DISCUSSION PAPER

// JOHANNES J. GAUL, FLORIAN KEUSCH, DAVUD ROSTAM-AFSCHAR, AND THOMAS SIMON

Invitation Messages for Business Surveys: A Multi-Armed Bandit Experiment

Leibniz Association

ZEW

Invitation Messages for Business Surveys
A Multi-Armed Bandit Experiment[*]

Johannes J. Gaul[†] Florian Keusch[‡] Davud Rostam-Afschar[§] Thomas Simon[¶]

December 11, 2024

Abstract

This study investigates how elements of a survey invitation message targeted to businesses influence their participation in a self-administered web survey. We implement a full factorial experiment varying five key components of the email invitation. Unlike traditional experimental setups with static group composition, however, we employ adaptive randomization in our sequential research design. Specifically, as the experiment progresses, a Bayesian learning algorithm assigns more observations to invitation messages with higher starting rates. Our results indicate that personalizing the message, emphasizing the authority of the sender, and pleading for help increase survey starting rates, while stressing strict privacy policies and changing the location of the survey URL have no response-enhancing effect. The implementation of adaptive randomization is useful for other applications of survey design and methodology.

Keywords: Adaptive Randomization, Reinforcement Learning, Nonresponse, Email Invitation, Web Survey, Firm Survey, Organizational Survey

JEL Classification: C11, C44, C93, D83, M00, M40

1   Introduction

Voluntary business surveys, like household surveys, often suffer from low participation. Even worse, recent studies indicate a downward trend in response rates for business surveys (König et al., 2021; Küfner et al., 2022; Pielsticker and Hiebl, 2020). This decline must be considered in light of the growing reliance on web-based surveys. For researchers, web surveys offer significant advantages—such as speed, flexibility, scalability, and cost-efficiency—but they are also particularly prone to lower response rates compared to other survey modes (Manfreda et al., 2008). Despite the opportunities associated with web surveys in business contexts, such as an increasing digitization of enterprises and higher web literacy among respondents, the limitations of web surveys remain notable (Daikeler et al., 2020). These limitations include the impersonal nature of online survey requests (Evans and Mathur, 2005), privacy concerns (Sax et al., 2003), organizational gatekeepers restricting access to decision-makers (Snijkers et al., 2013), and the complex, time-consuming response process (Haraldsen, 2018), which often results in a lack of capacity, motivation, or authorization to engage with the survey request (Langeland et al., 2023).

Low survey participation can be problematic as it may be linked to non-response bias (König and Sakshaug, 2023) and counteracting low response rates, for example, by increasing sample sizes, can significantly drive up costs and increase bias. Thus, researchers are encouraged to use cost-effective measures to increase response rates while being mindful not to introduce bias. Naturally, much of the focus has been on optimizing the content and layout of the questionnaires themselves. Yet, effective communication with potential respondents is also crucial for encouraging response due to a general lack of prior connection between a potential respondent and the organization conducting the survey (Einarsson et al., 2021). In this context, the design and layout of the survey invitation message are particularly important. If the invitation fails to spark the recipient's interest in responding, the best questionnaire design will be of no use.

Consequently, a large body of literature has tested the impact of several characteristics of invitation messages on participation behavior in self-administered surveys (e.g., Heerwegh and Loosveldt, 2002; Kaplowitz et al., 2012; Trespalacios and Perkins, 2016). However, these studies predominantly focus on household surveys, and their findings may not extend to the context of businesses due to their specific dynamics and unique constraints (Langeland et al., 2023). First, the flow of information within organizations is more complex than in household settings. There may be multiple potential respondents within a single organization, making it necessary to identify the most appropriate individual to complete the survey. In addition, the release of data often requires proper

authorization, which may involve coordination across different departments or hierarchical levels. Second, the cost-benefit analysis for businesses differs from that of individuals. While the benefits of participation tend to be less obvious, the costs are immediate and include the time and resources devoted to the survey as well as the opportunity cost of diverting attention from core business activities. Third, the response process itself can be intricate, sometimes requiring calculations or the retrieval of information that is not readily accessible, and may necessitate collaboration among multiple people to gather the necessary data.

The peculiarities of the business context, which make insights from household surveys not automatically transferable, must be addressed already in the initial communication with potential respondents, even if these peculiarities primarily affect the survey response process (Snijkers et al., 2013). Yet, determining the most effective way to phrase and design an invitation message for businesses remains an open empirical question, especially in light of the lack of any unifying theory of survey participation (Dillman, 2021; Keusch, 2015). Although various theoretical perspectives, such as social exchange theory, cognitive dissonance theory, and leverage-salience theory, suggest that certain elements of an invitation message may enhance participation, their relevance has been inconsistently assessed by survey methodologists (Keusch, 2015). Furthermore, traditional survey invitation message experiments test only a small set of alternatives, limiting conclusions about interaction effects and alternative design choices. At the same time, it is often prohibitively costly—given fixed terms and limited budgets—for experimental approaches to find the best message design choice from a comprehensive set of alternatives.

This study addresses these gaps by investigating how to design a survey invitation message that specifically targets business decision-makers and by offering an approach to mitigate key challenges in survey production. To this end, we conduct a sequential experiment that varies the content and layout of survey invitation messages issued by the German Business Panel (GBP). The GBP systematically surveys a probability sample of all legal entities operating in Germany. Over the course of our 15-week experimental phase, 176,000 firms opened their GBP survey invitation email, enabling a large-scale experimental setup. Our research design varies five key elements of the survey invitation message: (1) personalization, (2) emphasis on the authority of the sender, (3) survey link placement, (4) compliance with data protection, and (5) distinct request styles, either offering potential survey respondents the opportunity to share their opinion or issuing a plea for help. We implement each of these items in two different ways, resulting in an experiment with a full-factorial total of 32 ($= 2^5$) unique message alternatives.

One central innovation in our experimental design is the application of a reinforcement learning algorithm. This algorithm continuously adjusts the experimental group sizes toward better performing invitation messages—those that generate higher starting rates—unlike traditional setups with static group composition. As the experiment progresses, we apply a Bayesian decision rule known as *randomized probability matching*, which incrementally allocates more observations to the invitation messages with the highest likelihood of being optimal (Scott, 2010). To maximize output, the basic principle of this adaptive randomization, often referred to as multi-armed bandit (MAB) optimization, is straightforward: Allocate more efforts (observations) to actions (invitation message alternatives) that appear most rewarding.

Our results underscore the significant impact of invitation message design on participation rates in web surveys targeting business decision-makers. We find that our most effective survey invitation message can increase starting rates by up to 44%. Specifically, we demonstrate that personalization, stressing the authority of the invitation's sender, and pleading for help result in higher survey starting rates. By contrast, placing the URL near the top of the message and emphasizing strict data protection policies do not seem to significantly benefit survey participation. The latter result is particularly surprising as it suggests that an extended data protection statement does not unanimously heighten the trustworthiness of the survey invitation. Instead, extensive privacy assurances may also raise concerns about sharing sensitive information, reducing recipients' willingness to respond in the first place. Moreover, by employing adaptive randomization we were able to increase the number of survey starts by 6.66% compared to a traditional fixed-group design, illustrating the practical benefits of this dynamic approach in optimizing survey output.

We support our main findings with a series of supplementary analyses. First, we examine the effect of different invitation messages on the likelihood that a recipient not only starts but also completes the questionnaire. We demonstrate that once a respondent has begun answering the questionnaire, there is no significant difference in completion rates based on the invitation message received. Thus, the impact of optimizing survey invitations extends beyond simply increasing the survey starting rate and also increases the completion of the survey. Second, we investigate whether responses to the invitation messages vary based on firm size. Our analysis reveals that smaller businesses are particularly responsive to pleading frames and invitations from authoritative sources. By contrast, larger businesses exhibit higher starting rates when the survey is personalized, but emphasizing the GBP's compliance with privacy regulations appears to deter their engagement. Despite these heterogeneous treatment effects, we confirm that our experiment did not introduce

size-related nonresponse bias. Specifically, we find no significant difference in staffing levels between respondents and non-respondents across the different experimental treatments. Third, we show that our main findings are robust to the exclusion of participants from previous surveys conducted by the GBP. Finally, to address potential concerns about bias in the standard errors due to the adaptive randomization, we validate our findings by employing batched regressions (Zhang et al., 2020; Kemper and Rostam-Afschar, 2024).

We make several contributions to the literature. First, we provide evidence from a full-factorial experiment with 32 invitation messages. This allows us to simultaneously compare five individual message elements and to study their interaction effects. By contrast, prior studies have considered a maximum of 16 invitation message alternatives (Kaplowitz et al., 2012). Additionally, we examine the effect of emphasizing compliance with data privacy regulations, an area with limited prior evidence despite its relevance for any self-administered survey. Second, we respond to calls for experimental research in establishment surveys, given that findings from household surveys do not necessarily generalize to business contexts (Langeland et al., 2023). In this regard, we offer novel insights, demonstrating, for example, that personalization can enhance response rates even when using business names rather than person names. Third, we show how MAB optimization can be effectively implemented to increase survey starting rates by dynamically optimizing the allocation of invitation messages. This research design is made possible by the staggered roll-out of invitation messages, which enables a regular adjustment of the randomization scheme. In doing so, we provide a practical example of how MAB optimization can readily be applied in survey research, showcasing an approach that is transferable to other settings. For instance, this method may be used when the ex-ante elimination of alternative questions is not desirable or possible, or when survey content needs to be tailored to the characteristics of (potential) respondents.

We proceed as follows. In Section 2, we describe prior findings on survey invitation message design, while in Section 3 we introduce the concept of MAB optimization. In Section 4, we outline our experimental procedure. Our main findings are presented in Section 5. After providing supplementary analyses in Section 6, we conclude in Section 7.

2   Invitation Messages for Self-Administered Surveys

The decision to participate in a survey is linked to a trade-off between the benefits and costs, both of which can be influenced by the organization conducting the survey (Hill and Willis, 2001). While this cost-benefit framework applies to both household and business surveys, its specifics differ.

In household surveys, benefits might include monetary incentives or the value of a personalized, engaging survey experience (Fan and Yan, 2010). To reduce response burden in household surveys, organizations should carefully monitor factors such as the inclusion of sensitive or cognitively challenging questions and the length of the questionnaire (Hill and Willis, 2001). By contrast, business surveys rely less on monetary incentives, typically offering largely indirect benefits tied to broader economic mechanisms (Snijkers et al., 2013). Beyond the resources required to complete the survey, response burden for businesses often stems from complex computations, a lack of competence or authorization to answer certain questions, and, similar to household surveys, the overall length of the questionnaire (Haraldsen, 2018).

Beyond the characteristics of the questionnaire itself, which may alter the actual costs and benefits of participation, the design of survey invitations plays a critical role in shaping the *expected* costs and benefits perceived by potential respondents (Fan and Yan, 2010). Prior research from household surveys suggests that optimizing the structure and phrasing of survey invitations represents an effective strategy for increasing response rates (Keusch, 2015). Key design choices include personalizing invitation messages, emphasizing the power and status of the sender, placing the survey link strategically, stressing compliance with data protection, and adopting a distinct request style. In the following, prior research on each of these elements is reviewed and contextualized within the business setting.

Personalization.   Personalizing survey invitations has been widely investigated in household and student contexts, across different survey modes. In these studies, personalization is typically operationalized by addressing the recipient by their first and/or last name. In mail surveys, this approach has been shown to raise response rates (Heberlein and Baumgartner, 1978; Yammarino et al., 1991). Yet, evidence from web-based invitations is mixed. Several studies suggest that personalization enhances participation (Cook et al., 2000; Heerwegh, 2005; Heerwegh and Loosveldt, 2006, 2007; Heerwegh et al., 2005; Joinson and Reips, 2007; Muñoz-Leiva et al., 2010; Sánchez-Fernández et al., 2012; Sauermann and Roach, 2013), however, others do not (Kent and Brandal, 2003; Porter and Whitcomb, 2003; Trespalacios and Perkins, 2016; Wiley et al., 2009). Positive effects of personalization are often linked to reciprocity, where respondents feel a need to reciprocate a personal salutation (Dillman, 2007), or responsibility theory, which suggests that individuals feel more obligated to help if they perceive others as less able (Barron and Yechiam, 2002).

While these findings are already inconclusive, their applicability to business contexts remains uncertain. In business surveys, the flow of information is more complex: Gatekeepers often control

access to potential respondents, and multiple individuals within an organization may be qualified to answer the survey (Snijkers et al., 2013). This makes respondent selection less straightforward and typically beyond the direct control of the survey organization, complicating the use of personal names in salutations.[1]

Authority. Digital exchanges are typically characterized by high levels of anonymity. In the context of survey invitations, a lack of social information about the sender could lead to widespread disregard of the request. Stressing the high social status of the sender, however, may serve as a heuristic for decision-making, where recipients are more likely to trust and comply with requests from authoritative sources (Cialdini, 2001). While several studies identify a positive response rate effect of making authority salient in survey invitation messages targeting individuals and households (Guéguen and Jacob, 2002; Kaplowitz et al., 2012), others indicate no beneficial impact of emphasizing the high social status of the invitation's sender (Heerwegh and Loosveldt, 2006; Petrovčič et al., 2016; Porter and Whitcomb, 2003). In the business context, authority can stem from the mandatory nature of certain surveys conducted by governmental institutions or national statistical institutes. However, in voluntary surveys—such as the one used for this experiment—the effect of authority signals, like the sender's status or institutional power, remains an open question. While authoritative senders might increase the sense of obligation to participate as a contribution to the broader business community, businesses also perceive participation as a nonproductive cost (Snijkers et al., 2013), making the influence of authority uncertain.

URL Position. To increase participation in web-based surveys, organizations are advised to lower participation hurdles as much as possible (Crawford et al., 2001). If URL links are integrated into the invitation message, links with embedded identifiers are, thus, considered to achieve a higher willingness to participate as compared to links demanding the (burdensome) manual entry of an identification or password (Crawford et al., 2001). Beyond the URL's structural form, research is concerned with the optimal placement of the participation link within the invitation message. On the one hand, Couper (2008) recommends placing the URL near the top of the invitation to avoid requiring recipients to scroll down for accessing the survey. For businesses, where costs are an immediate concern, a top-placed URL may seem advantageous by reducing perceived effort. On the other hand, Kaplowitz et al. (2012), in a university context surveying students and faculty, find

---

[1]In a small-sample study, Ramirez (1997) addressed this issue by using telephone pre-contacting to *personally identify* and target the most suitable respondents within organizations, which subsequently increased starting rates. However, in large-scale surveys this approach seems not viable.

that placing the URL at the bottom of the invitation is associated with increased participation, as this placement encourages potential participants to read more of the invitation, thereby increasing the perceived trustworthiness of the request. In business surveys, building trust is crucial, as companies often question why they have been selected for participation and how their data will be used (Snijkers et al., 2013). A later URL placement that prompts careful reading of the invitation may help alleviate these concerns.

