person surveys, web surveys are "patient" and individuals who receive the invitation at an inconvenient time can simply keep the email and respond at a more convenient time.

Overall, theoretical considerations provide little insight regarding how important contact timing is, or which specific days of the week or times of the day are more effective than others. Despite this lack of predictions, empirical insights into these questions are of great importance for future survey researchers who invariably have to make decisions regarding the timing of survey invitations.

2.4. Dynamic design features

We use the notion of "dynamic design features" to capture aspects of the sequence of contacts over time, including the number of reminders, changes of contacts over time, and the delay between contacts.

2.4.1. Number of contacts and changes in contacts over time

Repeated contacts have been shown to increase response rates and it is often recommended to use as many as three reminders (cf. Cook et al., 2000; Dillman et al., 2009; Groves et al., 1992). Generally speaking, multiple contacts may increase response rates if recipients' subjective costs and benefits from responding change over time.

One reason for changes in costs and benefits over a sequence of contacts may be the repeated contact itself. In the particular context of online surveys, a repeated contact may signal to the recipient that the email is legitimate or that the survey is of particular importance, thus increasing the perceived benefits of the response (e.g., helping a serious researcher) while also reducing the cost (e.g., risk of the email being a phishing attempt). At the same time, repeated contacts signal the researcher's persistence and recipients may respond simply in order to avoid future reminders. A second source of changes in costs or benefits to responding is "transitory" factors that may affect individuals at particular points in time. For example, an individual may have an important project to finish when contacted at time t_1, leading to high opportunity costs that exceed the perceived benefits from responding to the survey request. When contacted again in t_2, however, the individual may have a different draw from the "cost distribution" (e.g., slow time at work) and the costs may be smaller than the perceived benefits, resulting in a response. Fig. 1 illustrates this mechanism.

While the above arguments suggest that repeated contacts will increase response rates, we expect additional benefits from sending repeated contacts that *change* over time, e.g., with respect to their timing or wording. One reason to expect additional benefits from such a "dynamic strategy" is that it may reinforce the exchange relationship between the researcher and the potential respondent. For example, reminders that vary in their wording may appear more genuine and signal that the researcher invests time and effort into the relationship, while repeated contacts that are identical may quickly become irritating.

A second reason to expect benefits from a dynamic strategy is that it exploits heterogeneity across respondents. For example, some individuals may be busy every Monday while others are particularly busy on Tuesdays. Systematically varying features of the repeated contacts (e.g., by sending the initial invitation on a Monday but the reminder on a Tuesday) increases the likelihood that a particular person responds positively to *at least one* of the contacts. Survey researchers may be able to draw on prior knowledge about their particular survey population to identify important dimensions of heterogeneity (e.g., age, gender, family status, socio-demographic status, etc.). They should consider how these dimensions are related to the effectiveness of particular design parameters, and whether changing these parameters might appeal to different segments of the population. To illustrate, if the survey population is known to be very heterogeneous with respect to family status, and if individuals with different family status are believed to strongly prefer different times of survey contact, then the timing of the survey contact should be changed over the survey life cycle to appeal to individuals with different family status.

2.4.2. Delay between contacts

Survey researchers using multiple contacts also need to consider the time delay between the contacts. In one of the few studies on the timing of follow-ups in online surveys, Deutskens et al. (2004) did not find significant differences in the response rate for follow-ups sent after 1 versus 2 weeks using a Dutch consumer sample. Dillman et al. (2009) recommend time lags of about 1 week.

We suggested above that the costs and benefits an individual considers when deciding whether to respond at a particular point in time may be affected by idiosyncratic factors (Fig. 1). Assuming that such factors are correlated over time, longer time lags are more likely to result in a relatively independent draw and may result in higher response rates. For example, if a person was on

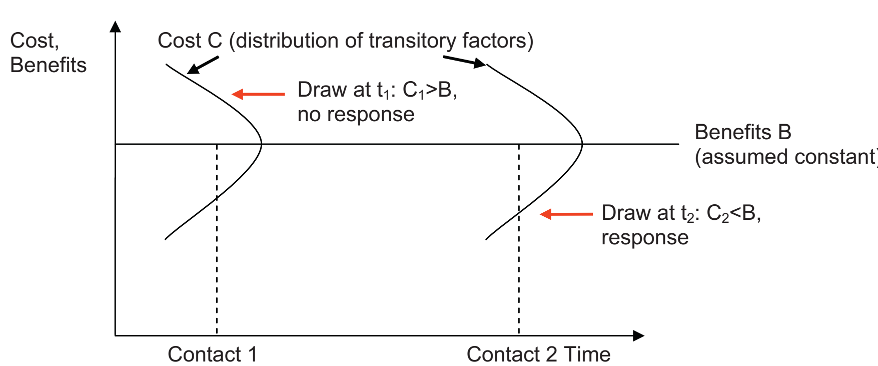

Fig. 1. Changes in "transitory" cost of responding over time.

vacation at time t_1 then she is less likely to still be on vacation 2 weeks later versus 3 days later. On the other hand, assuming that reminders are more effective in building a social exchange relationship than contacts that are perceived as "new", longer time lags may come at a cost if the repeated contacts are perceived as new rather than as reminders. Finally, very short time lags between reminders could be perceived as too "pushy", reducing the likelihood of a response. While it is not clear how these various effects play out, our discussion would suggest nonlinearities such that very short or very long delays may be less effective than medium delays.

2.5. Selective responding and response quality

Increasing response rates is generally desirable, but such increases should not come at the expense of increased nonresponse bias or lower response quality.

Higher response rates are generally thought to decrease nonresponse bias, yet nonresponse bias may actually increase if certain contact design features attract primarily certain subsets of the survey population (Groves et al., 2006; Groves and Peytcheva, 2008). We will consider three potential sources of such "selective responding". First, the use of financial incentives may lead to additional responses primarily from individuals who have stronger preferences for money than the population average, potentially leading to nonresponse bias with respect to preferences for money. The use of lottery incentives may also attract individuals who are less risk-averse (or more risk-loving) than the average person. Finally, we conjecture that the timing of surveys may affect the composition of the final sample with respect to demographic characteristics; in particular, contacts sent on evenings or on weekends may be less likely to yield responses from individuals with family than from individuals without family because the former may be busier at those particular times.

Certain contact design features may also affect the quality of the responses received. For example, Heerwegh (2005) suggests that personalization and associated privacy concerns may lead respondents to skip sensitive questions or to give more socially desirable answers. Similarly, respondents who participate because they are offered a financial incentive may be more likely to skip questions and to spend less time on the survey than respondents who participate because they are interested in the survey topic or who participate to fulfill their part of a social exchange with the researcher. Finally, it is possible that the timing of the survey invitation (e.g., day of the week or time of the day) affects how much time respondents are able to spend working on the survey, with associated effects on data quality.

In the following empirical analysis, we use an experimental design to examine the effects of static and dynamic contact design features on response rates. In addition, we investigate whether our design features lead to selective responding or affect the quality of the survey responses.

3. Sample, experimental design, and empirical strategy

3.1. Sample

We invited a sample of graduate students and postdoctoral researchers at tier-1 U.S. research institutions to participate in a web survey on their organizational environment, work activities, and career aspirations ("Science and Engineering PhD and PostDoc Survey", SEPPS). Details on the substantive content of the survey are provided in Sauermann and Roach (2012).

