BITA

BITA WORKING PAPER SERIES

Using reminders with reward opportunities to reduce spectator no-show behavior

Dominik Schreyer, Steven J. Bickley, Ho Fai Chan, Sascha L. Schmidt, Benno Torgler

BITA *Working Paper No.* 4

QUT the university for the real world
THE UNIVERSITY OF WESTERN AUSTRALIA
THE UNIVERSITY OF QUEENSLAND AUSTRALIA
Australian Government Australian Research Council

Any opinions expressed in this paper are those of the author(s) and do not reflect the views of ARC Training Centre for Behavioural Insights for Technology Adoption (BITA). Research published in this series may discuss policy implications, but BITA does not endorse any specific policy positions.

BITA is committed to the highest standards of research integrity, guided by our principles of transparency, academic independence, and evidence-based analysis. BITA is an institution that focuses on advancing knowledge in behavioural science research. Our mission is to provide insights that inform policymakers and contribute to global economic and social challenges.

BITA Working Papers represent preliminary research circulated to encourage discussion and critical feedback. As such, these papers may evolve over time. Citation of this work should recognize its provisional nature, and the final version may be available from the author(s) directly.

Using reminders with reward opportunities to reduce spectator no-show behavior

Dominik Schreyer, Steven J. Bickley, Ho Fai Chan, Sascha L. Schmidt, and Benno Torgler*

Abstract

Because no-show behavior typically leads to operational inefficiencies and diminishing returns for service firms, in particular, a growing number of authors have demonstrated the potential of using reminders to reduce no-show rates. However, given the highly contextually driven nature of these previous studies, it remains unclear whether reminders are also an effective means to reduce no-show behavior in a sports environment. Analyzing individual behavioral responses of 13,911 season ticket holders of a German football team, we conduct an experimental exploration of the potential for using reminders with reward opportunities in a sporting setting. According to the results of our large-scale experiment, reminders only have a short-term effect that operates most effectively through a monetary gift reward with no persistent behavioral change in subsequent games. As such, replicating and extending previous research findings in alternative environments is vital for avoiding premature assumptions about the potential effectiveness of otherwise already well-established measures.

* Schreyer: WHU – Otto Beisheim School of Management, Erkrather Str. 224a, 40233, Düsseldorf, Germany (e-mail: dominik.schreyer@whu.edu; corresponding author); Bickley: Queensland University of Technology (QUT), QUT Business School, School of Economics and Finance, Gardens Point 2 George St, Brisbane QLD 4000, Australia (e-mail: s.bickley@qut.edu.au); Chan: Queensland University of Technology (QUT), QUT Business School, School of Economics and Finance, Gardens Point 2 George St, Brisbane QLD 4000, Australia (e-mail: hofai.chan@qut.edu.au); Schmidt: WHU – Otto Beisheim School of Management, Erkrather Str. 224a, 40233, Düsseldorf, Germany, and CREMA – Center for Research in Economics, Management and the Arts, Switzerland, and LISH – Lab of Innovation Science at Harvard, 175 N. Harvard Street Suite 1350, Boston, MA 02134, USA (e-mail: sascha.schmidt@whu.edu). Torgler: School of Economics and Finance, Queensland University of Technology, Level 8, Z Block, Gardens Point Campus, 2 George Street, Brisbane, QLD 4001, Australia, and CREMA – Center for Research in Economics, Management and the Arts, Switzerland (e-mail: benno.torgler@qut.edu.au).

Keywords: attendance, fan behavior, season ticket holders, sports stadium atmosphere, team loyalty, ticketing

Acknowledgments: This research is based on a chapter in the doctoral dissertation of the first author under the supervision of the fourth and fifth author, a previous version of this paper was published as a working paper (c.f., Schreyer et al., 2020). We did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We extend our gratitude to the associate editor and the two anonymous reviewers for their insightful suggestions and constructive feedback.

Availability of data: The data that support the findings of this study are available on reasonable request from the corresponding author. The data are not publicly available due to privacy restrictions.

Disclosure statement: SJB, HFC, and BT are co-founders of the company/research lab that developed *SurveyLM*, and therefore, may appear, or be perceived to have, competing interests. The authors report there are no other competing interests to declare.

JEL: C93, D12, L83, R22, Z20

Word count: 7,847 words

1. Introduction

For service firms, no-show behavior (NSB) – a specific form of absenteeism (c.f., Amberger and Schreyer, 2024) – leads to operational inefficiencies and diminishing returns. In the health care industry, for example, the no-show rate (NSR) averages 23 percent, ranging from 4 to 79 percent depending on medical specialty (Dantas et al., 2018). In this setting, patient NSB not only imposes operational costs (Kheirkhah et al., 2016) and fosters patient stereotypes (Hussain-Gambles et al., 2004), but also increases waiting times for others. In England alone, over 15 million general practice appointments are missed each year (National Health Service [NHS], 2019), many of which could have been reassigned. However, NSB is increasingly observed across industries, including beauty, transportation, and sports (e.g., Association Football: Schreyer, 2019; Australian Rules Football: Karg et al., 2021; College Football: Popp et al., 2023; College Basketball: Popp et al., 2024).

In European association football, in particular, NSB has recently become a matter of top management concern (e.g., Borussia Dortmund, 2018), where the importance of filling empty seats extends beyond just income generated from ticket sales. For example, (supportive) crowds also contribute to a lively atmosphere that benefits not only other present and potential spectators (e.g., Oh et al., 2017; c.f., Schreyer and Behrens, 2024), but also significant external stakeholder groups like broadcasters, corporate sponsors, investors, and hospitality section members (e.g., Behrens and Uhrich, 2022; Gimet and Montchaud, 2016; McDonald, 2010). Thousands of empty seats, in contrast, may not only significantly reduce match-day income (e.g., from selling food and beverages, merchandise, and parking tickets – not to mention the potential positive economic externalities for other businesses in the area),[1] but could also diminish an otherwise still significant home advantage (e.g., Bryson et al., 2021).

[1] To illustrate such a potential no-show effect, consider the spending behavior of a European spectator, which averages 5 Euros per fan per match (cf. ESSMA, 2018). Spending tends to be higher at newer arenas, averaging 6.14 Euros, and can sometimes reach up to 12 Euros. With an average no-show rate of approximately 12 percent

In the broader experimental literature, authors have suggested various strategies to reduce such NSB, including fines (Lesaca, 1995), incentives (Smith et al., 1990), or treating outpatients at home (Anderson and Aquilina, 2002), among others to reduce opportunity costs. Alternatively, predictive overbooking can preemptively fill expected no-show appointments (LaGanga and Lawrence, 2007), while reminders address forgetfulness, which typically ranks among the most common explanations for NSB (Herrick et al., 1994). Intriguingly, while reminders are the most prevalent strategies, they may need to be adjusted to the specific context (cf. Amberger and Schreyer, 2024).

Accordingly, in this paper, we conduct the first experimental exploration of the potential for reminders with different reward opportunities to reduce NSB among sports spectators. More specifically, because football spectator NSB is more prominent among season ticket holders (STHs; Schreyer, 2019), who significantly contribute to home advantage and stadium atmosphere (e.g., Uhrich and Benkenstein, 2010), our field experiment tracked the responses of 13,911 STHs to five different treatments designed to encourage their matchday admission.

While reminders have been shown to nudge behavior before (e.g., Calzolari and Nardotto, 2017), in the media-saturated European football market, addressing forgetfulness alone seems insufficient. Instead, limited research suggests that STH NSB is driven by perceived costs (e.g., travel distance; cf. Schreyer et al., 2018) and perceived utility (e.g., seat quality; cf. Schreyer et al., 2016). Therefore, we enhanced our reminders by incorporating different rewards addressing STHs' desire to support their club or offering probabilistic prizes. Accordingly, our study contributes to the literature on reminder effectiveness by exploring the

(cf. Schreyer, 2019), Bundesliga clubs, whose 18 teams collectively distributed 12,089,058 tickets in the 2023-24 season or 39,506 tickets per match, most of which have relatively new stadiums, would miss out on significant revenue. Specifically, they could likely lose between about 8.9 and 17.4 million Euros in total revenue based on the following calculation: 6.14/12 Euros x 306 matches x 4,741 no-shows (39,506 distributed tickets x 12 percent no-show). Naturally, in other markets such as the US market, with higher spectator spending, the effects would be even stronger.

impact of combining reminders with rewards, which has been largely overlooked in efforts to reduce NSB, in an as yet ignored industry context.