Data Protection. Emphasizing strict data protection policies could also enhance the trustworthiness of a survey invitation. Information systems research indicates that the presence of a privacy statement on a firm's website increases the likelihood that customers will share personal information (Hui et al., 2007). Similarly, Al-Natour et al. (2020) show that privacy uncertainty is negatively associated with consumers' intention to download an app. If these observations apply to invitations for voluntary business surveys, highlighting the survey organization's strict data protection policies could encourage firms to participate. However, stressing data protection may also alert firms to the possibility that they could be sharing proprietary or sensitive information, potentially reducing their willingness to respond. Additionally, particularly in larger organizations, emphasizing data protection may introduce (perceived) procedural hurdles, as decisions about who is authorized to respond can delay or prevent the invitation from reaching the appropriate recipient, ultimately lowering response rates.

Offer vs. Plea Framing. Finally, survey participation behavior may be influenced by the way participation requests are framed. Researchers typically distinguish between an 'offer' frame, which invites recipients to share their opinions, and a 'plea' frame, which appeals for assistance. Considering the importance of the social norm of helping in online communities (Tanis, 2007), several studies suggest that framing participation requests in a way that leverages recipients' inherent willingness to assist can increase response rates in household surveys (Petrovčič et al., 2016; Porter and Whitcomb, 2003; Trouteaud, 2004). Yet again, this finding is not unanimously supported. For instance, Felix et al. (2011) find no significant differences in survey starting rates depending on the framing of the invitation. In business surveys, the relevance of the social norm of helping is less clear, as firms tend to view participation more transactionally, perceiving it as an investment in the response process that needs to pay off (Snijkers et al., 2013).

In sum, prior evidence on optimal survey invitation message design is largely inconclusive. Additionally, whether emphasizing a strict adherence to data protection rules increases survey participation

has—to the best of our knowledge—not yet been examined. While existing studies on survey invitation design are predominantly focused on household or student surveys, the GBP targets business decision-makers of firms operating in Germany, whose participation decisions likely differ substantially from those of individuals approached in their private sphere (Snijkers et al., 2013). Given the absence of a unifying theory on survey participation (Keusch, 2015), determining which elements of invitation messages in the context of business surveys influence participation rates remains an open empirical question.

3   Adaptive Randomization in Experiments

All prior studies examining the impact of invitation message design on survey response rates share a feature in their research design: They are carried out as non-adaptive experiments using fixed and balanced randomization. This approach is characterized by an experimental phase consisting of plain exploration, i.e., learning about the effectiveness of different treatment conditions. Crucially, any of the information that is gathered during the experiment remains disregarded for the purpose of the ex-ante defined randomization strategy, and the exploitation of potentially identified treatment effects, i.e., earning on the knowledge that has been gained, only occurs after the experiment has concluded.[2] This setup reflects a common evaluation criterion in research, where experiments are typically deemed efficient based on their *statistical power* to detect treatment effects (Breur, 2016).

Yet, in case of sequential experiments, sticking with the maxim of fixed and balanced randomization might sacrifice output, in our context survey starts, for the sake of exploration. Even if preliminary results clearly point to superior performance in one experimental group early in the trial, this data is typically not leveraged until after the experiment has concluded. This lack of flexibility and the disregard of interim results make traditional experiments less attractive for organizations and firms, which tend to evaluate experiments based on their *outputs* not only after, but also during the trial phase (Kaibel and Biemann, 2021).

In contrast to experiments using treatment and control groups with pre-determined, fixed sizes, response-adaptive randomization addresses these concerns by merging the exploration and exploitation phases of an experiment, so that any knowledge gathered during the experiment is already taken into account while it is still ongoing. To maximize some outcome, the basic idea of response-adaptive randomization is to assign more observations to treatments that appear most

---

[2]While *exploration* is also used to describe early research phases with undefined research questions or hypotheses, we retain the terms in their standard multi-armed bandit context. In this study, *exploration* refers to the learning phase of the experiment, where the algorithm seeks new information, and *exploitation* denotes the earning phase, where the knowledge gain is applied.

rewarding. The challenge of identifying a sequential randomization scheme that yields the largest possible (desirable) output from a set of potential treatments with unknown reward probabilities has become widely known as *multi-armed bandit* (MAB) problem—a metaphor referring to a gambler in a casino who needs to decide which slot machine to play for maximizing their monetary payoff.[3]

Determining response-enhancing design features in survey invitation messages directed at firm decision-makers represents a prime example of a MAB problem: While the reward probabilities (starting rates) of the individual message versions are unknown ex-ante, a goal of the GBP and, more broadly, of survey organizations is to increase overall survey participation.[4] Rather than relying on traditional static experimental designs, this objective can be achieved through adaptive randomization in sequential setups, for which a vast number of MAB algorithms have been introduced in the literature.[5]

From the array of available MAB algorithms, we implement a policy called *randomized probability matching* to manage the exploration-exploitation trade-off in our experiment. As introduced by Scott (2010), this algorithm belongs to the class of Bayesian decision rules, which have recently gained popularity for solving MAB problems (see, e.g., Ferreira et al. (2018), Kandasamy et al. (2018), or Schwartz et al. (2017)).[6] The general idea behind algorithms that apply a Bayesian decision rule is the following: Assuming some prior distribution on the parameters that characterize the reward distribution of each available option (in this case, invitation message alternatives), at every decision point, these options are chosen according to their posterior probability of being optimal (Agrawal and Goyal, 2012). In our experiment, this implies better-performing messages are sent out more often than those performing poorly. The term 'Bayesian decision rule' originates from the practice of updating the conjectured reward distributions after observing the successes and failures for each option according to Bayes' rule.

Optimizing survey invitations can be modeled as a Bernoulli bandit problem due to the binary nature of the primary outcome measure: Either a firm decision-maker starts the survey in response to receiving a particular invitation message or not. To formalize the concept of randomized probability matching for Bernoulli bandits, consider an experiment with $i \in \{1, \ldots, k\}$ arms that promise ex-

---

[3]When the available slot machines have varying mean reward probabilities, the gambler faces a choice between exploiting the machine with the highest expected payoff and exploring others to learn more about their reward potential. Pulling all available arms at equal rates mirrors an experiment with fixed and balanced randomization. By contrast, an adaptive randomization strategy yields a gambling sequence that increasingly favors machines with the highest expected rewards.

[4]From a bias perspective, solely increasing survey starts might of course not be the only objective. We test for potential bias induced by our randomization procedure in Section 6.2.

[5]Bouneffouf and Rish (2019) provide a detailed discussion of existing classes of MAB algorithms.

[6]Despite this recent trend in using Bayesian MAB algorithms, they have the longest tradition among the group of decision rules available for dealing with MAB problems. In fact, Bayesian decision rules date back to the seminal work by Thompson (1933), which is why they are commonly referred to as *Thompson sampling*.

ante unknown, independent reward probabilities $\theta_i \in \{\theta_1, \ldots, \theta_k\}$.[7] As priors, reward probabilities for each arm are assumed to follow a Beta distribution. Recall that any Beta distribution is defined within the interval $[0, 1]$ and is characterized by two parameters, $\alpha > 0$ and $\beta > 0$. The expected value of a Beta random variable $X \sim Be(\alpha, \beta)$ is given by $\alpha/(\alpha + \beta)$. Thus, the distribution is skewed to the left if $\alpha > \beta$ and skewed to the right otherwise. Larger values of α and β correspond to reduced variance in the distribution.

Prior to the start of the experiment, all arms i are assumed to have Beta priors $\theta_i \sim Be(1, 1)$, representing a uniform distribution over the interval $[0, 1]$. This premise reflects the lack of ex-ante knowledge about the arms' individual reward probabilities. What makes Beta distributions a convenient choice is that they are conjugate to Bernoulli likelihoods, meaning that if you assume a Beta prior and conduct an experiment with Bernoulli outcome, the posterior will also be Beta. More formally, let S_{i_t} denote the number of successes and F_{i_t} the number of failures observed for arm i until period t. The posterior distribution of θ_i is updated as $Be(\theta_i | 1 + S_{i_t}, 1 + F_{i_t})$. Accordingly, the joint posterior distribution of $\theta = (\theta_1, \ldots, \theta_k)$ is

$$p(\theta | \mathbf{s}_t) = \prod_{i=1}^{k} Be(\theta_i | 1 + S_{it}, 1 + F_{it}), \tag{1}$$

where $\mathbf{s}_t = (s_1, \ldots, s_t)$ represents the sequence of rewards observed up to time t across all arms. Following Scott (2010), by way of integration or simulation, the probability that arm i is optimal at time t can be derived as

$$\omega_{it} = \int_0^1 Be(\theta_i | 1 + S_{it}, 1 + F_{it}) \prod_{i \neq j} Pr(\theta_j < \theta_i | 1 + S_{jt}, 1 + F_{jt}) d\theta_i. \tag{2}$$

These probabilities determine the adaptive randomization scheme during the experiment, serving as distribution weights according to which each invitation message alternative is sent out. Their derivation in (2) reflects a simple intuition: While the integral of the Beta posterior for arm i's reward probability density function within $[0, 1]$ naturally equals 1, the probability of selecting arm i—and also its corresponding distribution weight—decreases as the posteriors of other arms indicate higher expected rewards based on their observed successes and failures.

Randomized probability matching is characterized by a number of desirable properties. First, calculating the individual distribution weights is rather simple, requiring only the input of successes

---

[7]In our experiment, invitation message alternatives represent these arms. Their individual starting rates constitute the associated reward probabilities.

(surveys started) and failures for each arm (invitation message alternative).[8] Second, randomized probability matching has been proven to be effective in maximizing the desired outcome. Scott (2010) shows that this decision rule outperforms simpler heuristics and other MAB algorithms. This might be due to the fact that using a Bayesian decision rule reduces the likelihood of becoming trapped with an early bad choice (Agrawal and Goyal, 2012). To further mitigate this risk, it is common to implement Bayesian algorithms with a burn-in phase, during which all arms are chosen at equal rates (Du et al., 2018; Kaibel and Biemann, 2021), and to impose a clipping constraint forcing a minimum distribution weight on each message alternative. By introducing a burn-in phase, the impact of outliers on the resulting sampling scheme is averaged away and statistical power is strengthened (Kaibel and Biemann, 2021).[9] Third, randomized probability matching is compatible with batch updating, i.e., when arms are not chosen one at a time. For these batches, treatment allocation occurs proportionally to the calculated distribution weights. Finally, randomized probability matching does not require discretionary tuning parameters to reduce exploration over the course of the experiment (Scott, 2010). Instead, the randomization scheme evolves endogenously based on each arm's posterior probability of being optimal.

4   Experimental Procedures and Data

Infrastructure. We implement our experiment within the infrastructure of the German Business Panel (GBP).[10] The GBP operates on a recurring six-month cycle, contacting a random sample of firms across Germany for which digital contact information is available. The contact pool is primarily sourced from Bureau van Dijk's Orbis database, a large commercial data platform providing extensive coverage of both public and private firms in Germany. The sampling process follows simple random sampling, and as Bischof et al. (2024) demonstrate, respondent firms closely resemble their population counterparts in key characteristics such as industry, firm size, and legal form. The target population comprises the universe of legal entities included in the official German Statistical Business Register. To address potential coverage error, the GBP provides survey weights calibrated to the marginal distributions of this target population.[11]

---

[8]However, numerical integration must be tested with great care. Some algorithms are approximate methods whose accuracy and stability are not always guaranteed. Our experiment is not affected by such issues during the 15-week analysis period.

[9]Alternatively, to reduce exploitation in favor of exploration, Kasy and Sautmann (2021) propose an algorithm which they call *exploration sampling*. Exploration sampling is based on Thompson sampling but replaces the Thompson distribution weights with transformed weights. This modification moves weight of the best performing arm to its closely competing arms.

[10]For a detailed description of the GBP's objectives, survey methodology, and scope, see Bischof et al. (2024).

[11]The GBP's *survey weights*, applied ex-post to adjust the respondent sample to the target population based on key characteristics, are not utilized in our experiment. These weights should not be confused with the *distribution*

For its web survey, the GBP engaged its entire contact pool exclusively via email, with invitations distributed daily over the course of each six-month interval, excluding weekends and public holidays. Each workday, a randomly drawn subset of firms is contacted, and non-respondents receive reminders after seven, 14, and 28 days. Importantly, firms are only contacted once per wave and are not reintroduced into the contact pool during the same period. The survey infrastructure of this study is fully web-based.[12] Thus, answering the survey is self-administered, typically taking respondents between five to 15 minutes to complete. No monetary incentives are provided for participation. Broadly speaking, the GBP's goal is to gather insights from firm decision-makers on topics related to accounting, taxation, and transparency regulation. Most responses come from top-level executives, including owner-managers, CEOs, and CFOs (Bischof et al., 2024).

For the purposes of our experiment, which ran from August 16, 2022, to November 25, 2022, we utilized a 15-week window within one of the GBP's six-month survey intervals. The structural features of the GBP make it an ideal environment for implementing an MAB approach. Distributing survey invitations in daily batches, rather than all at once, enables the use of adaptive randomization. Moreover, the GBP's infrastructure provides prompt feedback on successes (survey starts) and failures (non-participation), enabling us to track the performance of each invitation message. For every firm contacted, the system logs timestamps that record when the email was sent, opened, and whether the survey was started and completed, allowing for timely adjustments of the randomization scheme during the experiment. During the 15 weeks of the experiment, a total of 176,000 contacts opened their invitation message within one week after receiving the email, 7,833 started and 3,733 completed the survey.

Message Design Alternatives. We vary five invitation message characteristics that are potentially relevant for influencing survey starting rates in voluntary business surveys. As outlined in Table 1, each of these five treatment elements is implemented in two forms, which are coded as 0 or 1 for tracking purposes. By compiling all combinations of these characteristics, we generate 32 ($= 2^5$) distinct invitation message alternatives. Thus, our experiment has a full-factorial design.

A key distinction of this experiment, compared to prior studies, is its focus on a business survey context, which necessitates tailoring the personalization treatment for firm decision-makers. We consider an invitation to be personalized if it includes the name of the legal entity the respondent is reporting on (e.g., "[...] we would like to cordially invite you—on behalf of your business [*business*

---

weights in our adaptive experimental design, which determine the probability with which each invitation message is sent out in a given week.

[12]The GBP survey is implemented using the commercial software Qualtrics.