To obtain a sampling frame, we consulted the National Science Foundation's reports on earned doctorates (2008) and identified U.S. research universities with large doctoral programs in science and engineering fields. We selected a subset of institutions based primarily on program size while ensuring variation with respect to private/public status and geographic region.[4] From this set of over 30 universities we then developed a contact list by hand-collecting names and email addresses from listings provided on departments' websites. The final sample used for this study includes 24,651 individuals, covering 9 broad science and engineering fields. We distributed survey invitations by email using the Qualtrics survey system (see www.qualtrics.com). Each potential respondent received a unique survey link, allowing us to precisely track response behavior over time.

3.2. Experimental design

Each subject was assigned randomly into one of 25 experimental conditions.[5] Regardless of the condition, the survey was available for exactly 60 days from the time of the first contact and each subject received up to three reminder emails for a total of four contacts. Table 1 provides a detailed overview of the experimental conditions.

3.2.1. Static design features

3.2.1.1. Personalization (conditions 1–3). In addition to email addresses, department websites also included the first and last names of individuals. We used these names to create three treatments, differing only in the way subjects were addressed at the beginning of the survey invitation and of the reminders. The "no name" condition did not use any name but addressed subjects as "Dear Researcher". The "first name" condition addressed subjects using their first name only (e.g., "Dear John"), while the "first + last" condition addressed subjects using their full name (e.g., "Dear John Adams"). We used the "first name" personalization as default for all other blocks of conditions (see Table 1) because we expected this condition to yield the highest response rate.

3.2.1.2. Incentives (conditions 4–9). We created six conditions. Subjects in the "no pay" condition were told in the invitation email "We really appreciate your time in answering this survey". Subjects in the five pay conditions were additionally told "As a token of appreciation, we will enter you in the drawing for one of x \$$y$ amazon.com gift certificates upon completion of the survey", where x and y differed across conditions. The five pay conditions each had a total payoff of \$500, but differed in the chance of winning and in the size of the prize (i.e., 100x\$5, 50x\$10, 20x\$25, 10x\$50 and 5x\$100). We chose the particular lotteries to cover a wide range of payoffs and probabilities (the maximum payoffs/probabilities were 20x larger than the minimum values). As is common in lotteries, subjects were not told about the size of the subject pool (nor the number of respondents) and thus had to form a subjective estimate of the chance of winning. In addition to these six incentive conditions, we used a lottery offering 50x\$25 as the default for all other blocks of conditions.

3.2.1.3. Day of the week (conditions 10–16). We created seven conditions in which each contact occurred on the same day of the week,

[4] The largest institutions in our sample include MIT (6.15%), Purdue U (4.81%), U of Washington (4.71%), UC San Diego (4.71%), UC Berkeley (4.5%), Johns Hopkins (4.29%) and U of Illinois Urbana-Champaign (4.1%).

[5] To verify random assignment of subjects, we coded for each case the university, field, and degree status (e.g., PhD, postdoc) based on information from the website. After randomly assigning subjects to conditions, we found no significant differences in these characteristics across conditions. We additionally include these variables as controls in all regression models.

Table 1
Conditions, response rates, and mean comparisons.

Domain	Block	Condition	Cond #	N	Pers.	Incentives	Day of Week	Time of Day	Delay	Word. Chg.	Started	Sig. test (logistic)	Finished	Sig. test (logistic)	Finished/ Started
Static	**Personalization**	No Name (NN)	1	996	NN	50x$25	Sat, Tu, Th, Sun	12, 17, 10, 18	10, 9, 17	Y	**0.253**	Chi^2(2)=	**0.217**	Chi^2(2)=	**0.857**
		First Name (F)	2	992	F	50x$25	Sat, Tu, Th, Sun	12, 17, 10, 18	10, 9, 17	Y	**0.344**	19.46,	**0.293**	15.34,	**0.853**
		First+Last (FL)	3	992	FL	50x$25	Sat, Tu, Th, Sun	12, 17, 10, 18	10, 9, 17	Y	**0.303**	p<0.01	**0.263**	p<0.01	**0.867**
	Incentives	No pay	4	991	F	None	W, Th, Tu, Th	14, 14, 10, 21	8, 12, 21	Y	**0.323**		**0.254**		**0.787**
		Lottery:100x$5	5	983	F	100x$5	W, Th, Tu, Th	14, 14, 10, 21	8, 12, 21	Y	**0.296**	Chi^2(5)=	**0.250**	Chi^2(5)=	**0.845**
		Lottery:50x$10	6	989	F	50x$10	W, Th, Tu, Th	14, 14, 10, 21	8, 12, 21	Y	**0.345**	13.76,	**0.281**	16.61,	**0.815**
		Lottery:20x$25	7	988	F	20x$25	W, Th, Tu, Th	14, 14, 10, 21	8, 12, 21	Y	**0.359**	p<0.01	**0.305**	p<0.01	**0.848**
		Lottery:10x$50	8	976	F	10x$50	W, Th, Tu, Th	14, 14, 10, 21	8, 12, 21	Y	**0.318**		**0.263**		**0.829**
		Lottery:5x$100	9	989	F	5x$100	W, Th, Tu, Th	14, 14, 10, 21	8, 12, 21	Y	**0.357**		**0.311**		**0.873**
	Day of Week	Monday (Mo)	10	991	F	50x$25	Mo, Mo, Mo, Mo	14, 10, 20, 13	7, 7, 14	Y	**0.353**		**0.301**		**0.851**
		Tuesday (Tu)	11	988	F	50x$25	Tu, Tu, Tu, Tu	14, 10, 20, 13	7, 7, 14	Y	**0.347**		**0.294**		**0.845**
		Wednesday (W)	12	988	F	50x$25	W, W, W, W	14, 10, 20, 13	7, 7, 14	Y	**0.319**	Chi^2(6)=	**0.275**	Chi^2(6)=	**0.863**
		Thursday (Th)	13	993	F	50x$25	Th, Th, Th, Th	14, 10, 20, 13	7, 7, 14	Y	**0.366**	6.89, n.s.	**0.305**	4.52, n.s.	**0.835**
		Friday (F)	14	996	F	50x$25	F, F, F, F	14, 10, 20, 13	7, 7, 14	Y	**0.338**		**0.290**		**0.858**
		Saturday (Sat)	15	995	F	50x$25	Sat, Sat, Sat, Sat	14, 10, 20, 13	7, 7, 14	Y	**0.330**		**0.271**		**0.823**
		Sunday (Sun)	16	992	F	50x$25	Sun, Sun, Sun, Sun	14, 10, 20, 13	7, 7, 14	Y	**0.356**		**0.294**		**0.827**
	Time of Day	9 am	17	994	F	50x$25	W, Sun, Th, Tu	9, 9, 9, 9	11, 11, 19	Y	**0.325**	Chi^2(2)=	**0.285**	Chi^2(2)=	**0.876**
		14 pm	18	991	F	50x$25	W, Sun, Th, Tu	14, 14, 14, 14	11, 11, 19	Y	**0.345**	1.54, n.s.	**0.289**	0.23, n.s.	**0.836**
		21 pm	19	992	F	50x$25	W, Sun, Th, Tu	21, 21, 21, 21	11, 11, 19	Y	**0.350**		**0.294**		**0.841**
Dynamic	**Delay**	Long delay	20	991	F	50x$25	W, W, W, W	14, 10, 20, 13	14, 14, 21	Y	**0.324**		**0.273**		**0.844**
	Wording change	No word. change	21	865	F	50x$25	Sat, Th, F, W	14, 10, 20, 13	12, 15, 19	N	**0.240**		**0.207**		**0.861**
Other	Conditions to increase sample size and variation		22	993	F	50x$25	Sat, W, F, W	10, 14, 18, 10	11, 9, 19	Y	**0.334**		**0.287**		**0.858**
			23	995	F	20x$25	Sat, W, F, W	14, 16, 10, 14	11, 9, 19	Y	**0.319**		**0.280**		**0.880**
			24	992	F	50x$25	Sat, W, F, W	16, 10, 14, 19	11, 9, 19	Y	**0.342**		**0.284**		**0.832**
			25	999	F	20x$25	Sat, W, F, W	14, 16, 10, 14	11, 9, 19	Y	**0.310**		**0.271**		**0.874**
Total				24,651							**0.328**		**0.278**		**0.847**

Note: Highlighted cells indicate focal design parameters in a given block of conditions. Entries in the "day of the week" and "time of the day" columns indicate the design parameters chosen for each of the four contacts. Entries in the "delay" column indicate the time (in days) between the four contacts (i.e., 3 time intervals). The entry in column "Word. Chg." indicates whether the wording of contacts was changed in each of the four contacts (Y) or was kept constant (N).