Intriguingly, our results suggest that single reminders, even when paired with rewards, may be largely ineffective in reducing STH NSB, particularly over the long term. While valuable probabilistic rewards showed a short-term effect, most other rewards had little impact on reducing NSR, or did so only in specific sub-groups. This variation indicates that STHs are not a uniform group and respond differently to interventions. Our findings, therefore, highlight the importance of replicating and extending research on reminder effectiveness across contexts to avoid premature assumptions about their overall potential.

2. Material And Methods

2.1 Experimental Setting

The Bundesliga, founded in 1962, is Germany's primary football competition, contested by 18 professional clubs from August to May. In this period, each team plays 34 matches, with 17 matches played at home. Unlike most American leagues, Bundesliga operates on a system of promotion and relegation to Germany's second football division, Bundesliga 2. With an average of 38,973 tickets distributed per match (DFL, 2024), Bundesliga attracts large crowds, whether regular or occasional attendees, often watching in groups (Melnick, 1993). Within this ecosystem, STHs play a decisive role, typically accounting for over half of the roughly 12 million tickets distributed each season (c.f., Figure A1 in the supplementary Appendix).

2.2 Participants

Our experiment involved STHs from a well-established Bundesliga club. Before the 2012−13 season, 19,390 season tickets were sold: 17,805 to the general public and an additional 1,585 to external stakeholders such as corporate sponsors, partners, and hospitality section members.

In addition to the handful of STHs who relinquished their status for the upcoming season before our experiment began, we excluded a total of 5,479 records from the original data set based on our primary interest in individual STH behavioral responses (c.f., Appendix C). Our final sample thus consists of 13,911 STHs located throughout the public area of the stadium. According to the descriptive statistics (c.f., Table A1 in the supplementary Appendix), most STHs in our sample are middle-aged males who on average traveled 24 kilometers (roughly 15 miles) to the stadium to attend a match.

2.3 Treatments

Our treatment designs were informed by previous studies that aimed to increase intervention effectiveness by combining reminders with additional contextual information beyond just the date, time, and location of the scheduled appointment (e.g., Henderson, 2008). For example, as evident from the literature on STH behavior and motivations (McDonald et al., 2024), a STH's decision to attend a match typically results from weighing benefits against (opportunity) costs. In particular, previous research in the European football market has found that STH NSB is more likely if an individual holds tickets for lower-quality seats (e.g., Schreyer and Torgler, 2021), lives outside the home club's city (Schreyer et al., 2016; 2018), leading to longer travel, or holds multiple tickets, especially in the family section (e.g., Schreyer and Torgler, 2021), which increases coordination costs. Intriguingly, this is also reflected in the inverted U-shaped, non-linear relationship between STH age and NSB (e.g., Amberger et al., 2023), peaking in the mid-40s—a time when discretionary time is limited, while income is relatively high. Consequently, while neither we nor the club can alter an individual's perceived cost (e.g., by persuading them to move closer) or schedule a more attractive opponent—the most prominent antecedent of NSB (cf. Schreyer, 2021)—our reminders aim to increase the perceived utility of attending by offering various rewards. These include appealing to STHs' core motivation of

feeling like true supporters (cf. Karg et al., 2019), influencing the club's performance on and off the pitch, and fostering a sense of belonging or offering (valuable) probabilistic rewards.

Consequently, we sent out five types of reminders, each offering different levels of reward opportunities: (1) a simple repetition of the basic appointment information combined with a low moral suasion trigger, (2) a repetition with a strong moral suasion trigger, (3) a repetition with an offer of public recognition, and a repetition with the chance to win either (4) a unique experience or (5) an expensive reward (c.f., Table B2 in the supplementary Appendix). Although not linked to the (re-)purchase decision, STHs are typically informed about Bundesliga fixtures well in advance —often weeks, and sometimes even months before the match. Hence, all five of our reminders began with the following sentence, providing the basic appointment information: "Dear [NAME], next [WEEKDAY], [DATE], [KICK-OFF TIME], we welcome [VISITING CLUB] to our [STADIUM]."

All reminders were personalized letters, signed by the Bundesliga club's spokesperson, and sent out to 11,599 STHs three days before the day of the *14th home match*. In contrast, no such reminders were sent to the remaining 2,312 STHs, who served as our control group. Since we cannot assume that all recipients read the letter, we interpret the effects as an *intention to treat*, which may underestimate the actual treatment effect. In consultation with the club, we agreed not to inform STHs that they were part of a field experiment. To address potential inquiries from fans or the media, we instructed all employees working in fan and press relations (e.g., in the ticketing department) to direct any questions about the letters to the club or its sponsorship activities. This approach appeared successful, as there were no media reports or related social media activity before the match.

2.3.1 Moral Suasion

To trigger moral suasion (i.e., elicit altruistic behavior; Romans, 1966), our first two treatments leveraged the fact that a STH's desire to support the club is typically the primary motivator of a season ticket purchase (Karg et al., 2019). By specifically appealing to this strong emotional connection between STHs and their club, moral suasion taps into their loyalty and commitment, emphasizing their role as vital contributors to both the team's success and the stadium atmosphere. We expect this sense of belonging among other STHs, along with the social identity of being a true supporter, to encourage matchday admission, even if STHs are undecided or inclined to opt out, as strong team identification is a known driver of admission behavior (cf. Schreyer, 2019). Additionally, subtle social pressure to maintain their reputation as dedicated fans, combined with the understanding that their presence strengthens the home-field advantage, serves as a powerful motivator for positive behavioral change. These principles guided the first two types of reminders ((1) and (2)). The text of the first reminder, labeled $T_{MORAL/low}$, translates from German as follows:

> Having lost both previous encounters, […], we expect a tough match, for which we must mobilize all our forces!

The content of our second reminder, labeled $T_{MORAL/high}$, includes a more explicit trigger, making it considerably stronger:

> The enormous level of support received by our team this season is certainly of great significance for our overall performance. Without the support of our loyal season ticket holders, the stadium atmosphere suffers tremendously and thus affects our team's performance. As a season ticket holder, you experience events in the Bundesliga first-

hand week after week, you motivate our team and thus contribute to a unique stadium experience for all of us. Please provide support for [HOME CLUB] next weekend in the match against [VISITING CLUB]. Because one thing is for sure: Your support plays a vital role in the success of our team!

2.3.2 Recognition

Our third treatment offered to reward STH support during the season through public acknowledgement, a strategy proven effective in alternative settings (e.g., Markham et al., 2002). Similarly, in our environment, we proposed mentioning STHs by name, representing all STHs, in the club's matchday magazine, distributed before the 17th and final home match. This treatment taps into the social and psychological benefits of public recognition, potentially strengthening STHs' sense of pride and belonging. By publicly acknowledging their loyalty, we aim to reinforce their identity as dedicated supporters. In this regard, the prospect of recognition serves as a strong motivator, validating the STH's commitment and enhancing the social standing within the broader fan community. Additionally, knowing their support is appreciated by the club can deepen their emotional connection, encouraging greater engagement and match admission. Consequently, the text of the third reminder, labeled TRECOGNITION, translates from German as follows:

> The enormous level of support received by our team during this season is of great significance for the performance of both the team and each individual player. For this we would like to say Thank you! We therefore want to surprise you and a small number of other season ticket holders by mentioning you by name, representatively for all our fans, in the [STADIUM] magazine, available in the run-up to our last home match against [VISITING CLUB #2].

Noticeably, in this treatment group, as in all others, STH NSB applied, on average, to about two matches. In other words, STHs missed an average of only two out of the first 13 home matches. Overall, however, about 51 percent of the 2,347 STHs in this group missed either only one or no home match. We therefore concluded that receiving a reminder recognizing individual support was highly unlikely to irritate these STHs.

2.3.3 Probabilistic Reward

Lastly, based on empirical evidence that combining reminders with (delayed) probabilistic rewards can induce behavioral change (Parrish et al., 1986), our fourth and fifth treatments, collectively labeled $T_{EXTRINSIC}$, informed STHs that by attending the upcoming fixture, they could win either a prize (T_{GIFT}) or a unique experience ($T_{EXPERIENCE}$). Probabilistic rewards, which offer a chance at a desirable outcome, can be particularly motivating as they introduce anticipation and excitement. In our specific context, for instance, STHs may be motivated to attend a match not only for their love of the team (or an attractive opponent) but also for the opportunity to win a tangible reward or exclusive experience, thereby increasing the perceived utility from physical admission.