Table 1: Overview of Treatments

(1) Characteristic	(2) Acronym	(3) Coding: **0**	(4) Coding: **1**
Personalization	**P**	No mentioning of business name	Mentioning of business name right after salutation
Authority	**A**	Sender: Names (without titles) and GBP	Sender: Names with titles, universities and GBP
URL	**U**	Bottom: Before signature	Top: After brief introductory paragraph
Data Protection	**D**	No emphasis: Single sentence within a textblock	Emphasis: Separate textblock with two strongly phrased sentences
Message Frame	**M**	Offer	Plea

Notes: This table summarizes the five distinct message treatments that generate the experimental variation in our study. Column (1) lists the characteristics that are modified to create the set of invitation messages, with abbreviations for each characteristic shown in column (2). Columns (3) and (4) describe the two specifications for each characteristic, coded as 0 or 1. A combination of these acronyms and codes is used to reference a particular invitation message alternative. For example, the invitation message P0A1U1D0M0 excludes the recipient's business name, includes the senders' academic titles and university affiliations, places the participation link near the top of the message, does not emphasize data protection, and invites participants to share their opinions. For the exact wording of each treatment in its two versions, see Appendices B.1 and B.2 as well as B.3 and B.4.

name]—to participate [...]"). While surveys of individuals typically personalize invitations using the recipient's name, this method is not appropriate for business surveys, where multiple decision-makers, such as CEOs or CFOs, may be equally eligible to respond. As a result, targeting a specific individual risks addressing someone who may not be the most relevant respondent. Moreover, personal names may not be included in the sampling frame, or the available contact details could be outdated, further complicating the use of personal salutations.

In addition to personalizing the invitation message, we generate experimental variation by manipulating the authority of the invitation's sender. To signal higher authority, the email signature explicitly lists the full academic and honorary titles of the GBP's principal investigators, along with their university affiliations. By contrast, a lower-authority version presents the senders' plain names, omitting titles and affiliations. We also vary the placement of the survey URL link, displaying it either in the upper part of the message, after a brief introductory passage, or at the bottom, before the email signature. Furthermore, the invitation messages differ in their emphasis placed on data protection. One version briefly mentions the GBP's strict data protection policies within a larger text block, while another version features a separate text block with a bold headline, specifically highlighting the GBP's commitment to privacy protection. Lastly, we modify the phrasing of the participation request by altering the message's framing. In the *offer* condition, recipients are

invited to take part in the survey and share their insights on the perceived effects of tax and accounting-related regulation ("[...] we would like to cordially invite you [...]"; "The information you provide helps us to understand the effects [...]"; "Thank you for your participation!"). In comparison, the *plea* condition frames the message as a call for advice and help ("[...] we ask you for participation [...]"; "Only with your information can we help to understand the effects [...]"; "Please help by participating!").

As indicated in Table 1, we use a combination of acronyms and codes to reference the 32 distinct message alternatives in later analyses. For instance, the invitation message labeled as P0A1U1D0M0 does not include the recipient's business name, but it does mention the academic titles and university details of the sender. In this version, the participation link is positioned in the upper part of the message, and no emphasis is placed on data protection. Finally, this message offers recipients an opportunity to share their opinions rather than issuing a plea for help. The exact wording of each treatment variation is detailed in Appendices B.1 and B.2, which contain the original German versions of the invitation messages. Their English translation is provided in Appendices B.3 and B.4. An important caveat from prior literature is the observation that participation rates may be influenced not only by the content of an invitation message, but also by its length (Kaplowitz et al., 2012). To avoid such confounding effects, we refrain from modifications that materially alter the length of the invitations.

Timeline. When implementing a Bayesian decision rule for MAB optimization, it is generally advisable to incorporate a burn-in phase to lower the risk of getting locked into sub-optimal randomization decisions in the beginning of the experiment (Du et al., 2018). During this phase, all invitation message alternatives are distributed at equal rates. The optimal length of the burn-in phase involves a trade-off: While a longer phase reduces the likelihood of premature, inefficient exploitation, it may also result in avoidable foregone survey starts by not prioritizing better-performing messages sooner. Recommendations on the optimal length of the burn-in phase vary drastically depending on the research objective and the conjectured differences in effect sizes between the experimental groups. For instance, Kaibel and Biemann (2021) suggest that adequate burn-in phases range from only ten or 20 subjects per treatment condition, if an experimenter is primarily interested in identifying the most effective treatment condition, to half of the total sample size, if the experimenter aims at determining significant treatment effects for all experimental treatment conditions. Extended burn-in phases are particularly advisable when effect size differences are expected to be small and a substantial number of experimental groups are being compared. Based

on prior survey waves of the GBP, we know that starting rates are typically below 10%. While we expect the phrasing of invitation messages to influence firms' likelihood of participation, we anticipate that these effects will be moderate. To avoid any negative impact of the experiment on the number of survey starts stemming from inefficient, early MAB choices (that would be the result of unfortunate coincidence), we adopt a conservative strategy, extending the burn-in phase to four out of the overall 15 experimental weeks.

Figure 1 schematically illustrates the experimental procedure. During the initial four experimental weeks, all invitation messages are sent out at equal rates, i.e., the email distribution weights are constant and amount to $\frac{1}{32}$. During these four weeks, the procedure mirrors an experiment with fixed and balanced randomization. From week five onwards, we implement the MAB algorithm as outlined in Section 3. Each week, the algorithm evaluates the participation data for all invitation messages likelihood of maximizing survey starting rates. As a result, better-performing invitations are sent out more often as the experiment progresses. The MAB algorithm continuously incorporates additional data to refine the distribution weights. For instance, in determining the distribution weights for week five, the algorithm factors in the outcomes from the initial four weeks. For week six, it relies on data from the previous five weeks, and this pattern continues on a weekly basis throughout the experiment.

Data. The application of the MAB procedure requires an assessment of the participation statistics per invitation message alternative on a recurring basis. To determine the weekly distribution weights in our experiment, the MAB algorithm relies on two inputs: a vector that represents the number of firms who received and subsequently opened the survey invitation message, and another vector that indicates how many of these firms began answering the questionnaire after opening the respective invitation message.[13] The success measure for each invitation message alternative is defined by its starting rate, calculated as the proportion of firms that start the survey out of those that previously opened the respective message version. We use opened messages as the baseline for this calculation, as the invitation message can influence survey participation decisions only if it has been opened.

Operationally, this data is accessible via the GBP's survey infrastructure. At the end of each experimental week, we retrieve the distribution history of prior emails, which specifies whether firms received and opened the survey invitation, as well as whether they subsequently started answering the questionnaire. This data is then matched with stored information on which invitation message

---

[13]These inputs are equivalent to the ones presented in Section 3. The number of failures simply corresponds to the difference between the amount of trials and successes.

Figure 1: Stylized Experimental Procedure with Adaptive Randomization
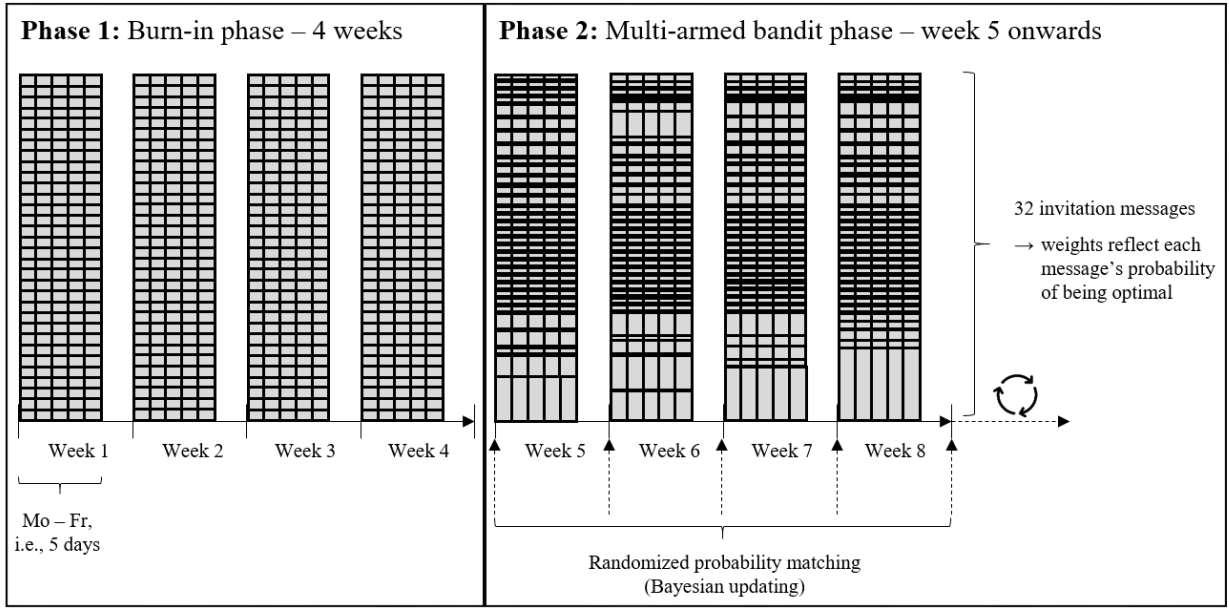
Notes: Each column in this stylized figure represents the distribution of the 32 invitation message alternatives on a single workday. The size of each rectangle illustrates the distribution weight of an individual message alternative. During the first four weeks—the burn-in phase—all messages are distributed equally. From week 5 onward, the MAB algorithm adjusts the weighting of each message version based on its posterior probability of being optimal. Within a given week, these weights remain constant. A clipping constraint ensures that each message alternative receives at least four allocations per day, with the remaining daily contacts distributed randomly among the other message versions according to their weights.

alternative was randomly assigned (according to the calculated distribution weights) to each business contact. By combining these data points, the MAB algorithm generates updated weights, which are applied to the distribution of invitation messages for the following week.

Realized Distribution Weights. Figure 1 illustrates our experimental procedures schematically. By contrast, Figure 2 displays the *realized* cumulative distribution shares of each invitation message alternative throughout the course of the experiment. During the initial four weeks—the burn-in phase—the distribution weights remained constant by design. In this 4-week period, approximately 1,500 firms opened each invitation message, resulting in a total of roughly 48,000 observations.

Beginning in week 5, the MAB algorithm assessed the success rate of each message version, defined as the proportion of opened emails that led to a survey start, and gradually allocated more weight to the more successful invitation alternatives. Over time, the message version P1A1U1D0M1 was particularly favored by the algorithm. Another example of a message that received a disproportionately high number of observations is P1A0U1D0M1. These shifting distribution weights reflect the observed differences in starting rates across the invitation message alternatives, as outlined in Section 5.1. Importantly, we ensured that at least four invitations of each message alternative were

Figure 2: Realized Cumulative Distribution Weights of Message Alternatives
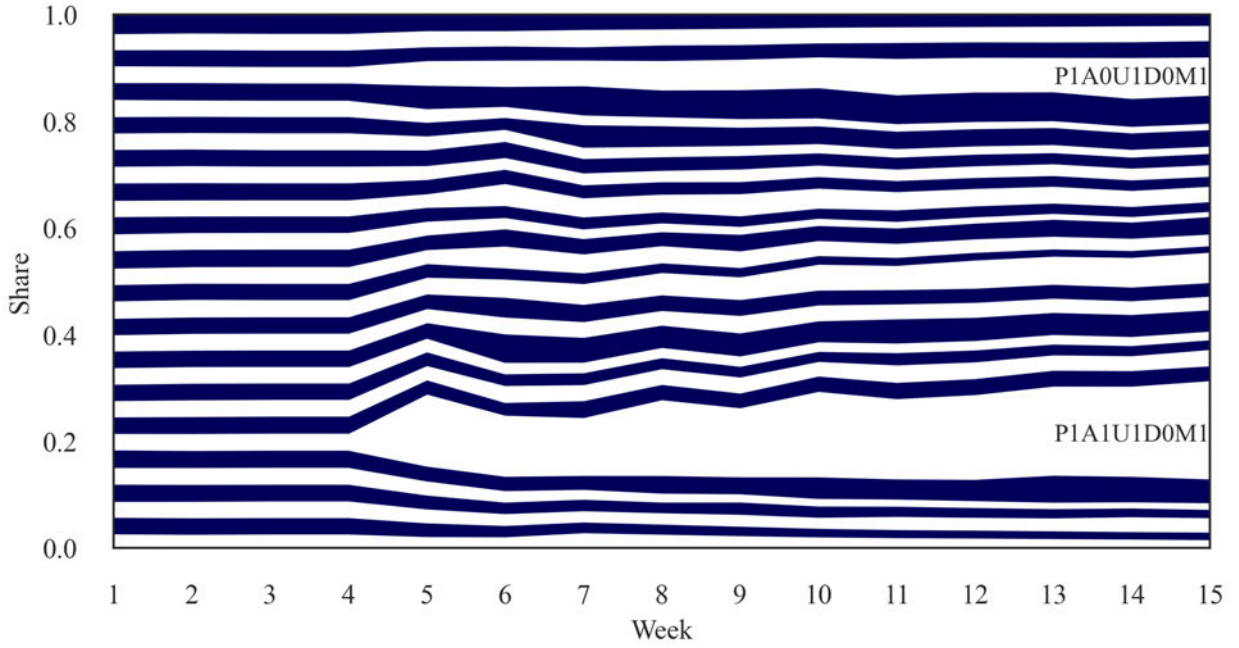
Notes: This figure shows the cumulative distribution share of the 32 invitation message alternatives over the course of the experiment. Each blue and white segment represents the cumulative distribution share of a specific message version. As per the experimental design, distribution weights remained constant at $\frac{1}{32}$ during the first four weeks. Afterward, the MAB algorithm adjusted the distribution, increasing the weight for better-performing alternatives and decreasing it for message versions with weaker performance.

sent per workday, placing a clipping constraint on the distribution weights to guarantee that each option was tested in every batch.

Randomization Checks. From the GBP's overall half-year contact pool, firms are randomly assigned to specific workdays throughout the survey wave. On each workday, these contacts are then randomly allocated to invitation message variants according to their weekly updated distribution weights. This procedure is intended to balance firm characteristics across the experimental groups. The effectiveness of our randomization process is validated in Figure A.1, which shows that the average number of employees per firm, logged due to the high skewness in firm size, and the geographic distribution of recipient firms, measured by the share located in the former East Germany, are comparable across all message variants sent during the optimization phase of the experiment. Figure A.2 provides a more granular view of the spatial distribution of sent invitations, ranked by each message version's cumulative distribution share. Naturally, as the share of invitations increases, more firms are reached, particularly in densely populated areas. However, the spatial pattern remains consistent, even as distribution weights expand, indicating no specific geographic clustering for any message variant.