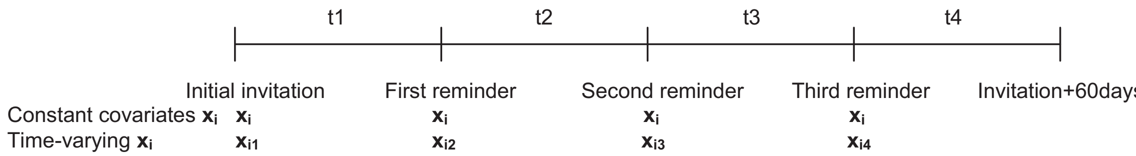

Fig. 2. Basic structure of discrete time hazard analysis.

respectively. For example, in the "Monday" condition, the initial invitation and all three reminders were sent on a Monday.

3.2.1.4. Hour of the day (conditions 17–19). We created three conditions where each of the four contacts was sent at the same time of the day (9 am, 2 pm, and 9 pm Eastern Standard Time, respectively). Given that our subjects live in different U.S. time zones (Eastern, Central, and Pacific Time) the invitation email arrived at three different local times (e.g., an email sent at 2 pm Eastern time will arrive at 1 pm Central time and 11 am Pacific time). Accordingly, our regression models use separate dummy variables for each hour at which a contact arrived (i.e., ranging from 6 am to 9 pm local time).

3.2.2. Dynamic design features

3.2.2.1. Number of contacts. Individuals in all conditions were contacted up to four times (one invitation, three reminders). We examine the benefits of multiple contacts by comparing response rates after the initial contact with those after the second, third, and fourth contact, respectively.

3.2.2.2. Delay between reminders. We varied the number of days between contacts (DELAY, continuous variable) across and within conditions. To examine the effects of long delays between contacts, we additionally created a separate condition "long delay" (condition 20 in Table 1) that is identical to the "Wednesday" condition except that all reminders were sent with an additional delay of 7 days (i.e., after 14, 14, and 21 days versus 7, 7, and 14 days for the "Wednesday" condition).

3.2.2.3. Change in contact wording. We changed the subject line as well as the wording of each reminder without conveying new substantive information for all conditions except condition 21.[6] As control group, condition 21 involved sending the same email in all four contacts.

3.2.2.4. Change in contact timing. To examine the effects of changing the timing of reminder contacts, we changed the contact day of the week and the hour of the day. We implemented this dynamic strategy in all conditions with respect to non-focal design parameters. For example, in each of the seven "day of the week" conditions, we kept the day of the week constant over time but sent each subsequent contact at a different time of the day (see Table 1).

Finally, we used the remaining subjects to add four additional conditions that were not part of any of the blocks discussed above. These conditions were designed to provide additional variation in design parameters that can be exploited in a pooled regression analysis (conditions 22–25, details in Table 1).

3.3. Econometric approach for response rate analysis

Our key dependent variable is survey completion, FINISHED, which equals one if a respondent clicked "next" on the final page of the survey and zero otherwise.[7] For auxiliary analyses, we also use STARTED, a dummy indicating whether a subject started the survey by clicking on the survey link.

As a first test of the effects of the static design parameters, we compare the response rates across conditions within each block (conditions 1–3, 4–9, 10–16, and 17–19 in Table 1). More specifically, we estimate logistic regressions for each block, using FINISHED or STARTED as dependent variables and including dummy variables indicating the experimental conditions. The resulting test statistics are reported in Table 1 and indicate whether there are significant differences in response rates across the experimental conditions within a given block.

We then estimate a series of discrete time hazard models to examine the effects of static as well as dynamic design features. For that purpose, we define four time periods (analysis time) based on the sequence of four survey contacts and a total survey window of 60 days. While some independent variables (e.g., incentives) remain constant over time for a given individual, others such as day of the week or hour of the day may vary across the four time periods (reflecting a "dynamic strategy"). For the analysis, the original 24,651 individual observations are then transformed into person-round observations, which are regressed on independent variables as well as round dummies using logistic regression (Allison, 1982).[8] Fig. 2 illustrates the setup of this analysis.

All regressions include controls for university fixed effects, field fixed effects, and subjects' degree status (e.g., postdoctoral researchers versus PhD students), which we derived from information available on university websites. In addition, we also coded a dummy variable that equals one if a contact was sent at the time that the subject's home institution was on spring break.

[6] For example, we changed the beginning of the email from "We are writing to ask for your participation in a research study on science and engineering careers" (initial contact) to "We recently asked you to participate in a brief research survey regarding the work experiences and career choices of junior scientists and engineers." (reminder 1) to "We ask you to please consider again participating in the UNC/Georgia Tech Science & Engineering survey." (reminder 2) to "We would like to make one final request for your participation in the national Science & Engineering Career Survey." (reminder 3).

[7] The survey used a paging design, placing a small number of questions on each page and asking respondents to click "next" to proceed to the next page. Respondents saw approximately 35 pages (depending on skip patterns) and took about 20 min to complete the survey.

[8] Survey response behavior could also be analyzed using continuous time models such as the Cox proportional hazards model. We chose discrete time models for several reasons. First, even though responses could theoretically occur continuously over time, they were heavily concentrated soon after a given contact, resulting in a relatively discrete response pattern (see discussion below and Fig. 3). Moreover, our focus is on the effect of experimental manipulations on the likelihood of a response at any time during a given round, and discrete time models provide a more parsimonious approach to examine this question than continuous time models that also consider the timing of responses within a given round. Finally, discrete time analysis is complicated since time itself is an experimental manipulation in our study (i.e., different delays of reminders); as a consequence, the baseline hazards of the different conditions cannot be assumed to have the same shape. We examined the robustness of our results by running selected models using both discrete and continuous time regressions and find very similar results.

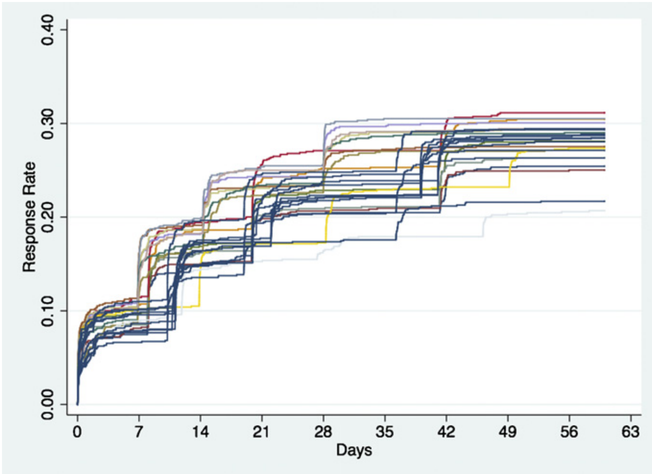

Fig. 3. Response rates (FINISHED) by condition. *Note*: for illustration only; please refer to Table 1 for a complete listing of conditions and response rates.