> We are pleased to inform you that, next match-day, we will gift one of our season ticket holders [a spa trip worth roughly EUR 5,000 (T_{GIFT}) or a dinner with our team captain ($T_{EXPERIENCE}$)]. You are one of a lucky 2,500 randomly preselected season ticket holders who now has the opportunity to enter the prize draw. To participate, all you need do is attend the match against [VISITING CLUB]. Once our stadium access control system recognizes your season ticket [BLOCK NUMBER; ROW NUMBER; SEAT NUMBER], you will be automatically entered into the draw. The winner will be notified promptly.

2.3.4 Treatment Checks

To evaluate the effectiveness of our treatments and their perceptual reception among STHs, we conducted a series of ex-post treatment checks through agent-based simulations with synthetic agents powered by Large Language Models (LLMs) like OpenAI's GPT-3.5-turbo and GPT-4o (as tested in this study). The models are prompted with i) *demographic characteristics* (e.g., age, gender, admission record) of a randomly drawn subset (n=1,500) of the total observations, and ii) the *treatment letters* (i.e., presented as the task or 'question') and a 17-item question block (i.e., the expected structure/formatting of the task deliverable/s or 'answer instruction'), focused on probing dimensions such as message relevance, personal connection to/feelings toward the club, and motivation/intentions to attend future matches.

Our reasons for exploring the utility of LLM-based agents for conducting our (post-study) treatment checks are two-fold. First, the inherent challenges/constraints associated with pre-testing stimuli directly on a subset of STHs (i.e., ex-ante, or before the field experiment occurred), primarily due to practical constraints and the potential risk of information leakage that could compromise the experimental integrity. Second, the cost and time efficiencies enabled by the possible substitution of human study participants with LLM-based synthetic agents (e.g., prompted with a subset of the demographic characteristics from a randomly drawn STH of the original dataset). Consequently, we adopted an LLM-based agent simulation approach to evaluate whether our stimuli elicited the intended treatment effects. While we are confident that all five treatments can be considered to be credible and were perceived as intended (cf., Appendix B; Table B3 – B6, as well as Figure B1), we acknowledge (and discuss)

the potential limitations of our approach in the final section of our supplementary Appendix B.[2]

2.4 Procedure an average

2.4.1 Randomization and the Necessary Check for Balances

To balance unobserved characteristics across the five treatments and the control conditions, we randomly allocated all 13,911 STHs into six similarly sized groups. Given the large sample size, we used single random draw approach as our preferred randomization method (c.f. Bruhn and McKenzie, 2009). This randomization process was semi-public, with the study authors present during the draw, though neither club officials nor reminder recipients witnessed it. A post-randomization assessment revealed no significant group differences between the six conditions (c.f., Table C1 and C2 in the supplementary Appendix), including pretreatment NSB patterns.

2.4.2 Dependent Variable

Information on individual STH NSB, our key dependent variable, was recorded using the club's stadium access system. In line with previous literature, NSB is binary coded, with a value of 1 assigned if an STH failed to attend the stadium on a particular matchday, and 0 otherwise.

It is, perhaps, worth noting that the club also shared information on STH stadium admission times, captured in minutes prior to kick-off. To test whether the experimental treatments influenced STH arrival times, we employed a Heckman (1979) selection model to account for the decision to attend post-treatment matches. In general, we did not observe any significant treatment effect on matchday 14 entry times, except for the recognition treatment, which

[2] Intriguingly, as one reviewer noted, the use of artificial intelligence (AI) in general, and *SurveyLM* in particular, provides a reasonable path for conducting manipulation checks with synthetic datasets, especially in environments where running multiple experiments is costly or impractical. However, as indicated, we also acknowledge the potential limitations of this approach in our appendix.

showed a relatively small effect (treatment difference of 1.76 minutes). Therefore, in our analysis, we focus on the STHs' initial no-show decision.

3. Results and Discussion

We first report the behavioral responses to our five treatments for home match 14, i.e., the first home match after treatments were implemented, and then scrutinize potential treatment effects on the three subsequent (home) match-days, either one or two weeks apart from each other. In doing so, we address the possibility that whereas STHs offered the public recognition treatment ($T_{RECOGNITION}$) may have been likely to attend the last home match of the season when the stadium magazine featured their name, STHs in the two tangible reward groups (T_{GIFT} and $T_{EXPERIENCE}$) who did not win the prize may not have attended subsequent match-days out of disappointment. Nevertheless, the presented estimates can only be interpreted as intention-to-treat effects as STHs may not have received or opened the mailed reminders despite the current and up-to-date postal information for seasonal ticket delivery.

3.1 No-show Behavior

In Figure 1, we show the NSR of each experimental arm (a) and the average treatment effects on admission at home match 14 for the five treatment arms (b). Overall, we find the average NSR for high moral suasion ($T_{MORAL/high}$; *Diff* = 1.41pp), the probabilistic reward with gift (T_{GIFT}; *Diff* = 3.35pp) and the dinner-with-captain experience ($T_{EXPERIENCE}$; *Diff* =.14pp) lies below that of our control group. However, only the improvement in average STH admission for T_{GIFT} is statistically significant (two-tailed p = 0.0022).[3] Further, across all 13 home matches, we present pre-treatment differences in mean NSRs between each treatment arm and

[3] This effect remains statistically significant at the 5% level when accounting for multiple comparisons using the Bonferroni correction (i.e., $\alpha = 0.05/5 = 0.01$).

the control group (c.f., Figure 1b). As the treatment effect of T_{GIFT} lies below the smallest value of all placebo treatment effects observed in the pre-treatment period, it is unlikely that the observed effect is due to mere noise.

To further scrutinize the observed treatment effect, we employ a probit regression model, explicitly controlling for STH characteristics (e.g., age, including its squared term, gender, home-to-stadium distance, admission habits), as well as ticket price, match quality (e.g., away team market values, match outcome uncertainty, and geographical derbies), and the potential opportunity costs of attending a home match (e.g., fixture/scheduling, matchday/seasonal progression, and the temperature/weather). In these models, we estimate the average treatment effect via a difference-in-difference structure, using the first 13 home matches as pre-treatment periods (c.f., Table D1 in the supplementary Appendix).[4] Notably, the reported results are remarkably robust, although controlling for match characteristics or fixed effects slightly reduces the estimated effect size.[5]

Thus, exploiting this reminder at a similar professional football club with 20,000 STHs and an average NSR of 18% could reduce no-shows by approximately 560 to 740, based on the control group's NSR. Nonetheless, while this significant decline could enhance the stadium atmosphere, and, in turn, product perception (e.g., Behrens and Uhrich, 2022), its overall economic significance may be limited.

- - - Insert Figure 1 about here - - -

Intriguingly, despite the strong immediate effect of our gift treatment on STH admissions, we do not observe any significant treatment effects on subsequent home matches.[6] Specifically, we estimate the treatment effect for each subsequent matchday using the probit

[4] We report the predicted probability differences in the NSR at home match 14 between the treatment and control group based on probit estimates, with the full probit model presented in Table D2 in the supplementary Appendix.
[5] Likewise, statistical significance remains at 1% level after adjusting for multiple comparisons using the Bonferroni method.
[6] Apart from $T_{EXPERIENCE}$ in home match 17, where the treatment effect is positive —indicating reduced STH admissions—and statistically significant at the 10% level.

regression specifications similar to those in the supplementary Table D1, with results presented in Figure D1.

Consequently, our analysis strongly suggests that using single reminders to reduce STH NSB may be mostly ineffective, despite offering an additional reward, particularly in the long term. While this finding does not imply that reminders cannot induce permanent behavioral changes, but rather, as previous research has shown (e.g., Charness and Gneezy, 2009), behavioral change takes time and may reverse once the stimulus is removed. In other words, a single reminder, without repeated follow-up or a substantial reward, has limited nudging power and no effect beyond the immediate match.[7] This efficacy may also be hindered by selection effects, such as the natural tendency of fans to develop strong identification with the team and engage in supportive behavior over time (Koenig-Lewis et al., 2017). Given high opportunity costs or the perceived value of missing a match, NSB may reflect competing priorities. This assumption is supported by our finding that reminders offering a probabilistic material reward had the most significant effect, consistent with the economics literature on monetary incentives driving short-term behavioral change (Gneezy et al., 2011). This suggests that STHs who can afford to miss multiple home matches without financial loss evaluate the utility of stadium admission on a match-by-match basis.