When evaluating the performance of each invitation message version and updating the distribution weights, we only consider a firm contact if the respective message was opened. Two concerns related to this procedure may arise. First, there could be inaccuracies in the survey software's ability to detect whether a message was opened, but this would only be problematic if the detection errors were message-specific. Second, there may be differences in the likelihood of individual messages being opened, for example, if spam filters respond differently to the placement of the URL link. If spam filtering or other message characteristics were affecting opening rates, we would expect systematic differences across message alternatives. Figure A.3 addresses these concerns and demonstrates that the opening rates are very similar for each message version. Both a Wald test and pairwise t-tests on the equality of observations across message versions show no significant differences (smallest p-value = 0.456), alleviating concerns about biases in opening rates.

5 Results

5.1 Which Invitation Messages Yield More Survey Starts?

To structure the analysis of the experiment's results, we first evaluate the performance of the individual invitation message alternatives. Figure 3 displays the linear prediction of each message's starting rate, conditional on participants opening the survey invitation email. The graph also highlights two mean values: the average realized starting rate (in red), which incorporates the varying distribution weights assigned through MAB optimization, and the equally-weighted mean starting propensity (in black), which assumes uniform distribution across all message alternatives, i.e., $\frac{1}{32}$. The latter mean value serves as a natural benchmark for assessing the performance of MAB optimization, which is discussed in greater detail in Section 5.3.

After 15 experimental weeks, we observe considerable differences in starting rates between the individual message alternatives. While the equally-weighted average starting propensity (in black) amounts to 4.2%, individual starting rates range from 3.4% to 4.9%. Thus, sending the best-performing invitation, P1A1U1D0M1, instead of the message alternative with the lowest starting rate, P0A0U0D1M0, increases the likelihood of a manager starting the survey upon having opened the invitation message by 43.9%.

An examination of Figure 3 and the confidence intervals for each invitation's predicted starting rate reveals a noteworthy trend: Messages with higher starting rates tend to have narrower confidence intervals, while weaker-performing messages show wider confidence bands. This pattern reflects the MAB optimization's weighting procedure, which allocates greater weight to messages associated with

Figure 3: Prediction of Starting Rates
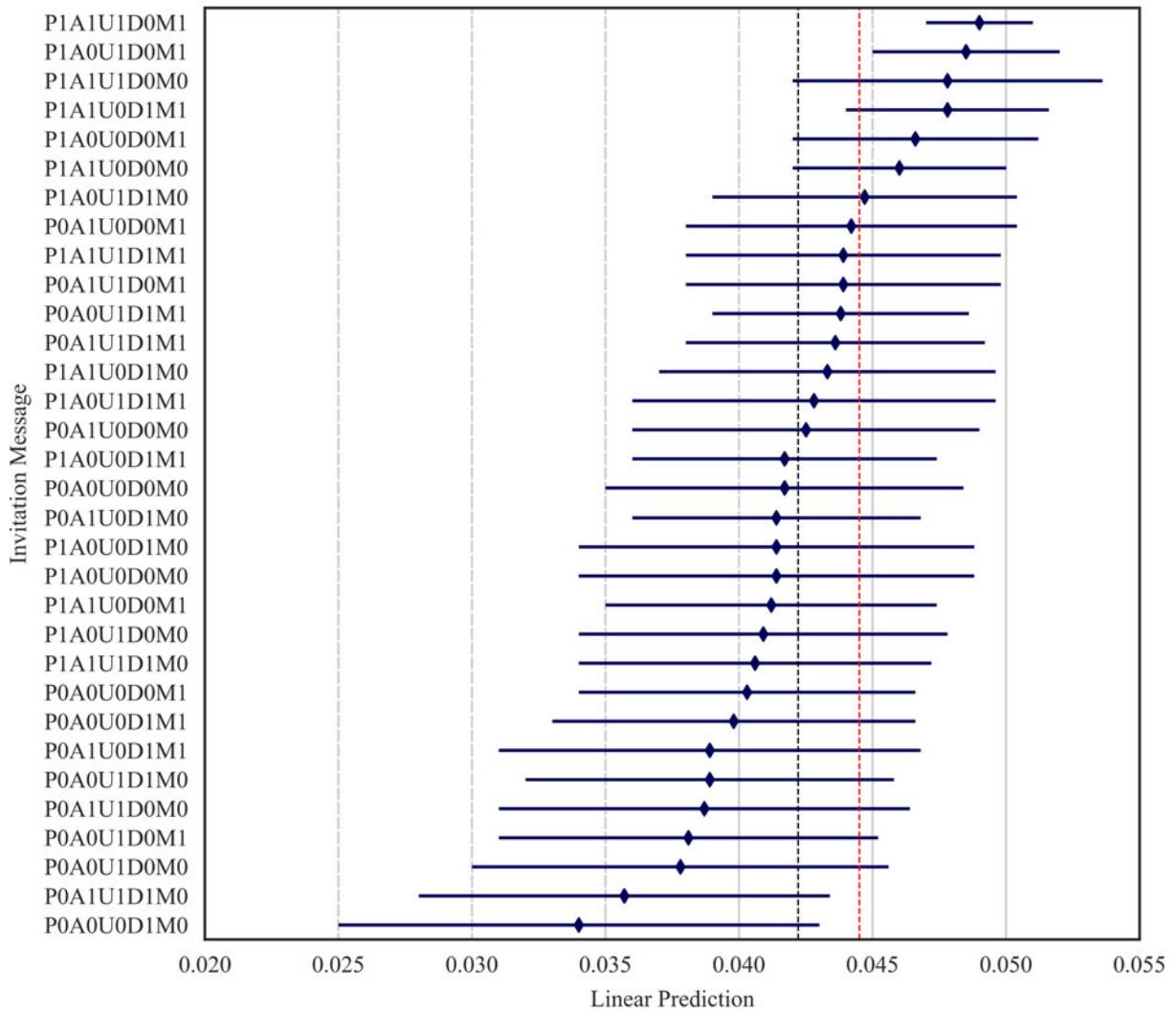
Notes: This figure presents OLS margins for the starting rates of each message alternative as well as their 95% confidence intervals with robust standard errors. The starting rate is defined as the share of firms that commence the survey after opening the invitation email. The two dashed vertical lines represent different mean values: the red line indicates the average starting rate based on the realized distribution during the experiment (with MAB optimization), while the black line shows the equally weighted average starting rate across all message versions. Thus, the black line approximates the starting rate of an experiment with fixed and balanced randomization.

higher survey starts.[14] For instance, the invitation message alternatives with the highest predicted starting rates (P1A1U1D0M1 and P1A0U1D0M1) received disproportionately high distribution weights through adaptive randomization, as shown in Figure 2.

---

[14]Under OLS, if the margin (the difference between the starting rates of two arms) is close to zero, standard inference is not valid, because the OLS estimates are asymptotically not normal. As proposed by Zhang et al. (2020), we therefore complement our analysis with batched OLS using the Stata command bbandits (Kemper and Rostam-Afschar, 2024). As illustrated in Figure A.4, our results are generally robust to using this alternative specification, which also mitigates concerns about unreliable confidence bands in our later conjoint analyses.

5.2 Which Treatments Increase Starting Rates?

Upon closer examination of Figure 3, it appears that the observed ranking of the invitation message alternatives is driven by differences in performance of the treatment elements, rather than occurring by chance. Notably, the message version with the highest average starting rate, P1A1U1D0M1, is the exact inverse of the worst-performing invitation, P0A0U0D1M0, with each treatment having the opposite specification. Moving beyond the comparison of individual messages, we next examine the specific treatment attributes underlying the message designs. Specifically, we conduct a conjoint analysis, regressing a binary variable indicating whether a firm started the survey after opening the invitation (1 if started, 0 otherwise) on the individual characteristics of the received message (P, A, U, D, M). Figure 4 presents the corresponding Average Marginal Component Effects (AMCEs), which indicate the causal impact of each treatment characteristic on the starting rate.[15]

The results of the conjoint analysis suggest that personalization, high authority, and pleading for help are associated with higher survey starting rates. Figure 4 shows that personalization significantly increases the starting propensity by 0.41 percentage points (p-value: <0.0001). High authority and pleading for help also increase starting rates by 0.16 (p-value = 0.0394) and 0.25 percentage points (p-value = 0.0036), respectively. In relative terms, these effects represent modest increases given the overall low baseline starting rate. For instance, compared to the marginal mean of 4.12% for unpersonalized invitations, personalizing the message increases the starting rate by approximately 9.95%.

The placement of the URL link, whether near the top or at the bottom of the invitation, does not significantly impact the starting rate.[16] By contrast, emphasizing data protection in the invitation is negatively related to the starting rate. Highlighting strict data protection policies reduces the starting propensity by 1.8 percentage points (p-value = 0.0336), which contradicts the expectation that such emphasis would alleviate privacy concerns and encourage participation. Instead, it seems that stressing compliance with data protection makes data security issues more salient—particularly in a business context—causing managers to be more cautious about sharing proprietary information. This heightened awareness might reduce their willingness to respond to the survey.

---

[15] The numerical estimates of this conjoint analysis are reported in Table A.1.

[16] These findings contrast with Kaplowitz et al. (2012), who report that placing the URL at the bottom of the invitation improves response rates. However, this discrepancy may stem from differences in operationalization. In the study by Kaplowitz et al. (2012), the URL in the 'top' condition appears immediately after the salutation, whereas in our 'top' condition, the URL follows an introductory paragraph. This less immediate placement may have diminished the distinction between the two placements, as the URL is still sufficiently embedded within the content to build trust and prompt engagement with the message.

Figure 4: AMCEs on the Starting Rate
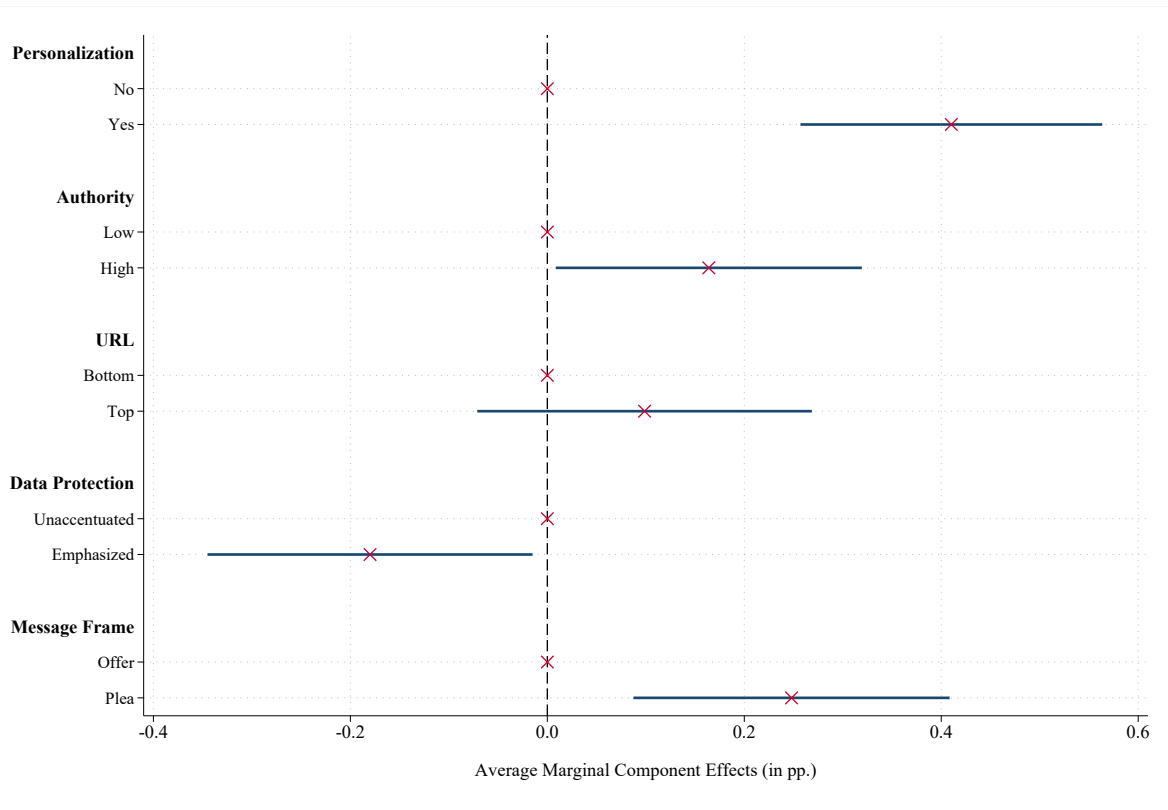
Notes: This figure presents Average Marginal Component Effects (AMCEs) from a conjoint analysis. The dependent variable is a dummy coded as one if a message recipient, who opened the invite, subsequently started answering the questionnaire (0 otherwise). The AMCE denotes the marginal effect of changing an invitation message's attribute averaging over the joint distribution of the remaining attributes. Whiskers indicate 95% confidence intervals with standard errors that are clustered at the message level. The number of observations (firms that opened the survey invitation message) amounts to 176,000.

5.3 MAB Optimization vs. Static Experiment

MAB optimization balances exploration and exploitation in sequential experiments and, thus, should improve outcomes as compared to experiments with static group composition. In this section, we quantify the gain in additional survey starts that we achieved through reinforcement learning using randomized probability matching. For that purpose we construct a counterfactual group that mimics the properties of a fixed and balanced randomization scheme. By design, in a static experiment the distribution weights per invitation message would have remained constant $(\frac{1}{32})$ throughout the entire experimental phase.

Because the burn-in phase uses fixed and balanced randomization, optimization—and thus any performance gains—begins only after this period, when the algorithm adjusts distribution weights. Using each message's mean starting rate, we predict the expected number of survey starts under a purely fixed and balanced randomization scheme and compare this to the realized survey starts under the MAB procedure. Table A.2 contains the detailed results. Relative to the 5,735 survey

21

starts that were recorded after the burn-in phase of the experiment, a static design would have been expected to generate 5,377 starts over the same period. As intended, MAB optimization reduced the number of foregone responses by prioritizing better performing invitations. Specifically, MAB optimization resulted in 358 additional survey starts, representing a 6.66% increase compared to a traditional static randomization setup. These gains have to be viewed in light of the heavily exploration-focused design of the experiment. A more exploitation-focused design with a shorter burn-in phase, lower batch size, and fewer arms could have increased starts further.