4. Results

Fig. 3 illustrates the development of (unadjusted) response rates over the 60 days for each condition and shows that final response rates varied considerably, ranging from 20.7% to 31.1%.[9] Fig. 3 also illustrates the different time delays between contacts and shows that responses tended to occur quickly after a particular contact. Across all conditions, we received 50% of responses within 4 h of a contact and 90% within 66 h. Each of the four contacts significantly increased response rates by an average of 9.6, 7.7, 5.6, and 4.9 percentage points, respectively. Table 1 lists the blocks of experimental conditions with their respective response rates (STARTED and FINISHED). Across all conditions, the mean of STARTED is 32.8% and the mean of FINISHED is 27.8%. The share of respondents who finished the survey after clicking on the survey link is 84.7%.

Table 2 reports the regression results. All coefficients in Table 2 are reported as odds ratios; odds ratios greater than 1 indicate a positive effect, odds ratios smaller than 1 indicate a negative effect.

4.1. Static design features

Models 1–5 in Table 2 estimate models separately for the blocks of conditions designed to examine the static design factors.

4.1.1. Personalization

Model 1 shows that personalization significantly increased the odds of a response. As predicted, using the first name only resulted in a higher response rate than using the full name (48% versus 24%; $Chi^2(1) = 3.73$, $p = 0.053$). This result suggests that a very high degree of personalization in a formal sense (i.e., first and last name) may be perceived as less personal than the first name alone.

4.1.2. Incentives

As predicted, the "no pay" and "100*x*$5" conditions had the lowest response rates; the response rate was highest for the condition with the largest prize and the lowest chance of winning (odds ratio of 1.32, $p < 0.01$). Model 3 collapses all five incentive conditions and shows that providing incentives resulted in higher response rates than not providing incentives. However, a closer inspection shows a nonlinear pattern, with no significant effect of the 10*x*$50 lottery. To complement this analysis, model 7 uses the full sample and also includes cases that were offered a sixth lottery (50*x*$25; i.e., the same size of the prize as the 20*x*$25 lottery, but a larger number of prizes). We find that this lottery significantly increases response rates compared to no incentives (odds ratio 1.265). While the effect is somewhat larger than that of the 20*x*$25 lottery, the difference is not significant ($Chi^2(1) = 0.13$, $p = 0.72$). One possible interpretation is that respondents pay more attention to the size of the prize than to the number of available prizes, perhaps because there is no information on the number of participants and thus on objective probabilities. Additional work is needed to separate the effects of chance and the size of the prize; however, our results suggest that a fixed budget for lottery prizes is more effective if used for a small number of large prizes than for a large number of small prizes.

4.1.3. Day of the week

We find no significant differences in response rates across days of the week, although Wednesday and Saturday seem to be slightly worse than the other days (model 4). The absence of an effect of timing could reflect that subjects do not perceive significant differences in the costs or benefits of responding on different days. Alternatively, it is possible that there are such differences in costs or benefits but that subjects who receive the survey at an inconvenient time postpone their response until they have time to respond. To examine the latter possibility, we compared the percentage of respondents who completed the survey on the day of the contact across conditions. Fig. 4 shows that subjects were less likely to respond on the same day if an invitation arrived on the weekend versus a weekday. These differences across days are statistically highly significant ($F(6) = 12.20$, $p < 0.001$), suggesting that weekends are less convenient for subjects. However, rather than decline participation if approached on a weekend, respondents postponed their participation, resulting in no significant differences in response rates across days of a contact.[10]

4.1.4. Time of the day

Model 5 shows no significant differences in response rates across times of the day ($Chi^2(8) = 3.44$, $p = 0.90$). Again, we can examine the timing of responses to understand whether this result reflects that all times of the day are similarly convenient for respondents or whether respondents systematically postpone their response to a more convenient time. We find that the median response delay (time between invitation and response) for the evening emails was approximately 12 h, compared to only 3–4 h for emails sent at other times of the day. Thus, many of the subjects who received their invitation in the evening responded on the next day, leading to longer response delays but no significant difference in final response rates. Model 7 shows the results for the pooled sample (combining conditions 1–25) and allows us to estimate a full set of time dummies. Again, we find no significant differences across times of the day ($Chi^2(15) = 10.52$, $p = 0.79$).

[9] In line with the recommendations by the American Association of Public Opinion Research (2009), our definition of the response rate includes in the denominator emails that were returned as non-deliverable (6.3% of the sample).

[10] We cannot determine when subjects actually read their email. Therefore, our result could reflect that respondents read their email soon after it was received (Sunday) but postponed their answer (Monday). Alternatively, it could also mean that respondents read and answered the email at a later point (both on Monday). Both mechanisms have similar implications for the survey researcher in that contact timing seems to matter little.

Table 2
Discrete time hazard models (logit; odds ratios reported).

	Design feature	Variables	Blocks of conditions						Full sample	Full sample
			1 FINISHED	2 FINISHED	3 FINISHED	4 FINISHED	5 FINISHED	6 FINISHED	7 FINISHED	8 STARTED
Static	Personalization	No name (omitted)								
		First name	1.477**						1.472**	1.520**
		First + last name	1.237*						1.243*	1.229*
	Incentives	No pay (omitted)								
		100*x*$5		1.001					0.995	0.912
		50*x*$10		1.136					1.122	1.088
		20*x*$25		1.272**					1.236**	1.112
		10*x*$50		1.046					1.046	0.979
		5*x*$100		1.316**					1.301**	1.157
		Pay yes (any level)			1.152*					
		50*x*$25							1.265**	1.159*
	Day of week	Monday (omitted)								
		Tuesday				0.992			1.097	1.096
		Wednesday				0.937			0.962	0.954
		Thursday				1.040			0.966	0.999
		Friday				0.968			0.977	0.967
		Saturday				0.886			0.892	0.915
		Sunday				0.986			0.966	0.996
	Time of day	6 am (omitted)								
		7 am							0.836	0.889
		8 am					1.224		0.991	1.049
		9 am					0.867		0.967	1.009
		10 am							0.993	1.027
		11 am					1.012		0.973	1.024
		12 noon							0.970	1.029
		1 pm					1.387		0.935	1.024
		2 pm					0.851		1.010	1.052
		3 pm							1.007	1.052
		4 pm							0.926	0.964
		5 pm							0.976	1.021
		6 pm					0.929		1.013	1.120
		7 pm							0.877	1.049
		8 pm					1.197		0.845	0.979
		9 pm					0.967		1.066	1.138
Dynamic	Delay	Delay: short (omitted)								
		Delay: long						0.967		
		Delay (continuous)							0.964	0.981
		Delay_squared							1.001	1.000
	Changes over survey cycle	Change wording							1.364**	1.445**
		Change day							1.052	1.042
		Change hour							1.049	1.064
Controls		Round fe (3)	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
		Degree fe (3)	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
		Field fe (8)	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
		University fe (36)	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
		Spring Break	0.815	1.033	1.030	1.125	1.086	1.636*	1.101	1.125*
		Observations	10,550	20,733	20,733	24,138	10,412	6,878	86,367	84,047
		Chi-square	112.57	132.39	117.67	206.04	111.99	118.93	605.12	655.82
		df	53	56	52	57	57	52	85	85

Note: Models 1–6 are estimated using cases in the experimental conditions designed to study the focal design parameter (see Table 1). E.g., model 1 is estimated using cases in conditions 1–3. Models 7 and 8 use the pooled sample, i.e., conditions 1–25. The unit of analysis is a person-round.
* Significant at 5%.
** Significant at 1%.