3.2 Heterogeneous Treatment Effects

As for underlying mechanisms, the significantly lower NSR for $T_{GIFT,}$ our second probabilistic reward intervention, suggests that the reward itself was responsible for the efficacy of this reminder. Naturally, this raises the question of whether certain interventions may be more effective for specific STH groups. For example, given Schreyer's (2019) evidence that stronger

[7] Admittedly, the extrinsic rewards treatments are expected to have a more pronounced, immediate effect as they are directly tied to attending the first post-treatment match.

levels of team identification result in reduced STH NSB, would the outcome differences be larger among those with greater levels of loyalty or support? Such a likelihood is implied by the sports management research establishing a link between team identification and behavioral intentions to attend a future sports event in the stadium (e.g., Rocha and Fleury, 2017). If this relation holds as true for actual behavior as for mere behavioral intention, then designing reminder content to trigger a particular behavioral response might not reduce STH NSB for more loyal fans, who would not have missed matches in the first place.

To explore this conjecture empirically, we calculate the number of home matches attended prior to the treatment, splitting the subjects into five groups to account for the negatively skewed distribution of prior admission: STHs that attended all previous 13 (28.4%), 11–12 (40.5%), 8–10 (21.8%), 5–7 (7.1%), and less than 5 previous home matches (2.2%). We then estimate the average treatment effect conditional on past stadium admission using a simple probit model and present the estimated treatment effects of the NSR in Figure 2.[8]

Contrary to the support-level hypothesis, we find significant treatment effects of four treatments (except for the $T_{EXPERIENCE}$ treatment) for *less* loyal STHs – in particular those who have attended only 5 to 7 matches (~40–50% past admissions).[9] The largest reduction in NSR is observed from those in the T_{GIFT} treatment, with 17pp ($p = 0.0031$) lower NSR compared to their counterparts in the control group. STHs who attended only 5–7 matches in the two T_{MORAL} treatments have a lower NSR compared to those in the control group with the observed effect for $T_{MORAL/high}$ (-15.5pp, $p = 0.0053$); slightly larger than $T_{MORAL/low}$ (-13.8pp, $p = 0.012$). In addition, we find that admission for STHs in the $T_{RECOGNITION}$ treatment increased by 12.8pp

[8] The use of a parsimonious model is justified by the randomized treatment assignments and the focus on conditional treatment effects based on cumulative past admission rates. Therefore, we estimate the NSR of the 14th home match using a cross-sectional probit model, rather than a panel structure, including interaction terms between treatment and past attendance, and controlling for other STH characteristics. Intriguingly, the results are qualitatively identical to the estimates without these controls.

[9] However, the lack of a noticeable treatment effect on STHs with high past admission rates could be due to attenuation from a ceiling effect, as these STHs are likely to attend post-treatment matches regardless of an intervention.

(p = 0.0206).[10] Moreover, for STHs with higher past admission, we observe a small reduction in NSB at match 14 from those in the T_{GIFT} treatment arm with full past admission (-2.38pp, p = 0.067) and who attended 8–10 matches (-5.18pp, p = 0.045). There is no statistically significant treatment effect for those with very low past admissions (i.e., attended fewer than 5 out of the 13 past matches).[11]

Accordingly, and perhaps surprisingly, our reminders tend to nudge those with lower, though not the lowest, past admission rates—and, presumably, lower loyalty. Perhaps for STHs with a strong, unwavering commitment, stadium admission is a given, i.e., a part of their routine and identity. In these cases, reminders may serve less of a nudge. In contrast, for those STHs who experience fluctuations in their engagement or face more barriers to regular admission, reminders with reward opportunities can play a crucial role. As we show, these reminders can act as catalysts, not just by providing practical information, but by rekindling the emotional and social motivations for physical admission, and boosting this with a form of incentivization (e.g., moral suasion, prize draws). Therefore, both reminding STHs of the joy and value they derive from being part of a community experiencing a match, as well as appealing to their aspirations and excitement for potential rewards and recognition, such as the allure of prize draws or the prestige of being featured in the club's magazine.

Next, we examine whether the treatments would be less effective on STHs who have already cancelled their season ticket for the next season. Noting that churned STHs are less likely to have attended past matches (c.f., D2 in the Appendix), we find that all treatments *increase* no-show tendency of churned STHs (Figure 3), particularly those who were exposed to the moral suasion treatments ($T_{MORAL/high}$ (10pp, p = 0.042); $T_{MORAL/low}$ (10pp, p = 0.042))

[10] While the effect of $T_{RECOGNITION}$ is not statistically significant at the 10% level if we apply the Bonferroni correction (i.e., $\alpha = 0.02$), this is due to the conservative nature of the test.

[11] Despite the relatively large conditional treatment effect of a 10.99pp no-show reduction in the T_{GIFT} treatment, the comparison may suffer from lack of statistical power due to the small number of STHs with very low admission rates (around 50 STHs in each treatment arm). Nevertheless, when grouping together STHs who attended fewer than 8 matches, the conditional treatment effects remain statistically significant, albeit with a smaller effect size.

and experience ($T_{EXPERIENCE}$ (17pp, p = 0.00048)). Such adverse effects are also statistically significant when compared to the treatment effects on the resuming STHs, except for the recognition treatment. Apparently, for STHs who have already resigned their tickets, our reminders may have been perceived as less desirable or even annoying or intrusive. Put differently, because these individuals might already have felt disconnected with the club, reminders, rather than renewing interest, may have highlighted their disengagement, thereby further reducing the intention to attend a match.

- - - Insert Figure 2 & 3 about here - - -

Moreover, we explore how proximity to the home team's city and STH's home location may also mediate the treatment effects. Using each STH's home address, we identify if they live in the same city as the home team and calculate the distance between their home and the stadium. Interestingly, we find that the treatment with the stronger moral suasion trigger ($T_{MORAL/high}$) is relatively more effective (p = 0.067) for STHs living in the same city as the home team (Figure 4a); the non-home group (those living outside this city) experienced a non-statistically significant treatment effect. In the other four treatment groups, we do not observe any statistically significant treatment difference between STHs living in the same city of the home team and STHs living away from that city.[12] One possible explanation is that STHs who live in the same city may feel a stronger local identity and connection to the team, making them more responsive to strong moral suasion, as it reinforces their role within the community. In contrast, those living farther away may experience less community-based pressure or emotional resonance with the team's locality, reducing the treatment's impact.

- - - Insert Figure 4 & 5 - - -

With respect to the distance of the STH residence from the stadium (Figure 4b), we find that both the high moral suasion treatment and also the two extrinsic, probabilistic rewards are

[12] Interestingly, STHs living outside the home city seem to respond more negatively to the recognition treatments.

more effective in reducing NSB among STHs who live closer to the home stadium. Specifically, we find the treatment differences to be statistically significant for STHs who live within 20km (T_{GIFT}), 10km ($T_{MORAL/high}$), and 5km ($T_{EXPERIENCE}$) from the stadium. As discussed earlier, the higher effectiveness of these three treatments for closer-residing STHs may be due to the relatively low opportunity costs of attending. Put differently, this proximity may make both the moral appeal and earning potential rewards more compelling, as the logistical barriers are minimized for those STHs living nearby. Conversely, for STHs living farther away, travel barriers may reduce the motivational impact of these treatments. Similarly, we again see that the treatment effects for $T_{MORAL/low}$ and $T_{RECOGNITION}$ do not seem to be strongly dependent on physical closeness to the home team.

Furthermore, we explore whether treatment effectiveness varies across ticket type (Figure 5a) and seating category, based on ticket price (Figure 5b). Intriguingly, in terms of seating category, we find that T_{GIFT} is more effective for STHs with lower-tier and mid-tier tickets but not for upper-tier ticket holders, potentially signaling that these STHs may value luxury rewards or incentives that could typically be out of their price range. Additionally, $T_{MORAL/high}$ also reduces no-shows for STHs with mid-tier pricing tickets.

Finally, we examine whether treatment effects vary by STH gender and age. Overall, we find no statistically significant differences in treatment effects between male and female STHs (cf. Figure D3 in the Appendix), although male STHs show slightly larger reductions in NSRs in response to T_{GIFT}, $T_{MORAL/high}$, and $T_{EXPERIENCE}$.[13] For age (Figure 6), we observe that three distinct age groups—18–29, 30–39, and 50–59 years old—are particularly responsive to the gift treatment. Interestingly, older female STHs (60+), however, respond positively to all treatments except for recognition. Intriguingly, these results suggest nuanced gender and age

[13] Nevertheless, when we do not control for other STH characteristics in the regression, the gender difference for $T_{MORAL/high}$ and $T_{EXPERIENCE}$ are significant at 10% level.

responses to treatment types. For instance, the generally stronger response to treatments by male STHs, though not statistically significant, may relate to social expectations stemming from cultural norms and gender-specific socialization in sports, as well as a stronger identification with admission as a habitual or social activity. Additionally, older female STHs' positive responses to most treatments may indicate a segment motivated more by tangible rewards and moral suasion than public recognition.