6 Supplementary Analyses

6.1 Treatment Effects on the Completion Rate

Our main analysis has shown that specific characteristics of a survey invitation message targeted at businesses influence their propensity to start answering the questionnaire. In this section, we examine whether these effects persist beyond the decision to start the survey, focusing on the

Figure 5: AMCEs on the Completion Rate
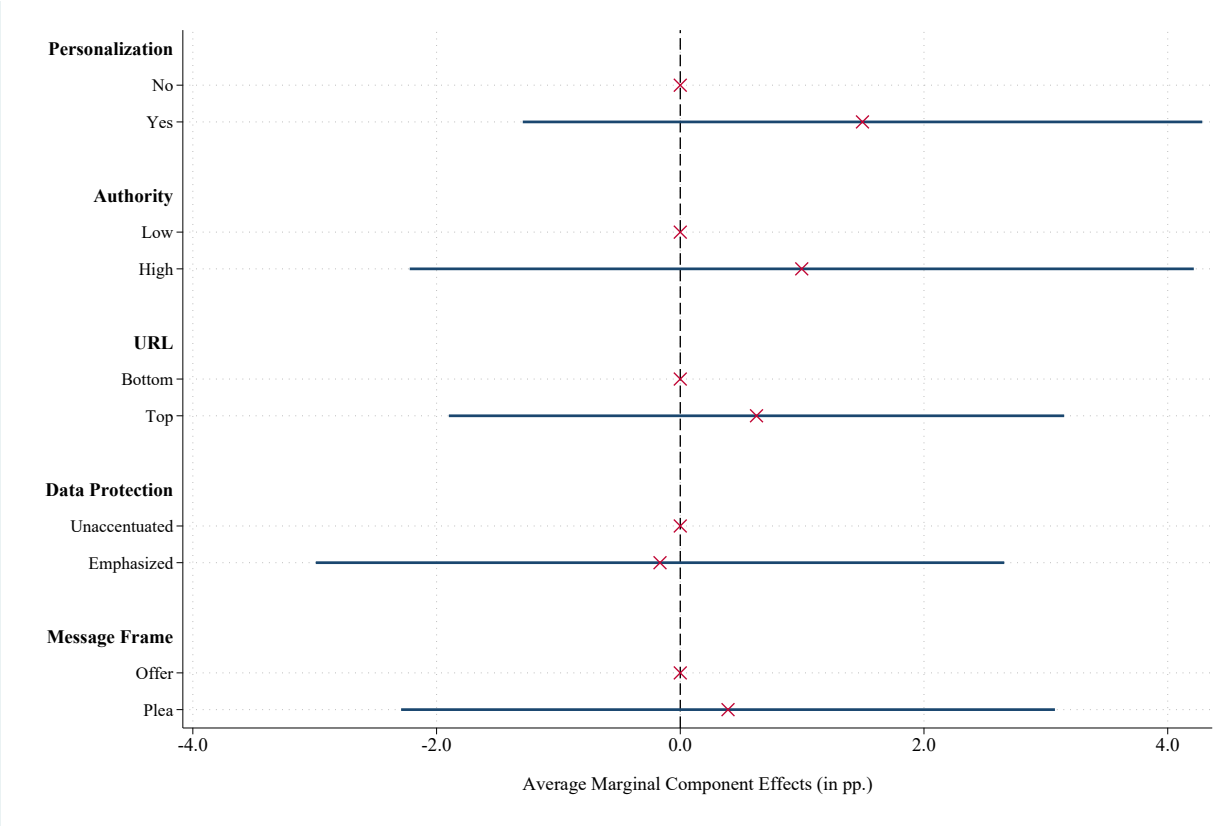
Notes: This figure presents AMCEs from a conjoint analysis, where the dependent variable is a binary indicator set to one if a survey respondent completed the questionnaire after starting the survey (0 otherwise). Whiskers indicate 95% confidence intervals, with standard errors clustered at the message level.

likelihood of completing it. To do so, we condition our analysis on businesses that began the survey and replace the dependent variable with a binary indicator set to one if a firm decision-maker completed the questionnaire. Figure 5 presents the corresponding AMCEs from a conjoint analysis on the completion rate.

The results indicate that, once a manager has started the survey, the specific characteristics of the invitation message no longer significantly affect the probability of completing the questionnaire. None of the AMCEs are statistically significant at conventional levels, suggesting that, after the decision to start, message attributes do not influence the likelihood of survey completion. Thus, considering both the impact on survey starts and completions, improvements in message design ultimately increase the total number of completed responses by motivating more managers to begin the survey, without negatively affecting the likelihood of finishing it once started.
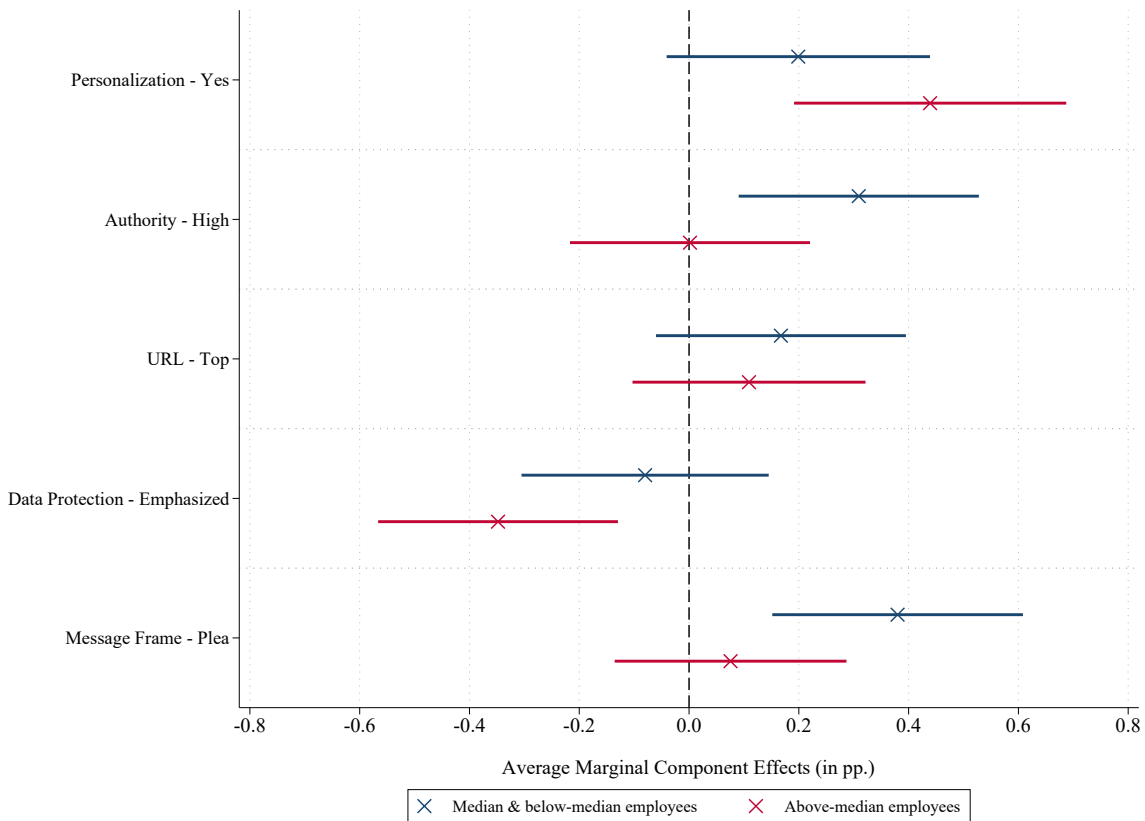
6.2 Heterogeneous Treatment Effects

This section explores heterogeneous treatment effects, focusing on whether firms of different sizes respond differently to specific invitation message characteristics. To investigate this, we split the sample according to the number of employees, using staffing data available from Bureau van Dijk's Orbis database for 138,380 out of the 176,000 firms that opened their invitation message. The median number of employees within this sample is four. Figure 6 presents the AMCEs of different invitation message characteristics on the starting rate for firms with employee counts at or below the median as well as for firms with above-median employees.

The results indicate that smaller firms appear to be particularly receptive to messages issued by a source with high authority or framed as a plea for help. By contrast, larger firms are more likely to engage with personalized invitations, while emphasizing data protection appears to reduce their willingness to participate. One possible explanation for this finding is that larger firms face higher risks when sharing sensitive information. Additionally, highlighting data protection may create the impression that employees may not be authorized to respond, introducing perceived procedural hurdles that can discourage survey participation.

In Appendix A, we analyze whether these heterogeneous treatment effects result in any nonresponse bias. The statistics in Table A.3 show that the staffing numbers between firms that started the questionnaire and non-starting firms do not significantly differ across the treatment characteristics.[17] In aggregate, these effects did not introduce nonresponse bias, likely due to the high number of observations in each experimental group and the moderate size of the treatment effects. Using

---

[17]Hack and Rostam-Afschar (2024) provide evidence that the composition of respondents is stable even across days.

Figure 6: AMCEs Depending on a Firm's Size

Average Marginal Component Effects (in pp.)

× Median & below-median employees × Above-median employees

Notes: This figure presents the Average Marginal Component Effects (AMCEs) from two separate conjoint analyses, where the outcome variable is an indicator equal to one if a respondent initiated the survey and zero otherwise. The sample is stratified by firm size, based on the number of employees as reported by Bureau van Dijk's Orbis database for 138,380 of the 176,000 survey recipients. The median number of employees is four. Error bars represent 95% confidence intervals, with standard errors clustered at the message level. Unlike previous figures, this one omits the baseline levels, but instead highlights the marginal effect of activating each attribute for the two subgroups.

contextualized bandits, messages could be tailored to recipients based on characteristics like firm size (Hoffmann et al., 2023). However, using this extension is more likely to introduce bias.

6.3 Excluding Prior Participants

Due to the panel structure of the GBP, some businesses approached during this experiment had participated in previous survey waves. In principle, potential confounding effects from prior participation are mitigated through our randomization procedure, which ensures that prior participants and non-participants are proportionally distributed across the experimental groups. Additionally, we conduct a robustness test that excludes all 1,208 firms that had previously completed a GBP questionnaire from consideration.

The results of the corresponding conjoint analysis are shown in Figure 7. They remain largely consistent with those in the main specification. Personalization and pleading for help significantly

Figure 7: AMCEs Given No Prior Participation
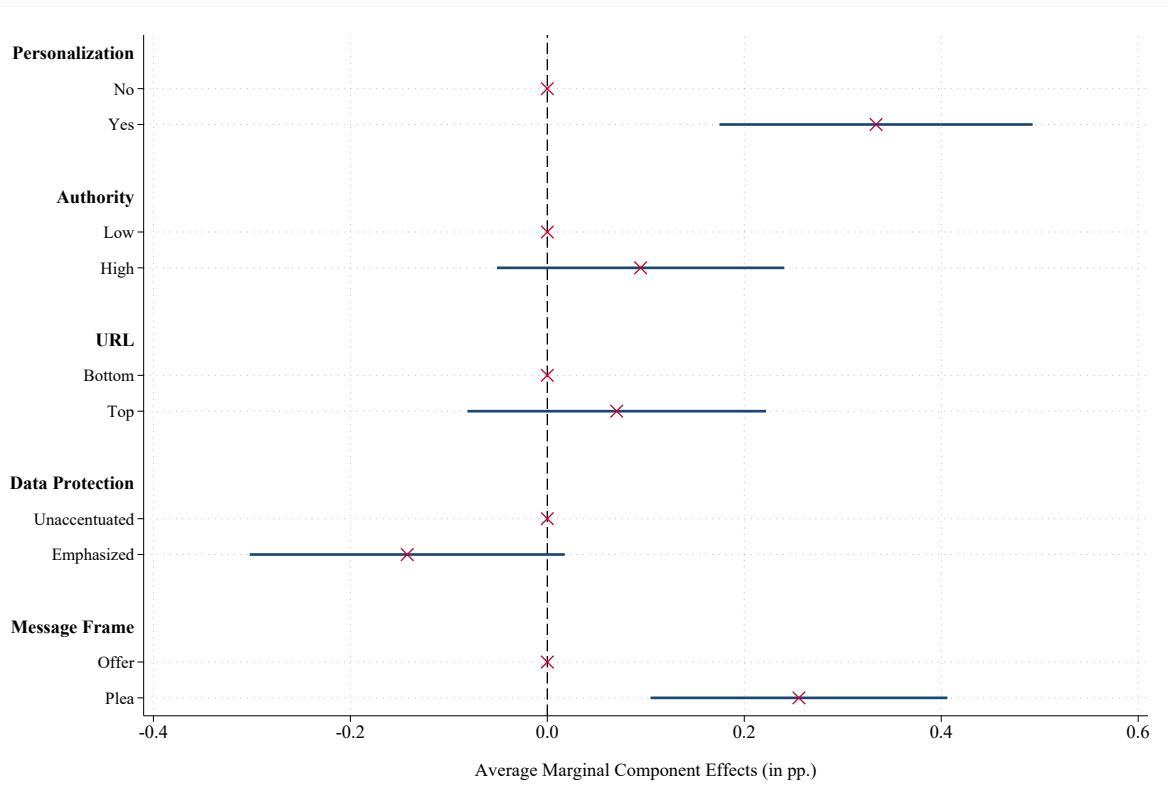
Notes: This figure presents AMCEs from a conjoint analysis. The dependent variable is a binary indicator set to one if a respondent who opened the invitation subsequently started the questionnaire (0 otherwise). The analysis is limited to businesses that had not participated in any prior GBP survey. Of the 176,000 recipients who opened the invitation, 1,208 had previously completed a GBP survey and are excluded from this analysis. Whiskers represent 95% confidence intervals, with standard errors clustered at the message level.

increase response rates compared to unpersonalized invitations and offer frames. The effects of emphasizing authority and data protection are no longer statistically significant at the 5% level. Nonetheless, the analysis still suggests that emphasizing data protection in the invitation message does not improve survey participation.

7 Conclusion

This study investigates which elements of an email invitation encourage business decision-makers to participate in a self-administered survey. Understanding this is important because findings from household surveys do not automatically transfer to the business context, where unique constraints influence the decision-making process to engage with a survey request. In our full-factorial experiment, we find that personalizing the message, highlighting the sender's authority, and framing the invitation as a plea for help enhance the likelihood that managers begin the survey. Importantly, personalization is achieved by referencing the company name rather than an

individual, underscoring the adaptability of this strategy in a corporate context. In contrast to these response-enhancing effects, changing the location of the survey URL within the email has no discernible impact on participation rates. Notably, emphasizing data protection in the invitation message appears to discourage participation, especially among larger firms. This may be due to concerns about sharing sensitive information or the perception that responding requires additional internal authorization, raising perceived procedural hurdles.