4.2. Dynamic design features

The lower portion of models 6 and 7 in Table 2 show the effects of different delays between contacts and of changes in design features.

4.2.1. Delay between contacts

A first analysis compares conditions 12 and 20, which were exactly the same except that condition 20 (“long delay”) had an additional week between each contact. Model 6 shows no significant difference in response rates between the two conditions. To explore potential nonlinear effects, we include a continuous variable DELAY (in days, ranging from 7 to 21) as well as DELAY_SQUARED in the pooled regression (model 7). We find no significant effects, suggesting that – at least within the range of our DELAY measure (7–21 days) – the delay between contacts does not affect response rates.

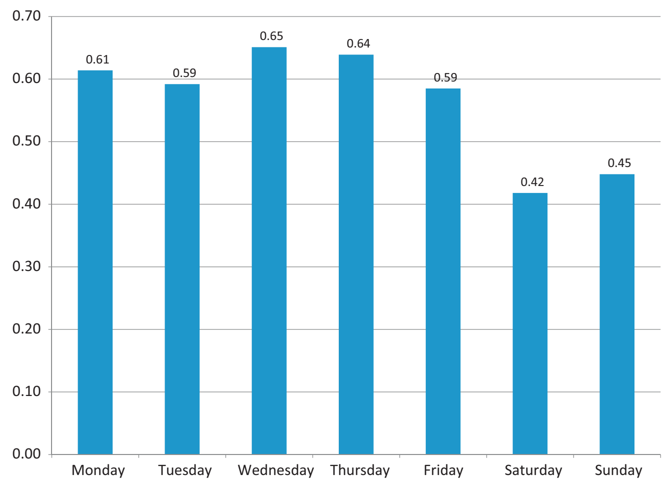

Fig. 4. Share of responses received (FINISHED) on the day of the contact.

4.2.2. Changes in contact wording

Changing the wording of reminder emails significantly increased response rates, leading to an overall increase in the odds of a response by 36% (model 7). Given that the objective informational content of the different contacts was virtually the same, we suggest that changes in wording resulted in a higher response rate primarily because subjects were more likely to interpret contacts as sent by a "real" researcher who invested effort in the relationship. In times of large amounts of spam and zero marginal mailing cost of sending an additional reminder, the effort going into writing a new email may serve as a signal to respondents that sets serious researchers apart from "spammers".

4.2.3. Changes in timing of contacts

Recall our earlier finding that day of the week and hour of the day had little effect on response rates. This finding could reflect that timing does not affect whether a particular individual responds, but it would also be consistent with heterogeneity across individuals such that even though different subsets of individuals have strong preferences for different times, aggregate response rates do not differ in a cross-sectional analysis of our conditions because each condition appeals to some (but different) individuals. In the latter case, a dynamic strategy with respect to survey timing should significantly increase response rates. However, model 7 shows no effect of changing contact days of the week or times of the day. Thus, it appears that the timing of contacts really did not matter to potential respondents, perhaps because those who received an invitation at an inconvenient time could respond later.[11]

4.3. Interactions between design parameters

It is conceivable that the benefits of certain design parameters depend on the choices regarding other design parameters. For example, if individuals have different daily routines on weekends versus weekdays, certain times of the day may be more effective on weekends but less effective on weekdays. Although our experimental design is not fully crossed due to the large number of design parameters, we are able to examine some key interactions. To reduce the number of interaction terms involving contact timing, we aggregated the large number of time of day and day of week dummies into a smaller set of variables, distinguishing MORNING (contact before noon), AFTERNOON (contact occurred between noon and 5 pm), and EVENING (contact occurred after 5 pm). Similarly, we distinguish WEEKDAY (contact occurred during a weekday) and WEEKEND.

Our analyses show no significant interactions between personalization and delay, incentives and delay, contact timing and delay, or incentives and contact timing (Table A1 in Appendix A). However, we find a positive and significant interaction between EVENING and WEEKEND (odds ratio 1.21), suggesting that survey invitations sent on a weekend are more effective when sent in the evening compared to the morning (omitted category). Overall, however, these analyses suggest that the effects of design features are largely independent from each other. Thus, while survey researchers have to make decisions about all design features, this task is somewhat simplified because choices regarding one feature can be made more or less independently from choices regarding other features.

4.4. Effects on STARTED

Our empirical analysis focused on the impact of survey design features on survey completion (FINISHED). To supplement this analysis, model 8 in Table 2 provides insights into the effects of design parameters on the likelihood of subjects starting to answer the survey (STARTED). We again find no effect of timing. Similarly, changing contact timing over the survey life cycle does not have an effect. Three interesting effects stand out, however, especially in conjunction with our main analysis. First, personalization has strong positive effects on STARTED and the size of these effects is comparable to that in the FINISHED regressions. Thus, personalization increases response rates primarily by encouraging more individuals to start working on the survey. Second, we find that incentives have only weak effects on STARTED, compared to quite strong effects on FINISHED. Thus, incentives increase response rates primarily by encouraging individuals who have started a survey to actually finish it. Finally, we find that changing the wording of invitation emails strongly increases STARTED, suggesting that the observed positive effect of this dynamic strategy on FINISHED

[11] In model 7, the coefficients on the dummy variables indicating dynamic strategies reflect the average effects of changing design parameters across all rounds. To account for the possibility that the dynamic strategy has greater benefits in later rounds, we interacted the dummy variables indicating changes in wording, day, and time of the day with round dummies (Cleves et al., 2008). Only the interactions with change_wording were significant. More specifically, the odds of a response were higher by a factor of 1.38 in round 2, 2.59 in round 3, and 1.97 in round 4, respectively, when contact wording was changed versus kept constant. Including the interactions did not affect other coefficients and we report the simpler models without interactions. Full results are available upon request.

Table 3
Measures for analysis of selective responding and response quality.

Construct	Measure/survey question
Importance of pay (imppay)	"When thinking about an ideal job, how important is each of the following factors to you: - Financial income (e.g., salary, bonus)" 5-point scale; mean: 3.94
Risk preferences (risklove)	"Imagine you had the choice between winning $1000 for sure or winning $2000 with a 50% chance. Please indicate which option you prefer". Sliding scale from 0 (strongly prefer $1000 for sure) to 10 (strongly prefer 50% chance to win $2000); mean: 3.02
Children (children)	Dummy = 1 if respondent had at least one child under the age of 18; mean: 0.13
Male (male)	Dummy = 1 if respondent is male; mean: 0.62
Item nonresponse (nmis)	Number of missing items out of a set of 20 randomly drawn items (including some items with skip patterns); mean: 4.16
Gave open ended answer (openend)	One survey question asked respondents how much they would hypothetically value publishing when employed in industrial R&D. We then asked respondents why they would (not) value publishing. We coded the variable openend = 1 if respondents entered text into the open-ended answer field, 0 if the field was left empty; mean: 0.43
Time spent on the survey (timespent)	The survey system records survey start and end times; the variable timespent reflects the difference between the two. We exclude cases with a duration of more than 60 min because these individuals likely interrupted their work on the survey. Mean: 23.08
Gave follow up email (gaveemail)	At the end of the survey, we asked respondents to provide an email for a follow-up study in about 2 years. The dummy gaveemail indicates whether a respondent complied with this request, conditional upon getting to the last page of the survey. Mean: 0.81
Socially desirable response (impsoc)	"When thinking about an ideal job, how important is each of the following factors to you: - Contributing to society through my research" 5-point scale We use this measure to examine potential effects of contact design features on socially desirable responding (cf. Moorman & Podsakoff, 1992). More specifically, we suggest that respondents may think of higher ratings on this item as more socially acceptable than low ratings. Mean: 4.13

largely results from increases in the likelihood that a respondent begins to work on the survey.