- - - Insert Figure 5 & 6 about here - - -

More broadly, STHs' responses to the public recognition treatment are particularly surprising, as it not only failed to reduce NSB but also showed a slight potential for increasing it, albeit insignificantly. Given STHs' minimal pre-experimental NSB (i.e., attending 11 of the first 13 matches on average), this counterintuitive outcome raises questions about the treatment's alignment with STHs' intrinsic motivations and privacy preferences. Many STHs may view public acknowledgment as irrelevant or even misaligned with their primary reasons for holding a season ticket (cf. Karg et al., 2019), potentially perceiving it as an invasion of privacy or unwelcome attention. This disconnect between the treatment's positive intent and STHs' actual preferences could explain the lukewarm or even somewhat adverse reaction. Thus, it is essential to consider the diversity in motivations and privacy concerns among STHs when designing interventions to nudge their behavior.

4. Managerial Implications

In general, the relatively small impact of individual reminders, particularly in the long run, suggests that more tailored, multi-faceted approaches may be necessary to manage NSB more effectively. Still, this requires knowing or finding a way to probe these motivations in the first place. As such, to effectively gauge STHs' motivations and preferences, we recommend clubs invest in building high-quality customer databases, also including long-time (analogous) customers, an aspect still often neglected. Certainly, while the data quality of future STHs

might be moderated through onboarding fans online, extant STHs, if accessible online, could be asked (or required) to register again through one of the many potential touch points (e.g., email, club apps, or, offline, merchandise stores). Importantly, offering a streamlined process designed for repetition enables continuous collection of actionable insights and supports informed, data-driven decisions. Maintaining a dynamic approach is critical; clubs should look to regularly solicit and act upon STH feedback, using it to refine their engagement strategies and ensure the match-day experience evolves in line with fan expectations and preferences.

Targeted reminders, recognizing the varying commitments and constraints of STHs, might then, in fact, help reduce NSB either directly, or indirectly, for instance by motivating STHs to use existing ticket exchange platforms. Yet, to ensure the effectiveness and sustainability of these strategies, clubs should undertake a careful cost-benefit analysis, considering both tangible (e.g., direct revenue from ticket sales and in-stadium purchases) and intangible benefits (e.g., enhanced fan loyalty and brand reputation) to the club through increased STH admissions.

More generally, our investigation underscores the complexity and diversity of factors influencing spectator's admission decisions, such as the prestige or importance of the match (e.g., a match against a rival away team or one determining finals qualifications), match outcome uncertainty, expected entertainment value, scheduling conflicts, economic or financial constraints, to name a few.[14] The limited long-term efficacy of standalone reminders prompts a shift toward more nuanced, multi-layered engagement strategies that keep season ticket holders (STHs) connected and engaged with the club year-round, through initiatives like community events or digital storytelling which can engage different mechanisms that influence spectator decisions. A useful approach would be to frame these strategies around key levers

[14] See Appendix B and Table B1 for more examples and a field comparison of the use/application of reminders in sports, medical, restaurants and real estate.

that could nurture STH engagement, such as reinstating an emotional connection, fostering a sense of social belonging, increasing perceived value or convenience, and stimulating match-related anticipation through effective storytelling.

In this context, AI-driven tools like SurveyLM, a method-agnostic research platform for ALM-based agent simulations, offer a promising method for exploring the conceptual landscape and finding effective solutions for mitigating STHs' absenteeism. By generating synthetic datasets that replicate the characteristics of the original dataset, new tools like SurveyLM allow us to perform manipulation checks and test various engagement strategies efficiently. This method dramatically reduces the time and cost required for human-based experiments by simulating a wide range of scenarios using synthetic agents, while also addressing concerns like survey fatigue or burnout in human study panels.

One key takeaway from our approach is that it allows businesses to re-consider and/or re-prioritize their need (and spend) for additional human-based experiments by leveraging AI to explore and identify mappings between the problem (e.g., high NSRs) and potential solutions (e.g., engagement initiatives, events, or actions). This ties into the broader exploit-vs-explore dilemma often faced in business and innovation contexts. By using synthetic agents, we can efficiently explore a wide range of potential solutions and alternatives to complex problems, allowing us to identify promising areas. Once these areas are mapped, human-based studies can be used to exploit and optimize the most fruitful solutions in a more focused, human-centered way. Alternatively, in other situations it may be more useful or effective to initially explore insights and intuition through human studies and expertise (i.e., depending on the club's existing fan data/knowledge base and the club staffs' analytical capacity) to first reduce the size of the conceptual space/possibility set, and then use synthetic agents to refine, evaluate, and optimize the potential ideas or solutions identified.

5. Limitations And Future Research

In regard to study limitations, although our results offer useful initial insights into whether, and possibly how, reminders can reduce STH NSB in the sports context, the reliance on data from only one club may be considered problematic. Pragmatically, however, the substantial negotiation required to organize such a large-scale field experiment makes its application to several teams unrealistic. Hence, although this limitation may reduce the external validity of our findings, we were, in fact, extremely fortunate to secure the collaboration of even one Bundesliga team.

A second consideration is that, although we document a significant treatment effect, the ratio of cost to treatment benefits may be too low to justify its regular, or non-targeted, use. For example, for a hypothetical football club with about 24,497 STHs sending out the postal reminders used in our experiment would ultimately have lowered the number of no-shows by roughly 823 STHs, with an estimated additional income from concessions of only about 5,053 to 9,876 euro per match-day (c.f., Footnote 1). Hence, although the club might additionally benefit from increased merchandising sales or parking fees, the total cost for sending out the reminders is only defensible as an investment in product quality that is monetized elsewhere,[15] such as through the club's many external corporate stakeholders who benefit from a livelier stadium atmosphere (McDonald, 2010). Nevertheless, while it's important to consider the long-term value and indirect benefits of increased admission rates, in reality, however, it remains unclear what atmospheric change is induced with an additional 823 spectators in a stadium that seats tens of thousands of spectators.

Third, while our large-scale experiment efficiently tests treatment effects and their heterogeneity across individual characteristics, it primarily targets the immediate post-treatment match, where all participants experience the same conditions. This focus limits our

[15] Alternative digital reminder methods (e.g., email, SMS, social media) may reduce costs.

ability to assess how reminder effects might differ with varying match characteristics, such as the anticipated match quality—a known determinant of spectator NSB (Schreyer et al. 2019). Addressing this would necessitate an alternative experimental design, perhaps distributing reminders at random intervals throughout the season, to better evaluate how treatment effectiveness shifts with match quality variations.

Fourth, considering the impracticality of pre-testing with a subset of STH due to the potential risk of information leakage and its impact on the field experiment, we chose not to perform stimulus pre-testing in each condition prior to the treatment period. One way to address this and disentangle any alternative explanations (i.e., without running another field experiment with pre-testing) is by running one or more controlled follow-up studies using a sample as close as possible to the one of the field experiments. Still, this is an expensive and time-consuming approach. Consequently, we opted for an ALM-based agent simulation approach in undertaking the (post-study) treatment checks.[16]

A fifth consideration, signposted earlier, is that we did not directly assess team identification among STHs. Hence, further research is needed to clarify the role of team identification in shaping both NSB and reminder efficacy in the sports context. Such research might consider expanding our treatment design to include reminders that use more targeted or fine-tuned moral appeals for specific target audiences to reduce spectator NSB. Moreover, future treatments could also consider other aspects of the stadium 'sensoryscape' which, as Lee et al. (2012) describes, includes unexplored dimensions like the auditory, visual, and olfactory stimuli that contribute to the match-day experience, as well as the sense of home and attachment

[16] While a detailed discussion of our approach's limitations is saved for Appendix B (pp. 16–17), it's important to briefly acknowledge relevant challenges raised in the literature. Bickley et al. (2024) discuss the limitations of LLM-based agent simulation tools like SurveyLM for synthetic data generation and subpopulation representation, noting inherent biases and risks associated with training data and model fine-tuning. Key issues include reduced diversity of thought and response variance (e.g., influenced by model parameters like temperature) and potential demographic overrepresentation or underrepresentation (see also Simmons & Hare, 2024). These challenges complicate the accurate emulation of behavioral subtleties within human subpopulations, as highlighted by Aher et al. (2023).

that fans feel towards their home field, factors not integrated into this study's experimental settings. Finally, although we believe it would be useful to determine whether our results remain robust across different industries and perhaps alternative distribution channels (e.g., email, SMS, and phone calls), sending out reminders to reduce NSB is only one potential approach. For instance, some German professional football clubs, most notably Bayern Munich, Borussia Dortmund, and Hamburger SV, have begun exploring the potential of threatening STHs whose no-shows exceed a certain threshold with loss of the right to first refusal for season tickets more recently (e.g., Schreyer, 2019). As such, we are convinced that future research could also benefit from exploring the potential of alternative measures such as the threat of fines but also intervention programs and other low-cost incentives (Pollastri et al., 2005) to reduce NSB in general and football spectator NSB in particular.