Caution is warranted when transferring these findings to other contexts, as this study examines a German corporate environment. For example, operationalizing high authority in different cultural settings may involve approaches other than referencing academic titles or university affiliations. Additionally, the economic significance of our results must be carefully considered. The effects we observe are modest. This is expected, given the generally low levels of engagement in web surveys, the subtle nature of the tested interventions, and the large number of message alternatives. In fact, a large number of groups in a factorial experiment mechanically leads to smaller effect sizes. In large-scale surveys like the GBP, even small improvements in starting rates can lead to meaningful increases in overall responses. This does not imply, however, that smaller surveys cannot benefit from these findings. In contrast, many MAB algorithms including Thompson sampling are designed to work in settings with few observations per batch. Generally, the decision to implement these strategies involves balancing the potential gains against the costs of application. For instance, leveraging a high-authority sender or adjusting the framing of the invitation requires minimal effort and textual adjustments, promising cost-effective gains. In comparison, personalization, while impactful, entails higher administrative costs, such as obtaining company-specific data, and thus requires careful weighing of the expected benefits and costs.

Our experimental design is innovative by using MAB optimization rather than relying on traditional fixed and balanced randomization. The core idea of MAB optimization is to balance learning about the performance of experimental groups (exploration) with capitalizing on the knowledge gain already during the experimental phase (exploitation). To manage this trade-off, we employ a Bayesian decision rule, randomized probability matching, that offers several advantages. This approach not only maximizes output more effectively than simpler heuristics but also maintains acceptable levels of statistical power (Scott, 2010). With a sufficiently large burn-in phase, a clipping constraint that ensures a minimum level of exploration, and a sufficiently large batch size (we recommend at least 20 observations), the algorithm minimizes the risk of prematurely locking into suboptimal choices. The input required is simply the successes and failures of each experimental

group, and because the decision rule does not require tuning parameters, it avoids assumptions about reducing exploration over time. Applying this decision rule, we were able to increase survey starts by 6.66% compared to a traditional fixed and balanced randomization scheme.

While MAB optimization offers advantages, it also involves administrative costs and requires specific prerequisites for implementation. First, MAB optimization is only feasible in sequential experiments where the randomization scheme can be updated. In settings where this is not naturally possible, partitioning the sample or identifying outcomes that can be measured early may offer a way to enable sequential updates. Second, MAB optimization relies on having a clear output to optimize, which may not always be the case in experiments lacking well-defined research questions or hypotheses. When an experiment is output-oriented, it becomes crucial to carefully select the right variable for optimization. In our experiment, focusing on the number of survey starts per opened message appeared to be a natural choice. However, if certain message characteristics had negatively affected completion rates, using the starting rate as optimization variable would have been suboptimal. Third, MAB optimization requires timely data collection and an infrastructure that allows to adjust the randomization scheme dynamically. This necessitates not only the recurring calculation of distribution weights, but also the integration of randomization scheme adjustments into the experimental process, generating implementation costs. Fourth, when margins are small— particularly with many arms—usual hypothesis tests based on OLS fail. Therefore, it is useful to report the results of the batched OLS estimator as well (Zhang et al., 2020; Kemper and Rostam-Afschar, 2024). Lastly, most algorithms assume that the reward distributions remain stable over time. If seasonal response patterns or other non-stationary factors come into play, the algorithm may over-explore and underperform (Liu et al., 2023). Addressing this would require additional modeling of time dynamics, imposing further analytical and administrative challenges.

Despite these challenges, MAB optimization is a promising alternative to fixed and balanced randomization, particularly in experiments where maximizing output is a priority. Researchers often assess the efficiency of experiments by focusing on statistical power, whereas firms and organizations are more concerned with maximizing outcomes. MAB optimization balances these two objectives and, thus, provides a way forward to conducting more experiments within business contexts and surveys. Potential fields of application lie, for instance, in the area of questionnaire design, adaptively randomized information provision, or vignette experiments. These future applications might consider extensions of our experimental design. One possible extension is the use of contextual bandits, which incorporate additional information, such as the characteristics of survey participants, into the

decision-making process. For instance, tailoring the randomization scheme based on firm size could have further enhanced the effectiveness of our experiment. However, contextualizing the bandit requires prior (data) knowledge about the firms and introduces additional complexity. Since a larger number of experimental groups reduces the differences between effect sizes, taking into account contextual variables in addition could lead to less conclusive results. Furthermore, in scenarios where identifying the best-performing experimental group quickly is the priority, pure exploration bandits might be a useful alternative, even when it implies sacrificing short-term output. Generally, MAB optimization might become a cost-effective option for conducting experiments in surveys once commercial software providers begin integrating this feature into their platforms.

References

Agrawal, S. and Goyal, N. (2012). Analysis of Thompson sampling for the multi-armed bandit problem. In Mannor, S., Srebro, N., and Williamson, R. C., editors, *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 39.1–39.26, Edinburgh, Scotland. PMLR.

Al-Natour, S., Cavusoglu, H., Benbasat, I., and Aleem, U. (2020). An empirical investigation of the antecedents and consequences of privacy uncertainty in the context of mobile apps. *Information Systems Research*, 31(4):1037–1063.

Barron, G. and Yechiam, E. (2002). Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, 18(5):507–520.

Bischof, J., Doerrenberg, P., Rostam-Afschar, D., Simons, D., and Voget, J. (2024). The German Business Panel: Firm-level data for accounting and taxation research. *European Accounting Review*, pages 1–29.

Bouneffouf, D. and Rish, I. (2019). A survey on practical applications of multi-armed and contextual bandits. *arXiv preprint arXiv:1904.10040*.

Breur, T. (2016). Statistical power analysis and the contemporary 'crisis' in social sciences. *Journal of Marketing Analytics*, 4(2):61–65.

Cialdini, R. B. (2001). The science of persuasion. *Scientific American*, 284(2):76–81.

Cook, C., Heath, F., and Thompson, R. L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and Psychological Measurement*, 60(6):821–836.

Couper, M. P. (2008). *Designing effective Web surveys.* Cambridge University Press, Cambridge.

Crawford, S. D., Couper, M. P., and Lamias, M. J. (2001). Web surveys: Perceptions of burden. *Social Science Computer Review*, 19(2):146–162.

Daikeler, J., Bošnjak, M., and Lozar Manfreda, K. (2020). Web versus other survey modes: an updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology*, 8(3):513–539.

Dillman, D. A. (2007). *Mail and Internet surveys: The tailored design method.* John Wiley & Sons, New York.

Dillman, D. A. (2021). Towards survey response rate theories that no longer pass each other like strangers in the night. In Brenner, P. S., editor, *Understanding survey methodology: Sociological theory and applications*, volume 4, pages 15–44. Springer, Cham.

Du, Y., Cook, J. D., and Lee, J. J. (2018). Comparing three regularization methods to avoid extreme allocation probability in response-adaptive randomization. *Journal of Biopharmaceutical Statistics*, 28(2):309–319.

Einarsson, H., Cernat, A., and Shlomo, N. (2021). Reducing respondent burden with efficient survey invitation design. *Survey Research Methods*, 15(3):207–233.

Evans, J. R. and Mathur, A. (2005). The value of online surveys. *Internet Research*, 15(2):195–219.

Fan, W. and Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26(2):132–139.

Felix, L. M., Burchett, H. E., and Edwards, P. J. (2011). Factorial trial found mixed evidence of effects of pre-notification and pleading on response to web-based survey. *Journal of Clinical Epidemiology*, 64(5):531–536.

Ferreira, K. J., Simchi-Levi, D., and Wang, H. (2018). Online network revenue management using thompson sampling. *Operations Research*, 66(6):1586–1602.

Guéguen, N. and Jacob, C. (2002). Solicitation by e-mail and solicitor's status: A field study of social influence on the web. *CyberPsychology & Behavior*, 5(4):377–383.

Hack, L. and Rostam-Afschar, D. (2024). Understanding firm dynamics with daily data. Technical report. CRC 224 Economic Perspectives on Societal Challenges Discussion Paper 593 and CRC 266 Accounting for Transparency Working Paper Series No. 155.

Haraldsen, G. (2018). Response processes and response quality in business surveys. In Lorenc, B., Smith, P. A., Bavdaž, M., Haraldsen, G., Nedyalkova, D., Zhang, L.-C., and Zimmermann, T., editors, *The Unit Problem and Other Current Topics in Business Survey Methodology*, pages 155–176. Cambridge Scholars Publishing, Newcastle upon Tyne.

Heberlein, T. A. and Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, 32(4):447–462.

Heerwegh, D. (2005). Effects of personal salutations in e-mail invitations to participate in a web survey. *Public Opinion Quarterly*, 69(4):588–598.

Heerwegh, D. and Loosveldt, G. (2002). An evaluation of the effect of response formats on data quality in web surveys. *Social Science Computer Review*, 20(4):471–484.

Heerwegh, D. and Loosveldt, G. (2006). An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in web surveys. *Journal of Official Statistics*, 22(2):191–210.

Heerwegh, D. and Loosveldt, G. (2007). Personalizing e-mail contacts: Its influence on web survey response rate and social desirability response bias. *International Journal of Public Opinion Research*, 19(2):258–268.

Heerwegh, D., Vanhove, T., Matthijs, K., and Loosveldt, G. (2005). The effect of personalization on response rates and data quality in web surveys. *International Journal of Social Research Methodology*, 8(2):85–99.

Hill, D. H. and Willis, R. J. (2001). Reducing panel attrition: A search for effective policy instruments. *Journal of Human Resources*, 36(3):416–438.

Hoffmann, M., Picard, B., Marie, C., and Bied, G. (2023). An adaptive experiment to boost online skill signaling and visibility. Working paper, retrieved from: bertillepicard.github.io/documents/papers/AnAdaptiveExpeToBoostOnlineSkillsSignaling.pdf (10-04-2024).

Hui, K.-L., Teo, H. H., and Lee, S.-Y. T. (2007). The value of privacy assurance: An exploratory field experiment. *MIS Quarterly*, 31(1):19–33.

Joinson, A. N. and Reips, U.-D. (2007). Personalized salutation, power of sender and response rates to web-based surveys. *Computers in Human Behavior*, 23(3):1372–1383.

Kaibel, C. and Biemann, T. (2021). Rethinking the gold standard with multi-armed bandits: machine learning allocation algorithms for experiments. *Organizational Research Methods*, 24(1):78–103.

Kandasamy, K., Krishnamurthy, A., Schneider, J., and Poczos, B. (2018). Parallelised bayesian optimisation via thompson sampling. In Storkey, A. and Perez-Cruz, F., editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 133–142. PMLR.

Kaplowitz, M. D., Lupi, F., Couper, M. P., and Thorp, L. (2012). The effect of invitation design on web survey response rates. *Social Science Computer Review*, 30(3):339–349.

Kasy, M. and Sautmann, A. (2021). Adaptive treatment assignment in experiments for policy choice. *Econometrica*, 89(1):113–132.

Kemper, J. and Rostam-Afschar, D. (2024). Earning while learning: How to run batched bandit experiments. Unpublished.

Kent, R. and Brandal, H. (2003). Improving email response in a permission marketing context. *International Journal of Market Research*, 45(4):1–13.

Keusch, F. (2015). Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection. *Management Review Quarterly*, 65(3):183–216.

König, C. and Sakshaug, J. W. (2023). Nonresponse trends in establishment panel surveys: Findings from the 2001–2017 IAB establishment panel. *Journal for Labour Market Research*, 57.

König, C., Sakshaug, J. W., Stegmaier, J., and Kohaut, S. (2021). Trends in establishment survey nonresponse rates and nonresponse bias: Evidence from the 2001-2017 IAB establishment panel. *Journal of Official Statistics*, 37(4):931–953.

Küfner, B., Sakshaug, J. W., and Zins, S. (2022). Analysing establishment survey non-response using administrative data and machine learning. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185:310–342.

Langeland, J., Ridolfo, H., McCarthy, J., Ott, K., Kilburg, D., CyBulski, K., Krakowiecki, M., Vittoriano, L., Potts, M., Küfner, B., Sakshaug, J. W., and Zins, S. (2023). Recent findings from experiments in establishment surveys. In Snijkers, G., Bavdaz, M., Bender, S., Jones, J., MacFeely, S., Sakshaug, J. W., Thompson, K. J., and van Delden, A., editors, *Advances in Business Statistics, Methods and Data Collection*, pages 437–467. John Wiley & Sons.

Liu, Y., Van Roy, B., and Xu, K. (2023). Nonstationary bandit learning via predictive sampling. In *International Conference on Artificial Intelligence and Statistics*, pages 6215–6244. PMLR.

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., and Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International journal of market research*, 50(1):79–104.

Muñoz-Leiva, F., Sánchez-Fernández, J., Montoro-Ríos, F., and Ibáñez-Zapata, J. Á. (2010). Improving the response rate and quality in web-based surveys through the personalization and frequency of reminder mailings. *Quality & Quantity*, 44:1037–1052.

Petrovčič, A., Petrič, G., and Manfreda, K. L. (2016). The effect of email invitation elements on response rate in a web survey within an online community. *Computers in Human Behavior*, 56:320–329.

Pielsticker, D. I. and Hiebl, M. R. (2020). Survey response rates in family business research. *European Management Review*, 17(1):327–346.

Porter, S. R. and Whitcomb, M. E. (2003). The impact of contact type on web survey response rates. *The Public Opinion Quarterly*, 67(4):579–588.

Ramirez, C. (1997). *Effects of precontacting on response and cost in self-administered establishment surveys*. American Statistical Association, 1000-1005.

Sánchez-Fernández, J., Muñoz-Leiva, F., and Montoro-Ríos, F. J. (2012). Improving retention rate and response quality in web-based surveys. *Computers in Human Behavior*, 28(2):507–514.

Sauermann, H. and Roach, M. (2013). Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. *Research Policy*, 42(1):273–286.

Sax, L. J., Gilmartin, S. K., and Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, 44(4):409–432.

Schwartz, E. M., Bradlow, E. T., and Fader, P. S. (2017). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4):500–522.

Scott, S. L. (2010). A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26(6):639–658.

Snijkers, G., Haraldsen, G., Jones, J., and Willimack, D. (2013). *Designing and conducting business surveys*. John Wiley & Sons.

Tanis, M. (2007). Online social support groups. In Joinson, A., McKenna, K., Postmes, T., and Reips, U., editors, *The Oxford handbook of Internet psychology*, pages 139–153. Oxford University Press, New York.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294.

Trespalacios, J. H. and Perkins, R. A. (2016). Effects of personalization and invitation email length on web-based survey response rates. *TechTrends*, 60:330–335.

Trouteaud, A. R. (2004). How you ask counts: A test of Internet-related components of response rates to a web-based survey. *Social Science Computer Review*, 22(3):385–392.

Wiley, J. B., Han, V., Albaum, G., and Thirkell, P. (2009). Selecting techniques for use in an Internet survey. *Asia Pacific Journal of Marketing and Logistics*, 21(4):455–474.

Yammarino, F. J., Skinner, S. J., and Childers, T. L. (1991). Understanding mail survey response behavior a meta-analysis. *Public Opinion Quarterly*, 55(4):613–639.