4.5. Selective responding and response quality

We now examine whether our contact design features selectively appeal to certain types of respondents ("selective responding") [12] and whether design parameters affect the quality of the actual survey responses (Deutskens et al., 2004; Groves and Peytcheva, 2008; Walsh et al., 1992). For this analysis, we use only cases who started the survey.

Ideally, an analysis of selective responding would draw on detailed information on the characteristics of both respondents and non-respondents. Unfortunately, our data on non-respondents is limited to their academic field, university, and degree status. Thus, we examine selective responding indirectly. Our approach is based on the assumption that random assignment into experimental conditions leads to a uniform distribution of individual characteristics such as family status or risk preferences in the *population* of subjects across all conditions. We then examine whether *respondents* differ significantly across conditions with respect to these variables. Any differences in the characteristics of respondents across conditions would suggest that certain contact design features selectively attracted particular types of individuals to respond. For example, any evidence that respondents in the incentive conditions have stronger preferences for risk than respondents in the no-incentive condition would suggest that our lottery incentives selectively attracted risk-loving individuals. Table 3 summarizes the measures of individual characteristics, which are derived from the actual survey responses. Table A2 (Appendix A) shows correlations.

Table 4 shows the results of regressions of the individual characteristics on experimental conditions as well as control variables. The particular types of regressions (e.g., probit, OLS) were chosen depending on the nature of the dependent variable and are indicated in Table 4. The results show no evidence that lottery incentives attract individuals with stronger preferences for money. Similarly, lotteries do not attract individuals with stronger preferences for risk (models 1 and 2). However, there is some evidence that survey timing affects who responds (model 3). More specifically, respondents with children are less likely than others to respond when approached on a Sunday, perhaps reflecting that they are busier on that particular day, presumably spending time with family.

In a second set of regressions, we examine the relationships between our design parameters and the quality of survey responses, using the quality measures described in Table 3. One concern was that financial incentives could attract additional respondents but that these respondents do not take the survey as seriously as respondents who are willing to participate without compensation, potentially resulting in more missing items and lower data quality. Mitigating that concern, model 5 shows that respondents who were offered financial incentives tend to have *fewer* missing items. Model 6 uses only those cases who finished the survey and shows that incentives have no significant effect on item nonresponse. Thus, respondents in the incentive conditions had fewer missing items because they were more likely to actually finish the survey; conditional upon finishing the survey, they skipped about as many items as other respondents. Respondents in the incentive conditions were also just as likely to provide an open-ended answer and spent about the same amount of time on the survey as respondents without incentives (models 7 and 8). Furthermore, respondents in the incentives conditions were significantly more likely to provide a follow-up email address than respondents in the no-incentives condition (model 9). Even though we explicitly asked for the email for follow-up purposes and not for the purpose of sending the gift certificate (we used the original email on file for the latter purpose), respondents may have provided the follow-up email because they assumed it was required for the drawing. Regardless of the underlying mechanism, this result suggests further benefits of using post-paid incentives. Finally, we examined whether personalization leads to differences in the response to a question that may trigger socially desirable response behavior (self-reported preferences for making a contribution to

[12] We thank two anonymous reviewers for this suggestion.

Table 4
Selective responding and response quality.

	Full 1 oprobit imppay	Full 2 ols risklove	Full 3 probit children	Full 4 probit male	Full 5 nbreg nmis	Finished 6 nbreg nmis	Full 7 probit openend	Finished 8 nbreg timespent	Finished 9 probit gaveemail	Full 10 oprobit impsoc
First name	−0.122	0.052	0.226	0.147	0.331	0.070	−0.077	0.009	−0.003	0.054
First + last name	−0.050	0.013	0.015	0.006	0.135	0.773*	−0.190	−0.016	−0.203	0.108
100*x*$5	−0.147	−0.208	−0.232	0.051	−0.472*	−0.793*	0.095	−0.056	0.185	0.063
50*x*$10	0.172	−0.465	−0.115	0.030	−0.167	−0.923	0.060	0.025	0.328*	0.124
20*x*$25	−0.053	−0.362	−0.105	−0.005	−0.587**	−0.134	0.194*	−0.005	0.320**	0.116
10*x*$50	−0.052	−0.174	−0.064	0.156	−0.465*	0.402	0.001	−0.045	0.167	0.267
5*x*$100	0.036	−0.058	−0.226	−0.104	−0.489*	0.106	0.067	−0.040	0.246	0.120
50*x*$25	−0.060	−0.337	−0.034	−0.040	−0.358	−0.031	0.109	0.010	0.252*	0.159
Tuesday	−0.028	−0.304	−0.204	−0.165	−0.249	−0.075	0.110	−0.025	0.071	0.092
Wednesday	0.049	−0.210	−0.183	−0.164	−0.174	−0.014	0.072	−0.010	−0.030	0.096
Thursday	−0.013	−0.319	−0.105	−0.140	−0.144	−0.675*	0.024	−0.036	0.025	−0.054
Friday	0.019	−0.369	−0.021	−0.079	−0.309	−0.536	0.012	−0.035	0.064	−0.105
Saturday	0.108	−0.375	−0.250	−0.157	−0.168	−0.509	0.111	−0.006	−0.041	−0.023
Sunday	−0.008	−0.220	−0.496**	−0.004	−0.087	−0.032	0.158	0.044	0.133	−0.007
6 am	0.008	0.139	0.455*	0.398*	−0.366	0.975	−0.018	−0.063	−0.005	−0.101
7 am	−0.077	−0.280	0.135	−0.047	0.414	0.085	−0.030	−0.001	0.121	−0.139
8 am	−0.093	0.621	−0.038	−0.097	0.129	0.433	−0.132	0.000	−0.159	0.000
10 am	−0.029	0.013	0.087	−0.139	0.177	0.096	0.038	−0.020	0.035	−0.157
11 am	−0.085	−0.058	0.022	−0.064	0.153	0.222	−0.019	−0.030	−0.026	0.011
12 noon	0.006	0.320	0.107	−0.116	0.089	0.027	0.115	0.034	−0.001	−0.119
1 pm	0.007	0.104	−0.061	−0.178*	0.164	0.263	0.071	−0.026	−0.062	−0.049
2 pm	0.061	0.092	−0.034	−0.141	0.038	−0.108	0.019	−0.016	0.076	−0.127
3 pm	−0.035	0.052	0.161	0.000	−0.072	−0.057	0.044	0.025	−0.102	−0.012
4 pm	−0.071	−0.013	−0.143	−0.058	0.376	0.682	−0.146	−0.046	−0.219	−0.208
5 pm	−0.024	−0.321	0.036	0.105	0.303	−0.302	−0.182	−0.007	−0.016	−0.282*
6 pm	−0.076	0.098	0.276	−0.011	0.133	−0.067	0.052	0.000	−0.035	−0.132
7 pm	−0.249*	0.140	0.077	−0.189	0.543*	−0.646	0.084	0.031	0.323*	−0.232
8 pm	0.000	0.081	0.041	−0.102	0.595**	−0.392	−0.081	0.050	0.239	−0.171
9 pm	0.012	0.138	−0.159	−0.036	0.064	0.081	0.096	−0.012	0.052	0.123
Delay	−0.005	−0.119	0.117*	0.067	−0.031	−0.564**	−0.022	−0.005	0.058	−0.014
Delay_squared	0.000	0.003	−0.004*	−0.001	0.001	0.018**	0.001	0.000	−0.003	0.000
Change wording	−0.063	−0.222	0.044	−0.164	0.068	−0.485	0.007	0.098**	−0.07	−0.201
Change day	−0.031	0.117	0.131	−0.012	0.232	0.221	0.041	0.024	0.154*	0.086
Change hour	0.027	0.181	0.079	0.153	−0.011	−0.078	0.068	0.027	0.124	0.221*
Round fe	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Degree fe	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Field fe	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
University fe	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.	incl.
Spring Break	−0.045	0.059	0.011	0.014	0.231	0.419	0.003	−0.043	−0.085	0.172
Constant		4.070**	−2.545**	−0.632	0.792	1.799	0.187	3.025**	0.539	
Observations	6091	6134	5796	6589	6740	5685	6740	5118	5920	2783