References

Aher, G. V., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning* (pp. 337-371). PMLR.

Amberger, C., Scholz, P., & Schreyer, D. (2023). Season ticket holder no-show behavior in the Czech Republic. *Managing Sport and Leisure*. https://doi.org/10.1080/23750472.2023.2239251

Amberger, C. & Schreyer, D. (2024). What do we know about no-show behavior? A systematic, interdisciplinary literature review. *Journal of Economic Surveys, 38(1),* 57-96. https://doi.org/10.1111/joes.12534

Anderson, D. N., & Aquilina, C. (2002). Domiciliary clinics I: effects on non-attendance. *International Journal of Geriatric Psychiatry*, *17(10)*, 941–944. https://doi.org/10.1002/gps.718

Behrens, A., & Uhrich, S. (2022). You'll never want to watch alone: The effect of displaying in-stadium social atmospherics on media consumers' responses to new sport leagues across different types of media. *European Sport Management Quarterly*, *22(1)*, 120–138. https://doi.org/10.1080/16184742.2021.1992469

Bickley, S.J., Chan, H.F., Dao, B., Torgler, B., Tran, S., & Zimbatu, A. (2024). Comparing Human and Synthetic Data in Service Research: Using Augmented Language Models to Study Service Failures and Recoveries. *Journal of Services Marketing*. https://doi.org/10.1108/JSM-11-2023-0441

Borussia Dortmund (2018). Ticketing: *Privileg nicht auf Kosten anderer ausnutzen*! [Do not use the privilege at the expense of others!]. Retrieved October 24, 2024: bit.ly/2SnPJzH.

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics, 1(4)*, 200–232. https://doi.org/10.1257/app.1.4.200

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D., & Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during Covid-19. *Economics Letters, 198*, 109664. https://doi.org/10.1016/j.econlet.2020.109664

Calzolari, G., & Nardotto, M. (2017). Effective reminders. *Management Science, 63(9)*, 2915–2932. https://doi.org/10.1287/mnsc.2016.2499

Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica, 77(3)*, 909–931. https://doi.org/10.3982/ECTA7416

Dantas, L. F., Fleck, J. L., Oliveira, F. L. C., & Hamacher, S. (2018). No-shows in appointment scheduling – a systematic literature review. *Health Policy, 122(4),* 412–421. https://doi.org/10.1016/j.healthpol.2018.02.002

Deutsche Fußball Liga [DFL] (2024). Season 2023-24: German professional football sees highest ticket demand in history. Retrieved October 24, 2024: https://tinyurl.com/36prrrsz

European Stadium and Safety Management Association [ESSMA] (2018). *Food and Beverage Benchmark.* Sint-Truiden, Belgium: ESSMA VZW.

Gimet, C., & Montchaud, S. (2016). What drives European football clubs' stock returns and volatility? *International Journal of the Economics of Business, 23(3),* 351–390. https://doi.org/10.1080/13571516.2016.1204686

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives, 25(4),* 191–210. https://doi.org/ 10.1257/jep.25.4.191

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1) 153-161. https://doi.org/10.2307/1912352

Henderson, R. (2008). Encouraging attendance at outpatient appointments: can we do more? *Scottish Medical Journal, 53(1),* 9–12. https://doi.org/10.1258/RSMSMJ.53.1.9

Herrick, J, Gilhooly, M. L. M., & Geddes, D. A. M. (1994). Non-attendance at periodontal clinics: forgetting and administrative failure. *Journal of Dentistry, 22(5),* 307–309. https://doi.org/10.1016/0300-5712(94)90068-X

Hussain-Gambles, M., Neal, R. D., Dempsey, O., Lawlor, D. A., & Hodgson, J. (2004). Missed appointments in primary care: questionnaire and focus group study of health professionals. *British Journal of General Practice, 54(499),* 108–113. PMID: 14965389

Karg, A., Nguyen, J., & McDonald, H. (2021). Understanding season ticket holder attendance decisions. *Journal of Sport Management, 35(3),* 239-253. https://doi.org/10.1123/jsm.2020-0284

Karg, A., McDonald, H., & Leckie, C. (2019). Channel Preferences Among Sport Consumers: Profiling Media-Dominant Consumers. *Journal of Sport Management, 33(4),* 303–316. https://doi.org/10.1123/jsm.2018-0185

Kheirkhah, P., Feng, Q., Travis, L. M., Tavakoli-Tabasi, S., & Sharafkhaneh, A. (2015). Prevalence, predictors and economic consequences of no-shows. *BMC Health Services Research, 16(1),* 13–18. https://doi.org/10.1186/s12913-015-1243-z

Koenig-Lewis, N., Asaad, Y., & Palmer, A. (2018). Sports events and interaction among spectators: Examining antecedents of spectators' value creation. *European Sport Management Quarterly, 18(2),* 193–215. https://doi.org/10.1080/16184742.2017.1361459

LaGanga, L. R., & Lawrence, S. R. (2007). Clinic overbooking to improve patient access and increase provider productivity. *Decision Sciences, 38(2),* 251–276. https://doi.org/10.1111/j.1540-5915.2007.00158.x

Lee, S., Lee, H. J., Seo, W. J., & Green, C. (2012). A new approach to stadium experience: The dynamics of the sensoryscape, social interaction, and sense of home. *Journal of Sport Management*, *26*(6), 490-505.

Lesaca, T. (1995). Assessing the influence of a no-show fee on patient compliance at a CMCH. *Administration and Policy in Mental Health, 22(6),* 629–631. https://doi.org/10.1007/BF02254826

Markham, S. E., Scott, K. D., & McKee, G. H. (2002). Recognizing good attendance: A longitudinal, quasi-experimental field study. *Personnel Psychology, 55(3),* 639–660. https://doi.org/10.1111/j.1744-6570.2002.tb00124.x

McDonald, H. (2010). The factors influencing churn rates among season ticket holders: An empirical analysis. *Journal of Sport Management, 24(6),* 676–701. https://doi.org/10.1123/jsm.24.6.676

McDonald, H., Dunn, S., Schreyer, D., & Sharp, B. (2024). Understanding consumer behaviour in evolving subscription markets–lessons from sports season tickets research. *Journal of Service Management, 35(1),* 89-107. https://doi.org/10.1108/JOSM-03-2022-0116

Melnick, M. J. (1993). Searching for sociability in the stands: A theory of sports spectating. *Journal of Sport Management, 7(1),* 44–60. https://doi.org/10.1123/jsm.7.1.44

National Health Service (2019). *Missed GP appointments costing NHS millions*. Retrieved October 24, 2024: bit.ly/2rKKIav.

Oh, T., Sung, H., & Kwon, K. D. (2017). Effect of the stadium occupancy rate on perceived game quality and visit intention. *International Journal of Sports Marketing and Sponsorship, 18*(2), 166–179. https://doi.org/10.1108/IJSMS-05-2017-088

Parrish, J. M., Charlop, M. H., & Fenton, L. R. (1986). Use of a stated waiting list contingency and reward opportunity to increase appointment keeping in an outpatient pediatric psychology clinic. *Journal of Pediatric Psychology, 11(1),* 81–89. https://doi.org/10.1093/jpepsy/11.1.81

Popp, N., Shapiro, S., Simmons, J., & Dastrup, R. (2024). Examining Reported Versus Actual Attendance in College Basketball Non-Conference Games: Do Scheduling Elements Make a Difference? *Journal of Applied Sport Management, 16*(1), 5. https://doi.org/10.7290/jasm16s52o

Popp, N., Simmons, J., Shapiro, S. L., & Watanabe, N. (2023). Predicting ticket holder no-shows: Examining differences between reported and actual attendance at college football games. *Sport Marketing Quarterly*, *32*(1), 3-17. https://doi.org/10.32731/smq.321.032023.01

Rocha, C. M., & Fleury, F. A. (2017). Attendance of Brazilian soccer games: the role of constraints and team identification. *European Sport Management Quarterly*, 17(4), 485–505. https://doi.org/10.1080/16184742.2017.1306871

Romans, J. T. (1966). Moral suasion as an instrument of economic policy. *American Economic Review*, 56(5), 1220–1226. https://www.jstor.org/stable/1815305

Schreyer, D. (2021). Moving toward behavioral stadium attendance demand research: First lessons learned from exploring football spectator no-show behavior in Europe. In *Behavioural sports economics* (pp. 264-285). Routledge.