Zhang, K., Janson, L., and Murphy, S. (2020). Inference for batched bandits. *Advances in Neural Information Processing Systems*, 33:9818–9829.

Appendix

A Extensions and Robustness

Table A.1: Conjoint Analysis

Dep. var.: dummy variable set to one if survey was started after invitation has been opened.

	AMCE				MM	
	Est.	SE (P-Value)	SE (P-Value)	SE (P-Value)	Est.	SE
Personalization						
No	−	−	−	−	0.0412	0.0007
Yes	0.0041	0.0011 (0.0001)	0.0010 (0.0001)	0.0008 (0.0000)	0.0461	0.0010
Authority						
Low	−	−	−	−	0.0432	0.0012
High	0.0016	0.0010 (0.1046)	0.0010 (0.1029)	0.0008 (0.0394)	0.0454	0.0012
URL						
Bottom	−	−	−	−	0.0434	0.0009
Top	0.0010	0.0010 (0.3317)	0.0010 (0.3278)	0.0008 (0.2446)	0.0453	0.0014
Data Protection						
Unaccentuated	−	−	−	−	0.0457	0.0012
Emphasized	-0.0018	0.0010 (0.0800)	0.0010 (0.0773)	0.0008 (0.0336)	0.0427	0.0009
Message Frame						
Offer	−	−	−	−	0.0423	0.0009
Plea	0.0025	0.0011 (0.0197)	0.0010 (0.0181)	0.0008 (0.0036)	0.0456	0.0011
Clustered SE	−	No	Respondent	Message	−	Message
N	176,000	176,000	176,000	176,000	176,000	176,000

Notes: This table presents Average Marginal Component Effects (AMCEs) and Marginal Means (MMs) from a conjoint anaysis. The dependent variable is a dummy coded as one if a respondent, who has opened the invitation message, subsequently started answering the questionnaire (0 otherwise). For the AMCEs, we show non-clustered standard errors as well as standard errors that are clustered on the level of the respondents and the message alternatives.

Table A.2: Static Experiment vs. MAB Optimization

|	(II) Mean Start. Rate (>Week4)	Static		MAB	
		(III) Distribution Weight	(IV) Predicted Starts	(V) Distribution Weight	(VI) Realized Starts
P0A0U0D0M0	0.0426	0.0313	170.85	0.0134	73
P0A0U0D0M1	0.0373	0.0313	149.45	0.0184	88
P0A0U0D1M0	0.0377	0.0313	151.37	0.0017	8
P0A0U0D1M1	0.0380	0.0313	152.55	0.0111	54
P0A0U1D0M0	0.0415	0.0313	166.61	0.0049	26
P0A0U1D0M1	0.0353	0.0313	141.77	0.0097	44
P0A0U1D1M0	0.0372	0.0313	149.16	0.0136	65
P0A0U1D1M1	0.0422	0.0313	169.45	0.0433	235
P0A1U0D0M0	0.0384	0.0313	154.15	0.0174	86
P0A1U0D0M1	0.0445	0.0313	178.55	0.0263	150
P0A1U0D1M0	0.0397	0.0313	159.13	0.0283	144
P0A1U0D1M1	0.0353	0.0313	141.53	0.0071	32
P0A1U1D0M0	0.0453	0.0313	181.54	0.0053	31
P0A1U1D0M1	0.0441	0.0313	176.83	0.0315	178
P0A1U1D1M0	0.0298	0.0313	119.39	0.0052	20
P0A1U1D1M1	0.0425	0.0313	170.28	0.0273	149
P1A0U0D0M0	0.0441	0.0313	176.97	0.0079	45
P1A0U0D0M1	0.0468	0.0313	187.73	0.0489	294
P1A0U0D1M0	0.0436	0.0313	174.79	0.0107	60
P1A0U0D1M1	0.0418	0.0313	167.69	0.0306	164
P1A0U1D0M0	0.0413	0.0313	165.82	0.0138	73
P1A0U1D0M1	0.0478	0.0313	191.62	0.0866	531
P1A0U1D1M0	0.0431	0.0313	172.98	0.0332	184
P1A0U1D1M1	0.0459	0.0313	184.20	0.0183	108
P1A1U0D0M0	0.0443	0.0313	177.70	0.0603	343
P1A1U0D0M1	0.0371	0.0313	148.67	0.0235	112
P1A1U0D1M0	0.0429	0.0313	172.20	0.0234	129
P1A1U0D1M1	0.0473	0.0313	189.83	0.0663	403
P1A1U1D0M0	0.0509	0.0313	204.05	0.0256	167
P1A1U1D0M1	0.0481	0.0313	193.08	0.2405	1,486
P1A1U1D1M0	0.0389	0.0313	155.88	0.0150	75
P1A1U1D1M1	0.0451	0.0313	180.81	0.0308	178
		1.0000	~5,377	1.0000	5,735

Notes: This table provides an approximation of the survey starts that have been gathered through MAB optimization relative to a static experiment with fixed and balanced randomization. Column (II) presents the mean starting rates per message alternative during the optimization phase of the experiment (week five to 15). Column (III) contains the distribution weights in a static experiment, which are constant per assumption ($\frac{1}{32}$). By contrast, column (V) shows the cumulative distribution weights that were realized through reinforcement learning following week 4. The predicted starts in column (IV) are calculated as the product of the mean starting rate (II), the distribution weight (III), and the number of opened invitation messages after the burn-in phase of the experiment (128,364).

Table A.3: Number of Employees by Response Status

	Starters			Non-Starters			Mean Diff.	
	N	Mean	SE	N	Mean	SE	T-Stat. Abs. Empl.	T-Stat. ln(Empl.)
P=1: Personalized	4,071	40.62	24.68	89,106	23.29	3.77	-0.6938	-1.5494
A=1: High Authority	3,573	45.32	28.13	79,036	25.43	4.57	-0.6980	-0.0027
U=1: Top	3,460	15.99	1.08	76,341	19.54	0.99	2.4109**	-0.8538
D=1: Emphasized	2,256	18.40	2.30	53,440	22.84	2.94	1.1889	0.3929
M=1: Plea	3,984	43.98	25.28	87,684	20.84	1.78	-0.9134	-1.0895

Notes: This table presents statistics from two-sample t-tests with unequal variances depending on whether a firm has started the survey or not. Employee data is available for 138,380 among the overall 176,000 message receivers. In terms of absolute staffing numbers, starters and non-starters only differ when receiving a URL that is displayed at the top of the invitation message (95% significance level). Since staffing numbers are highly skewed, we also log-transform the variable. When considering these natural logarithms, there are no significant differences between starters and non-starters. This analysis provides support that our experiment did not lead to a specific selection of survey respondents.

Figure A.1: Balancing Checks

(a) Employee Numbers
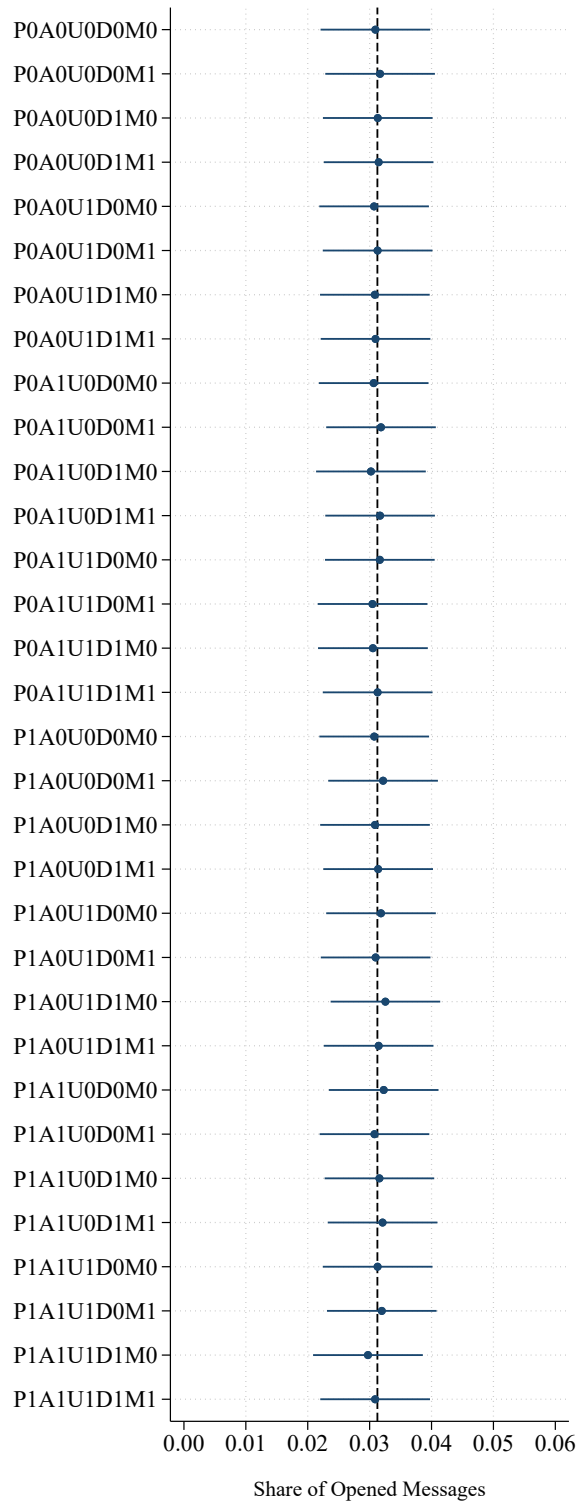
(b) Geographic Distribution
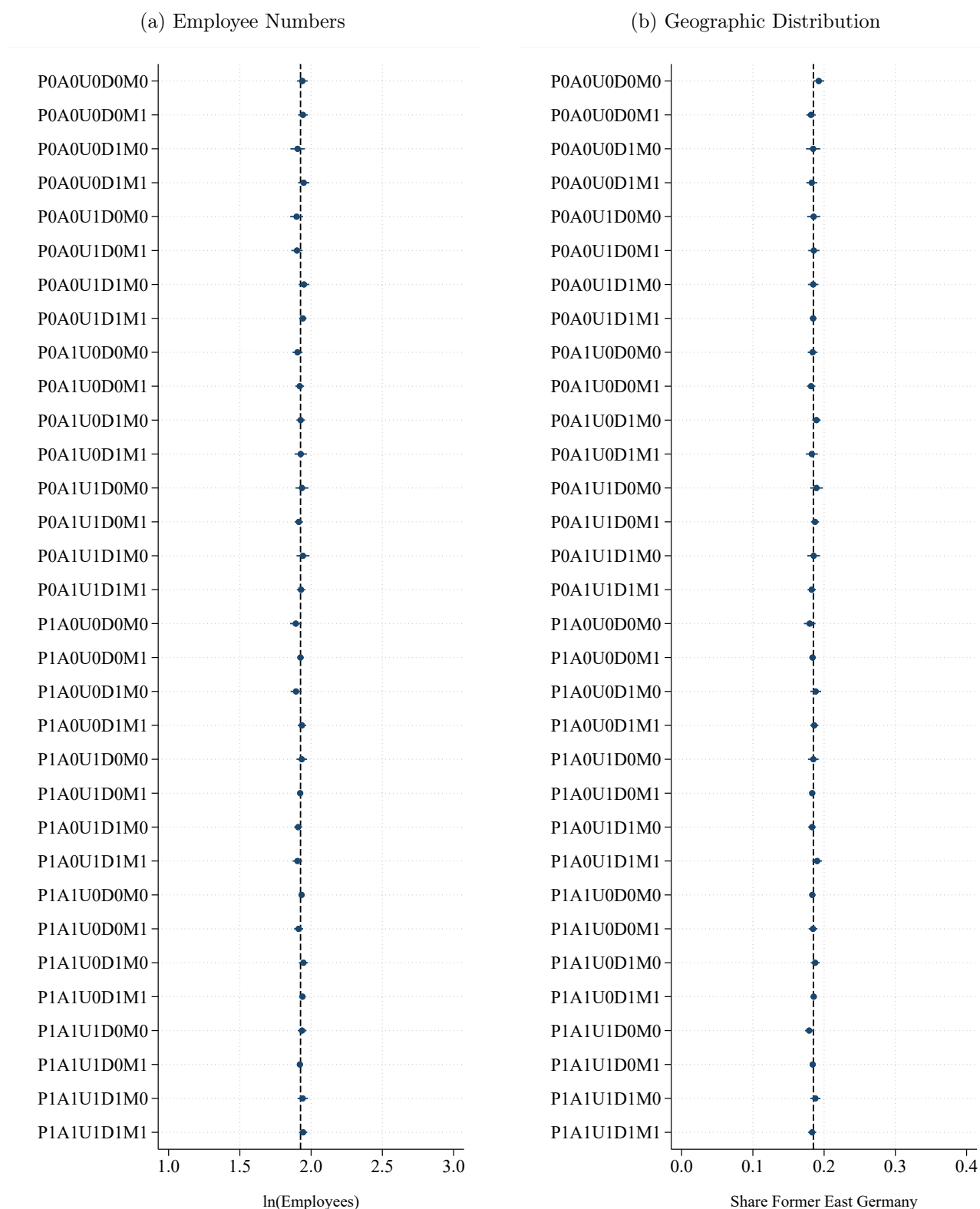
Notes: This figure evaluates the success of the randomization procedure in balancing key firm characteristics across invitation message alternatives during the experiment. The panels display (a) the average logged number of employees and (b) the proportion of invitations sent to firms located in the former East Germany for each message version. Whiskers represent 95% confidence intervals, and the black dashed line indicates the unweighted average across all message versions. Employee numbers are presented in logarithms to account for the high skewness in firm size across the sample, which includes both private and public firms. Wald-tests indicate no significant differences in staffing numbers or geographic distribution across the 32 message versions.