Note: Omitted categories: no name; no incentives; Monday; 9 am. Models 6 and 8 are estimated using the sample of respondents who finished the survey. Model 8 excludes individuals with timespent > 60. Model 9 is estimated using the sample of respondents who saw the last page of the survey (which included the request for a follow-up email).

* Significant at 5%.

** Significant at 1%.

society). We do not find evidence that personalization increases social desirability bias (model 10).

5. Summary and recommendations

Online surveys offer significant cost and speed advantages over conventional paper-based surveys. However, response rates tend to be low, limiting statistical power and raising concerns about sample selection bias and representativeness. We develop a generalized cost–benefits framework that explicitly considers heterogeneity in respondents' preferences for various design features. Building on this framework, we discuss potential effects of static contact design features, as well as of a "dynamic strategy" that systematically varies design features over the survey life cycle.

We tested the effectiveness of various design parameters by inviting over 24,000 junior scientists and engineers to participate in a survey on their organizational context, work activities, and career choices. To allow for causal inferences, we employed an experimental approach and randomly assigned subjects to conditions that differed with respect to static and dynamic design features.

Before we summarize our results and conclude with recommendations for survey researchers, it is important to consider the generalizability of our findings. Our sample included scientists and engineers working in the United States, and our results may not necessarily generalize to other populations. For example, while the timing of survey invitations did not have much of an effect in our study, it may be important in general population samples that include individuals who do not have regular internet access. Thus, while our results should be particularly valuable for survey researchers working in the area of science and innovation, future research is needed to examine the effectiveness of the various contact design features in other types of samples. More generally, readers seeking guidance in their survey efforts should consider

Table 5
Summary of results and recommendations for survey design.

Design feature	Definition/conditions	Effect on response rate (change in odds of response) and response quality	Recommendations for survey design
Static			
Personalization	Use name in the invitation; conditions are no personalization, first name only, first and last name	Positive effect on response rate; First and last name: +24% First name only: +48% Effect primarily by encouraging more individuals to start the survey	Personalize survey contacts, but ensure that personalization is appropriate for the specific sample (e.g., need to also consider seniority and cultural background of sample).
Lottery incentive	Offered lottery incentive for survey completion; conditions include no lottery, 100*x*$5, 50*x*$10, 20*x*$25, 10*x*$50, 5*x*$100	Positive effect on response rate; greatest for lottery offering (5*x*$100): +32%. Effect primarily by encouraging individuals to finish surveys they had started No negative effects on response quality	Use post-paid lottery incentives; easier to implement and more cost effective than pre-paid incentives, especially in large samples. For a given budget, a small number of large prizes is likely to be more effective than a large number of small prizes.
Day of week	Day that the respondent received invitation; conditions include all seven days of the week	Day received influences when subjects respond, but not whether they respond (i.e., less likely to respond immediately if received on weekend, but tend to respond on later days). Respondents with children are less likely to respond when approached on Sunday	While all days are similarly effective in terms of overall response rates, Sundays may lead to less representation of respondents with family. Also, consider whether the response delay on weekends may pose a problem (e.g., if substantive responses may differ between weekends and weekdays).
Time of day	Local time that the respondent received invitation; conditions varied from 6 am to 9 pm	Time received influences when subjects respond, but not whether they respond (i.e., less likely to respond immediately if invitation received in evening, but respond on next day)	All days of the week are similarly effective. If samples are very large, sending contacts at different times (and days) may overcome capacity constraints of outgoing servers and reduce the risk that incoming servers (e.g., large universities or firms) flag emails as "spam".
Dynamic			
Number of reminders	One, two, or three reminders after the initial survey invitation	Each of the three reminders significantly increased response rates	Use multiple reminders. Monitor benefits of reminders and consider trade-off between marginal increase in response rate and burden on recipients. Include "opt out" link in every contact.
Delay between contacts	Varied number of days between initial invitation and reminder emails, ranging from 7 to 21 days	No differences in response rates	Reminders between 7 and 21 days are equally effective in increasing response rates. Short delays speed up the data collection process. Short delays may also result in more comparability of responses between early and late respondents (especially if survey is intended to capture constructs that are highly volatile or depend on general time trends). Our results should not be generalized beyond the 7–21 day range; much shorter delays may result in lower response rates.
Change contact wording	Changed the wording of invitation and reminders without conveying new substantive information	Positive effect on response rate: +36% Effect primarily by encouraging more individuals to start the survey	Change the wording of each contact to maintain respondent attention and to signal effort and legitimacy of the survey study.
Change contact day and time	Changed days that respondents received reminders from the day that they received initial invitation. Similarly, changed times of the day between initial invitation and subsequent reminders	No effect on response rates	High response rates can be achieved even if reminders are sent on the same days of the week or same times of the day. However, changing the contact timing may be effective in very heterogeneous samples. We recommend a "dynamic strategy" given this potential upside, given the low implementation cost, and given the lack of a downside.

how their samples compare to ours and think carefully about potential differences in response patterns. The generalized cost–benefits framework outlined in the first part of this paper will be helpful in that effort.

Table 5 summarizes our findings and provides concrete recommendations for researchers seeking to increase response rates to web surveys. Our observations regarding significant benefits of some design parameters suggest effective levers to increase response rates. Our findings that other factors such as the timing of survey invitations matter little are also important; all survey researchers have to make decisions regarding these factors and our results suggest that they may usefully focus their time and effort on optimizing other design parameters that have greater impacts on survey participation.

Acknowledgements

We thank the Georgia Research Alliance/Kauffman Foundation for financial support. We are grateful for extremely helpful comments and suggestions from Chuck Eesley, Mary Frank Fox, David Kirsch, John Walsh, and two anonymous reviewers.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.respol.2012.05.003.