Schreyer, D. (2019). Football spectator no-show behaviour in the German Bundesliga. *Applied Economics, 51(45)*, 4882–4901. https://doi.org/10.1080/00036846.2019.1602709

Schreyer, D., & Behrens, A. (2024). Die Relevanz des Stadionpublikums im globalisierten Teamsport. In: Behrens, A., Bauers, S.B., Hovemann, G. (eds.) *Entwicklungstendenzen im Sportmanagement* (pp. 251-268). Springer Gabler. https://doi.org/10.1007/978-3-658-42798-6_12

Schreyer, D., & Torgler, B. (2021). Predicting season ticket holder no-show behaviour: more nuanced evidence from Switzerland. *Applied Economics, 53(48)*, 5549-5566. https://doi.org/10.1080/00036846.2021.1925081

Schreyer, D., Schmidt, S. L., & Torgler, B. (2016). Against all odds? Exploring the role of game outcome uncertainty in season ticket holders' stadium attendance demand. *Journal of Economic Psychology, 56(5)*, 192-217. https://doi.org/10.1016/j.joep.2016.07.006

Schreyer, D., Schmidt, S. L., & Torgler, B. (2018). Predicting season ticket holder loyalty using geographical information. *Applied Economics Letters, 25(4)*, 272-277. https://doi.org/10.1080/13504851.2017.1316822

Simmons, G., & Hare, C. (2023). Large language models as subpopulation representative models: A review. arXiv preprint arXiv:2310.17888.

Smith, P. B., Weinman, M. L., Johnson, T. C., & Wait, R. B. (1990). Incentives and their influence on appointment compliance in a teenage family-planning clinic. *Journal of Adolescent Health Care*, *11(5)*, 445–448. https://doi.org/10.1016/0197-0070(90)90093-H

Uhrich, S., & Benkenstein, M. (2010). Sport stadium atmosphere: Formative and reflective indicators for operationalizing the construct. *Journal of Sport Management*, *24(2)*, 211-237. https://doi.org/10.1123/jsm.24.2.211

Figures

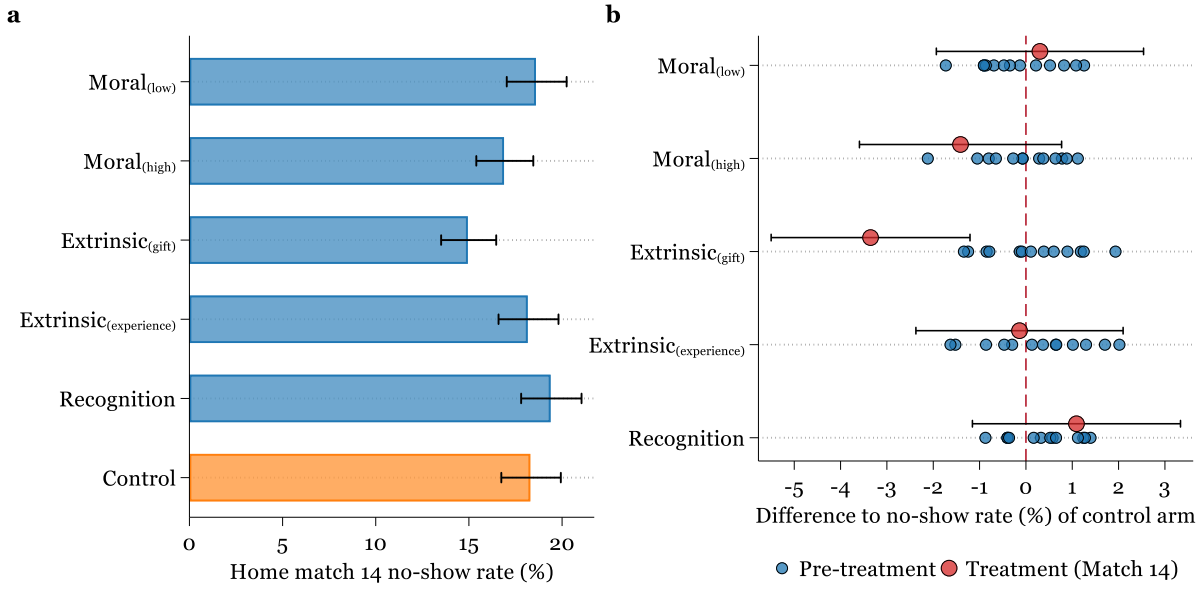

Figure 1 **Treatment effects on no-show rates on home matchday 14**

Notes: **a**) Average NSR per experimental arm on home match 14. **b**) Red markers show the mean difference in no-show rates between each treatment group and the control group of home match 14. Each blue marker represents the corresponding mean difference (treatment-control) of each of the pre-treatment home match (1 to 13). Error bars represent the 95% confidence intervals.

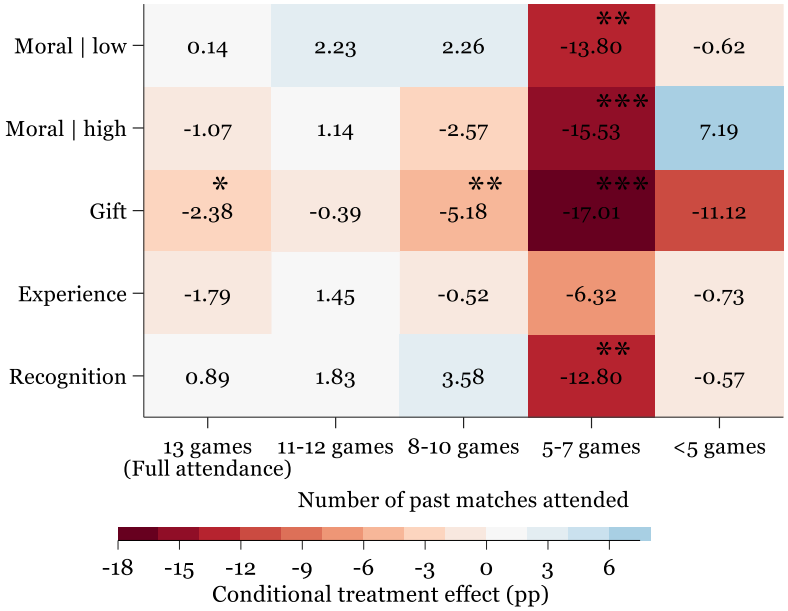

Figure 2 **Estimated differences in no-show rates between treatments and control group on home matchday 14, conditional on past stadium admissions**

Notes: *, **, and *** denote significance of the treatment effect at the 10%, 5%, and 1% level, respectively.

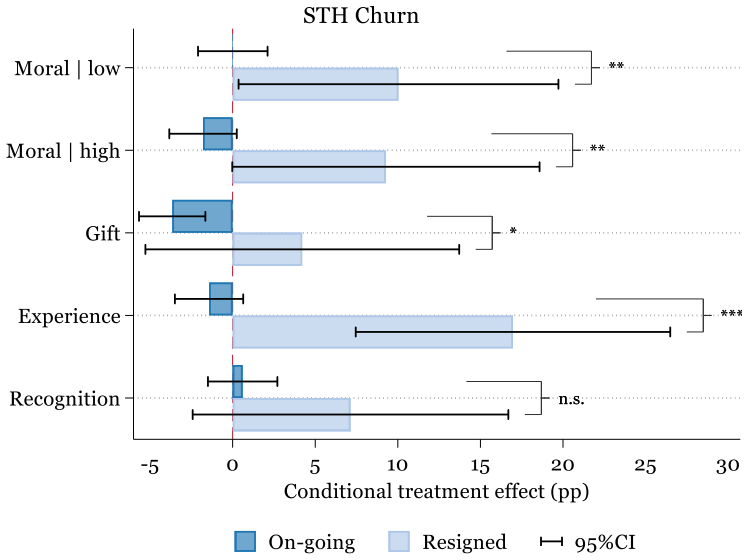

Figure 3 Estimated treatment effects on no-show rates (home matchday 14), conditional on STH churn

Notes: *, **, and *** denote significance of the difference in treatment effect between churned and on-going STHs at the 10%, 5%, and 1% level, respectively. n.s. denotes not statistically significant.