Figure A.2: Spatial Distribution of Sent Invitations

(a) Rank 1

P0A0U0D1M0

(b) Rank 4

P0A1U1D0M0

(c) Rank 8

P1A0U0D1M0

(d) Rank 12

P1A1U1D0M0

(e) Rank 16

P0A0U0D0M1

(f) Rank 20

P0A1U0D0M1

(g) Rank 24

P1A1U1D1M1

(h) Rank 28

P1A0U0D0M1

(i) Rank 32

P1A1U1D0M1
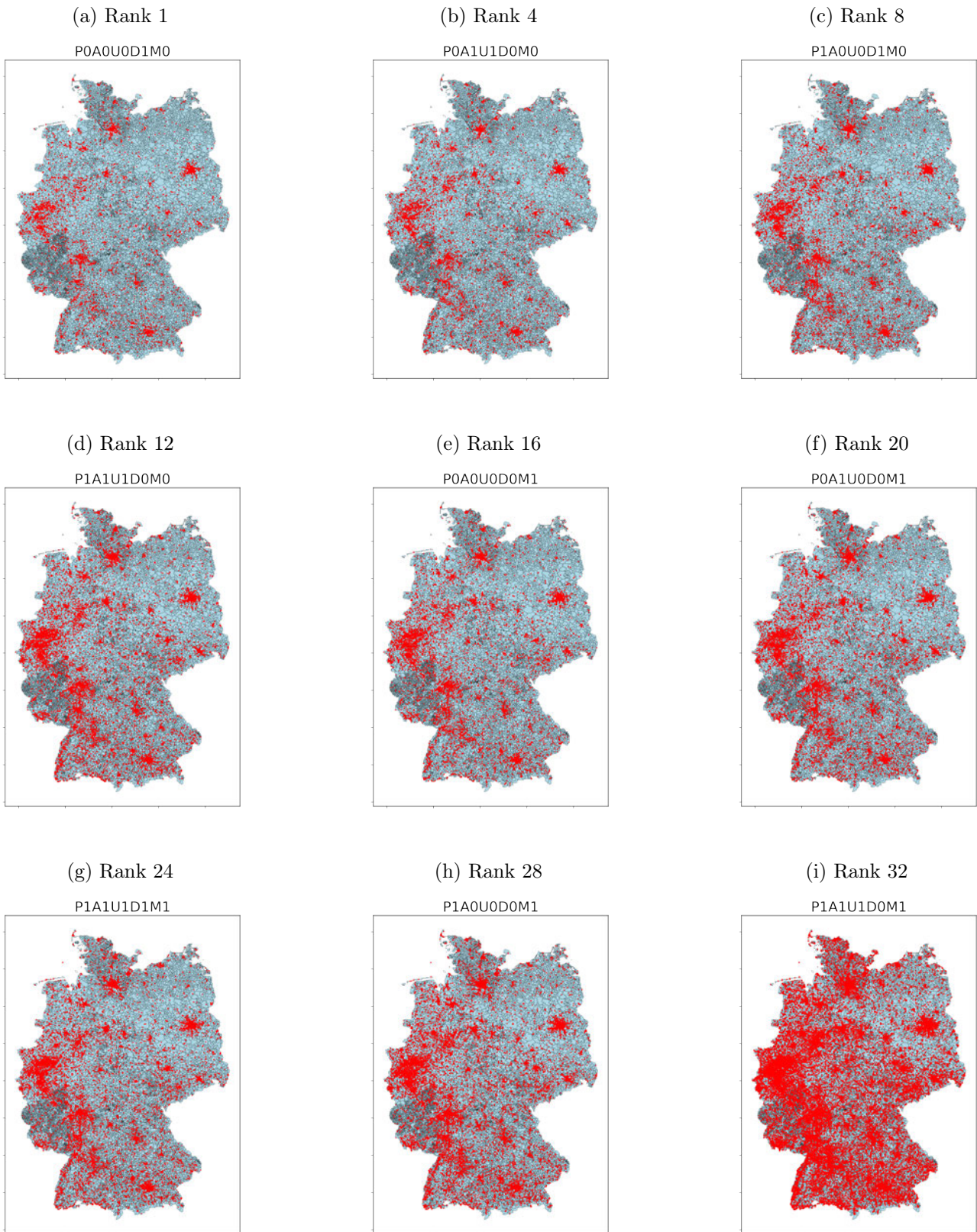
Notes: The maps in this figure display the spatial distributions of survey invitations sent to businesses located in Germany for nine selected message variants. The message versions are ranked by their cumulative distribution share in the experiment and shown in increments of four, also mapping out the lowest-ranked version (P0A0U0D1M0). Each red dot represents a firm that received a particular invitation message. As the cumulative distribution share increases, more firms are reached, particularly in densely populated areas. However, the spatial pattern remains consistent, indicating no specific geographic clustering for any message variant.

Figure A.3: Share of Opened Invitations per Message Variant

Figure A.4: Predicted Difference in Starting Rates using OLS and Batched OLS
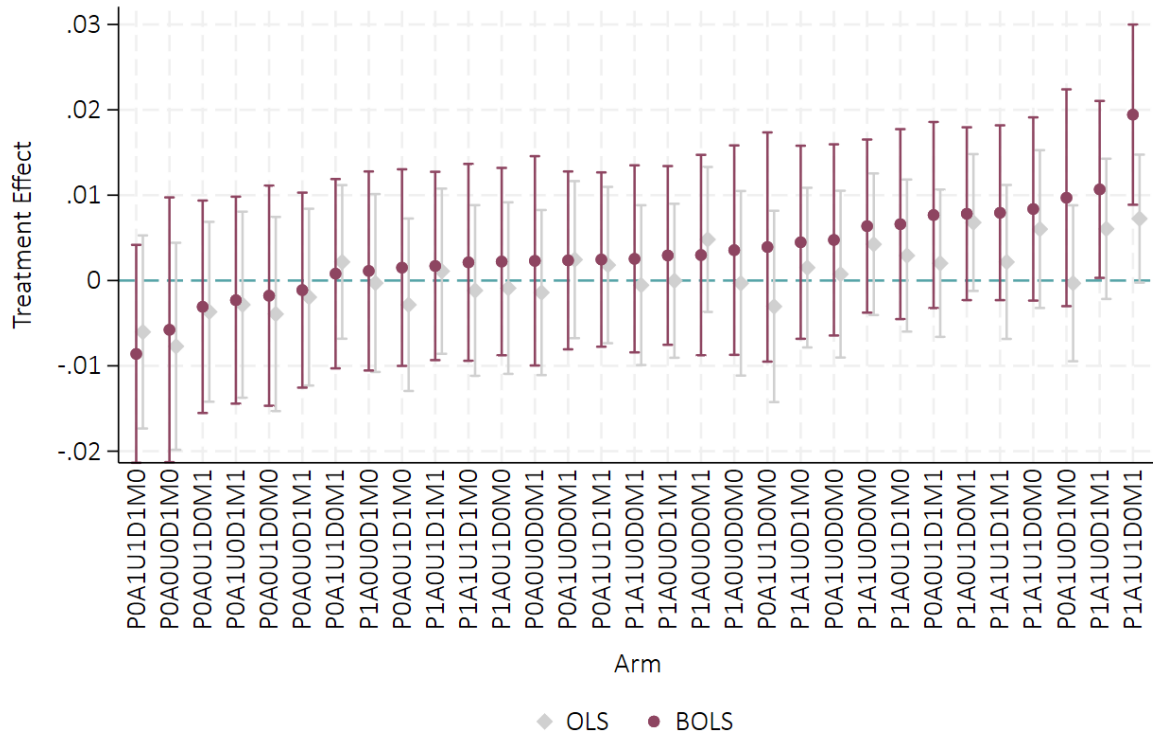
Notes: This figure illustrates margins, i.e., the difference in expected starting rates between a invitation message alternative and the reference message (P0A0U0D0M0) using Ordinary Least Squares (OLS) and batched OLS (BOLS) as proposed by Zhang et al. (2020). OLS predictions are shown in gray, while BOLS predictions are presented in red. Whiskers indicate 95% confidence intervals. The starting rate is defined as the percentage of firms that commence the survey after opening the invitation email. The results using OLS and BOLS are mostly comparable. For the message alternative P1A1U1D0M1, BOLS suggests a higher reward than OLS.

B Invitation Message Alternatives

Figure B.1: Invitation Message P0A0U0D0M0 (Original Language, German)

Sehr geehrte Damen und Herren,

wir möchten Sie herzlich zur Teilnahme an der aktuellen Unternehmensbefragung durch das German Business Panel einladen. Im Fokus der Erhebung stehen sowohl aktuelle Themen (z.B. die betriebswirtschaftlichen Folgen des Russland-Ukraine-Kriegs) als auch Fragen zu den Effekten von Regulierung in den Bereichen der Rechnungslegung und Unternehmensbesteuerung.

Ihre Sicht der Dinge ist ausgesprochen wichtig für die Forschung! Denn Ihre Angaben helfen, die Wirkungen und Nebenwirkungen politischer Maßnahmen auf Unternehmen in Deutschland zu verstehen. Die Bearbeitungszeit beträgt ca. 5 bis 15 Minuten.

Haben Sie noch Fragen?
Sie erreichen uns unter der Telefonnummer +49 (0) 621 181 3286 oder per Mail (gbpinfo@mail.uni-mannheim.de). Weitere Informationen sowie Ergebnisse bisheriger Befragungen finden Sie auf unserer Projektseite https://gbpanel.org/. Ihre Angaben werden vertraulich behandelt und die Forschungsergebnisse anonymisiert verwendet.

Sie können die Befragung unter folgendem Link starten:
URL-LINK

Wir danken Ihnen für Ihre Teilnahme!

Mit freundlichen Grüßen

Jannis Bischof, Dirk Simons, Caren Sureth-Sloane, Johannes Voget
Die wissenschaftliche Leitung des German Business Panels

_____

Sollten Sie nicht teilnehmen wollen, klicken Sie bitte auf folgenden Link:
URL-Link

Notes: This figure shows the invitation message P0A0U0D0M0 in its original German form. The highlighted passages indicate treatment characteristics that are altered in other message variants. This version is unpersonalized (green), has a low-authority sender (yellow), provides the participation link at the bottom (red), does not emphasize data protection (purple), and offers the opportunity to share opinions (grey). The message containing the exact opposite treatment specifications is presented in Appendix B.2. The English translation of this message is provided in Appendix B.3

Figure B.2: Invitation Message P1A1U1D1M1 (Original Language, German)

Sehr geehrte Damen und Herren,

wir bitten Sie – für Ihr Unternehmen [*Unternehmensname*] – um Ihre Teilnahme an der aktuellen Unternehmensbefragung durch das German Business Panel. Im Fokus der Erhebung stehen sowohl aktuelle Themen (z.B. die betriebswirtschaftlichen Folgen des Russland-Ukraine-Kriegs) als auch Fragen zu den Effekten von Regulierung in den Bereichen der Rechnungslegung und Unternehmensbesteuerung.

Sie können die Befragung unter folgendem Link starten:
URL-LINK

Ihre Sicht der Dinge ist ausgesprochen wichtig für die Forschung! Nur mit Ihren Angaben können wir dazu beitragen, dass Wirkungen und Nebenwirkungen politischer Maßnahmen auf Unternehmen in Deutschland verstanden werden. Die Bearbeitungszeit beträgt ca. 5 bis 15 Minuten.

Informationen zum Datenschutz
Ihre Angaben werden streng vertraulich nach der EU-Datenschutz-Grundverordnung (DSGVO) behandelt. Die Forschungsergebnisse werden ausschließlich in anonymisierter Form verwendet.

Haben Sie noch Fragen?
Sie erreichen uns unter der Telefonnummer +49 (0) 621 181 3286 oder per Mail (gbpinfo@mail.uni-mannheim.de). Weitere Informationen sowie Ergebnisse bisheriger Befragungen finden Sie auf unserer Projektseite https://gbpanel.org/.

Bitte helfen Sie mit Ihrer Teilnahme!

Mit freundlichen Grüßen

Prof. Dr. Dr. h.c. Dr. h.c. Caren Sureth-Sloane (Universität Paderborn), Prof. Dr. Jannis Bischof, Prof. Dr. Dirk Simons, Prof. Dr. Johannes Voget (alle Universität Mannheim)
Die wissenschaftliche Leitung des German Business Panels

_____

Sollten Sie nicht teilnehmen wollen, klicken Sie bitte auf folgenden Link:
URL-Link

Notes: This figure shows the invitation message P1A1U1D1M1 in its original German form. The highlighted passages indicate treatment characteristics that are altered in other message variants. This version is personalized (green), has a high-authority sender (yellow), provides the participation link near the top (red), emphasizes data protection (purple), and pleas for help (grey). The English translation of this message is provided in Appendix B.4

Dear Sir or Madam,

we would like to cordially invite you to participate in the current business survey conducted by the German Business Panel. The survey focuses both on current topics (e.g., the economic consequences of the Russia-Ukraine war) as well as on questions about the effects of regulation in the areas of accounting and corporate taxation.

Your perspective is extremely important for research! The information you provide helps us to understand the effects and side effects of political measures on businesses in Germany. The processing time is approx. 5 to 15 minutes.

Do you have any questions?
You can reach us by phone at +49 (0) 621 181 3286 or by email (gbpinfo@mail.uni-mannheim.de). Further information and the results of previous surveys can be found on our project page: https://gbpanel.org/. Your responses will be treated confidentially, and the research results will be used anonymously.

You can start the survey via the following link:
URL-LINK

Thank you for your participation!

Sincerely,

Jannis Bischof, Dirk Simons, Caren Sureth-Sloane, Johannes Voget
The scientific board of the German Business Panel

_____
Should you not wish to participate, please click on the following link:
URL-Link

Notes: This figure shows the invitation message P0A0U0D0M0 in its English translation. The highlighted passages indicate treatment characteristics that are altered in other message variants. This version is unpersonalized (green), has a low-authority sender (yellow), provides the participation link at the bottom (red), does not emphasize data protection (purple), and offers the opportunity to share opinions (grey).

Figure B.4: Invitation Message P1A1U1D1M1 (English Translation)

Dear Sir or Madam,

we ask you – on behalf of your business [*business name*] – for participation in the current business survey conducted by the German Business Panel. The survey focuses both on current topics (e.g., the economic consequences of the Russia-Ukraine war) as well as on questions about the effects of regulation in the areas of accounting and corporate taxation.

You can start the survey via the following link:
URL-LINK

Your perspective is extremely important for research! Only with your information can we help to understand the effects and side effects of political measures on businesses in Germany. The processing time is approx. 5 to 15 minutes.

Do you have any questions?
You can reach us by phone at +49 (0) 621 181 3286 or by email (gbpinfo@mail.uni-mannheim.de). Further information and the results of previous surveys can be found on our project page: https://gbpanel.org/.

Information on data protection
Your information will be treated strictly confidentially in accordance with the EU General Data Protection Regulation (DSGVO). The research results will be used exclusively in anonymized form.

Please help by participating!

Sincerely,

Prof. Dr. Dr. h.c. Dr. h.c. Caren Sureth-Sloane (University of Paderborn), Prof. Dr. Jannis Bischof, Prof. Dr. Dirk Simons, Prof. Dr. Johannes Voget (all University of Mannheim)
The scientific board of the German Business Panel

_____

Should you not wish to participate, please click on the following link:
URL-Link

Notes: This figure shows the invitation message P1A1U1D1M1 in its English translation. The highlighted passages indicate treatment characteristics that are altered in other message variants. This version is personalized (green), has a high-authority sender (yellow), provides the participation link near the top (red), emphasizes data protection (purple), and pleas for help (grey).

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//

IMPRINT