References

Allison, P.D., 1982. Discrete-time methods for the analysis of event histories. Sociological Methodology 13 (1), 61–98.
American Association for Public Opinion Research, 2009. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. AAPOR.
Anderson, C., 2003. The psychology of doing nothing: forms of decision avoidance result from reason and emotion. Psychological Bulletin 129 (1), 139–167.
Anseel, F., Lievens, F., Schollaert, E., Choragwicka, B., 2010. Response rates in organizational science 1995–2008: a meta-analytic review and guidelines for survey researchers. Journal of Business and Psychology 25 (3), 335–349, http://dx.doi.org/10.1007/s10869-010-9157-6.
Azoulay, P., Graff Zivin, J.S., Manso, G., 2011. Incentives and creativity: evidence from the academic life sciences. The RAND Journal of Economics 42 (3), 527–554.
Baruch, Y., Holtom, B., 2008. Survey response rate levels and trends in organizational research. Human Relations 61 (8), 1139.
Bosnjak, M., Tuten, T., 2003. Prepaid and promised incentives in web surveys: an experiment. Social Science Computer Review 21, 208–217.
Brostrom, A., 2010. Working with distant researchers—distance and content in university–industry interaction. Research Policy 39 (10), 1311–1320.
Bruneel, J., D'Este, P., Salter, A., 2010. Investigating the factors that diminish the barriers to university–industry collaboration. Research Policy 39 (7), 858–868.
Camerer, C., Hogarth, R.M., 1999. The effects of financial incentives in experiments: a review and capital–labor-function framework. Journal of Risk and Uncertainty 19, 7–42.
Church, A.H., 1993. Estimating the effect of incentives on mail survey response rates: a meta-analysis. Public Opinion Quarterly 57 (1), 62–79.
Cleves, M., Gould, W., Gutierrez, R., Marchenko, Y., 2008. An Introduction to Survival Analysis using Stata, 2nd ed. Stata Press.
Cobanoglu, C., Cobanoglu, N., 2003. The effect of incentives in web surveys: application and ethical considerations. International Journal of Market Research 45 (4), 475–488.
Cobanoglu, C., Warde, B., Moreo, P.J., 2001. A comparison of mail, fax and web-based survey methods. International Journal of Market Research 43 (4), 441–452.
Cohen, J., 1992. A power primer. Psychological Bulletin 112 (1), 155–159.
Cook, C., Heath, F., Thompson, R., 2000. A meta-analysis of response rates in web- or Internet based surveys. Educational and Psychological Measurement 60, 821–836.
Couper, M., 2000. Web surveys. Public Opinion Quarterly 64, 464–494.
Couper, M., Miller, P., 2008. Web survey methods: introduction. Public Opinion Quarterly 72 (5), 831–835.
Deci, E.L., Koestner, R., Ryan, R.M., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin 125 (6), 627–668.
Deutskens, E., DeRuyter, K., Wetzels, M., Oosterveld, P., 2004. Response rate and response quality of internet-based surveys: an experimental study. Marketing Letters 15 (1), 21–36.
Dillman, D.A., Smyth, J., Christian, L., 2009. Internet, Mail and Mixed-Mode Surveys, 3rd edition. John Wiley & Sons, Inc.
Ding, W., Choi, E., 2011. Divergent paths to commercial science: a comparison of scientists' founding and advising activities. Research Policy 40 (1), 69–80.
Fini, R., Lacetera, N., Shane, S., 2010. Inside or outside the IP system? Business creation in academia. Research Policy 39 (8), 1060–1069.
Fox, C.R., See, K.E., 2003. Belief and preference in decision under uncertainty. In: Hardman, D., Macchi, L. (Eds.), Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making. Wiley.
Frey, B.S., Oberholzer-Gee, F., 1997. The cost of price incentives: an empirical analysis of motivation crowding-out. American Economic Review 87 (4), 746–755.
Fricker, S., Galesic, M., Tourangeau, R., Yan, T., 2005. An experimental comparison of web and telephone surveys. Public Opinion Quarterly 69 (3), 370–392.
Goeritz, A., 2006. Incentives in web studies: methodological issues and a review. International Journal of Internet Science 1 (1), 58–70.
Groves, R., Cialdini, R., Couper, M., 1992. Understanding the decision to participate in a survey. Public Opinion Quarterly 56, 475–495.
Groves, R., Couper, M., Presser, S., Singer, E., Tourangeau, R., Acosta, G., Nelson, L., 2006. Experiments in producing nonresponse bias. Public Opinion Quarterly 70 (5), 720–736.
Groves, R., Peytcheva, E., 2008. The impact of nonresponse rates on nonresponse bias. Public Opinion Quarterly 72 (2), 167–189.
Groves, R.M., Heeringa, S.G., 2006. Responsive design for household surveys: tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society: Series A (Statistics in Society) 169 (3), 439–457.
Haeussler, C., 2011. Information-sharing in academia and the industry: a comparative study. Research Policy 40 (1), 105–122.
Haeussler, C., Colyvas, J.A., 2011. Breaking the Ivory Tower: academic entrepreneurship in Germany and the United Kingdom. Research Policy 40 (1), 41–54.
Heerwegh, D., 2005. Effects of personal salutations in invitations to participate in a web survey. Public Opinion Quarterly 69 (4), 588–598.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–292.
Kaplowitz, M.D., Hadlock, T.D., Levine, R., 2004. A comparison of Web and mail survey response rates. Public Opinion Quarterly 68 (1), 94–101.
Lacetera, N., Macis, M., 2010. Do all material incentives for pro-social activities backfire? The response to cash and non-cash incentives for blood donations. Journal of Economic Psychology 31 (4), 738–748.
Lazear, E.P., 2000. Performance pay and productivity. American Economic Review 90 (5), 1346–1361.
Moorman, R.H., Podsakoff, P.M., 1992. A metaanalytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behavior research. Journal of Occupational and Organizational Psychology 65, 131–149.
National Science Foundation, 2008. Survey of Earned Doctorates.
Piazza, T., 1993. Meeting the challenge of answering machines. Public Opinion Quarterly 57 (2), 219–231.
Porter, S., 2004. Raising response rates: what works? New Directions for Institutional Research 121, 5–21.
Porter, S., Whitcomb, M., 2003a. The impact of contact type on survey response rates. Public Opinion Quarterly 67, 579–588.
Porter, S., Whitcomb, M., 2003b. The impact of lottery incentives on student survey response rates. Research in Higher Education 44 (4), 389–407.
Rogelberg, S., Stanton, J., 2007. Understanding and dealing with organizational survey nonresponse. Organizational Research Methods 10, 195–209.
Roth, P., BeVier, C., 1998. Response rates in HRM/OB survey research: norms and correlates, 1990–1994. Journal of Management 24 (1), 97–118.
Rynes, S.L., Gerhart, B., Parks, L., 2005. Performance evaluation and pay for performance. Annual Review of Psychology 56, 571–600.
Sauermann, H., Cohen, W., 2010. What makes them tick? Employee motives and industrial innovation. Management Science 56 (12), 2134–2153.
Sauermann, H., Cohen, W.M., Stephan, P.E., 2012. Complementing Merton: the motives, incentives, and innovative activities of academic scientists and engineers, Working Paper.
Sauermann, H., Roach, M., 2012. Science PhD career preferences: levels, changes, and advisor encouragement. PLoS ONE 7 (5), e36307.
Shannon, D.M., Bradshaw, C.C., 2002. A comparison of response rate, response time, and costs of mail and electronic surveys. The Journal of Experimental Education 70 (2), 179–192.
Simsek, Z., Veiga, J., 2001. A primer on Internet organizational surveys. Organizational Research Methods 4 (3), 218–235.
Van Looy, B., Landoni, P., Callaert, J., van Pottelsberghe, B., Sapsalis, E., Debackere, K., 2011. Entrepreneurial effectiveness of European universities: an empirical assessment of antecedents and trade-offs. Research Policy 40 (4), 553–564.
Wagner, J.R. 2008. Adaptive Survey Design to Reduce Nonresponse Bias. Unpublished Doctoral Thesis, University of Michigan.
Walsh, J., Kiesler, S., Sproull, L., Hesse, B., 1992. Self-selected and randomly selected respondents in a computer network survey. Public Opinion Quarterly 56, 241–244.
Weeks, M., Jones, B., Folsom, R., Benrud, C., 1980. Optimal times to contact sample households. Public Opinion Quarterly 44 (1), 101–114.
Yammarino, F., Skinner, S., Childers, T., 1991. Understanding mail survey response behavior: a meta-analysis. Public Opinion Quarterly 55 (4), 613.