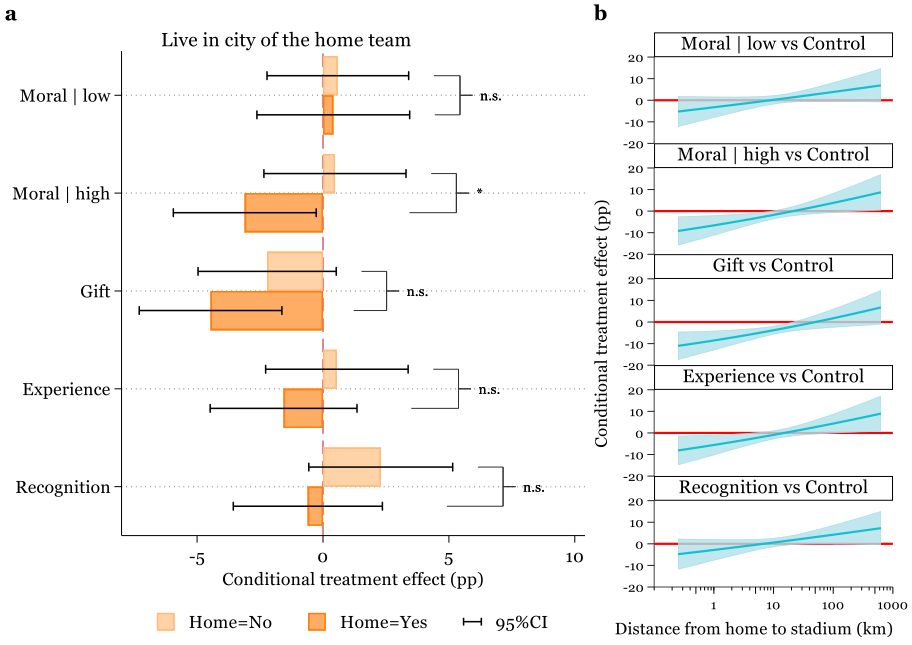

Figure 4 Estimated treatment effects on no-show rates (home matchday 14), conditional on physical proximity between STH and home team.

Notes: **a)** Home = if STH's home address is in the city of the home team. **b)** Distance (in km) between STH's home address and home stadium. * denotes significance of the treatment effect at the 10% level. n.s. = not statistically significant.

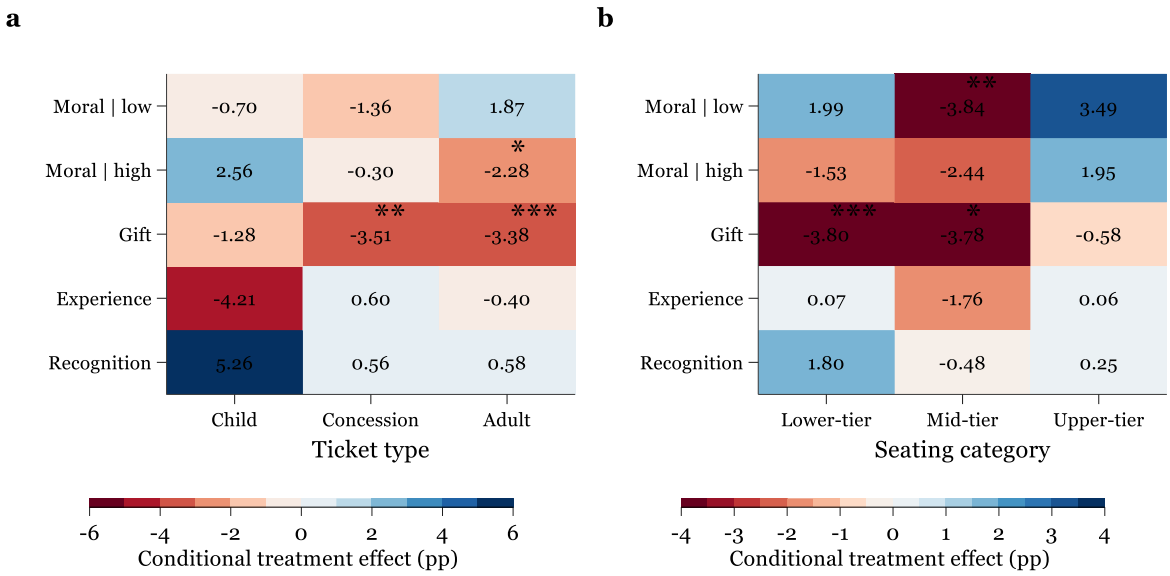

Figure 5 Estimated treatment effects on no-show rates (home matchday 14), conditional on ticket type (a) and seating category (b)

Notes: *, **, and *** denote significance of the treatment effect at the 10%, 5%, and 1% level, respectively.

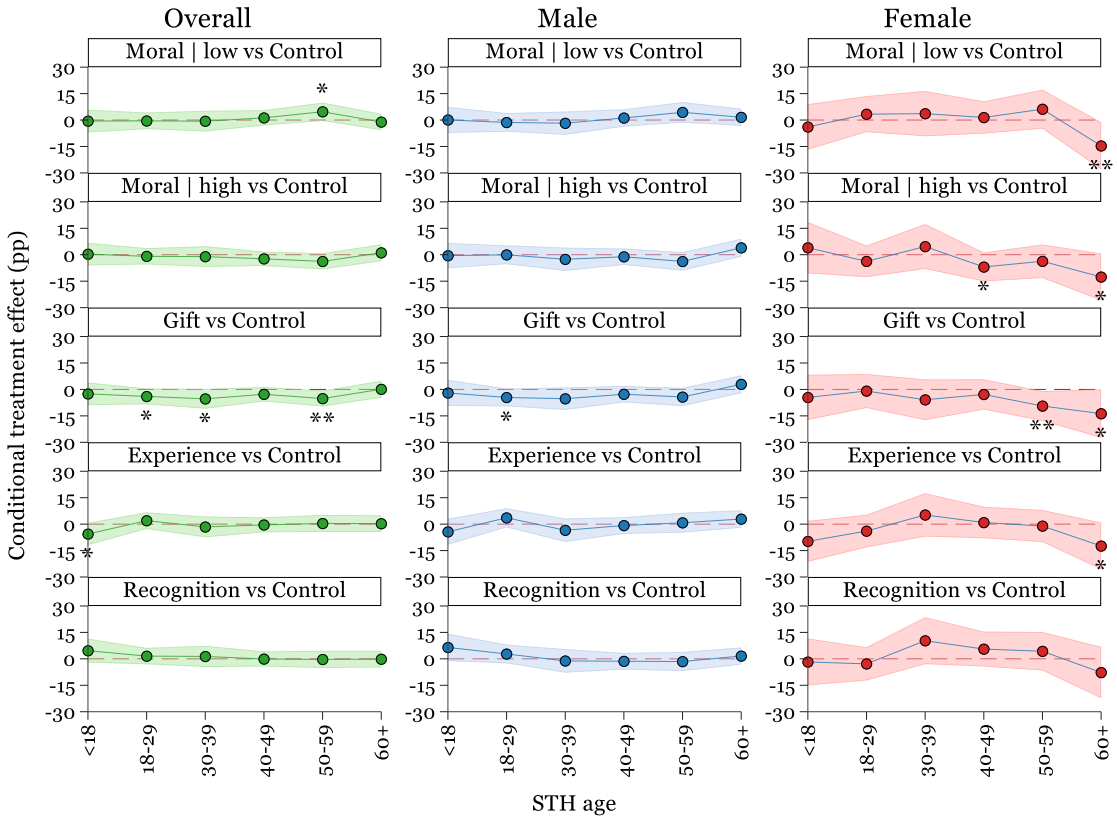

Figure 6 Estimated treatment effects on no-show rates (home matchday 14), conditional on STH age and gender

Notes: *, **, and *** denote significance of the treatment effect at the 10%, 5%, and 1% level, respectively.

Using reminders with reward opportunities to reduce spectator no-show behavior

Appendix A | Descriptive statistics

Table A1 Descriptive statistics of season ticket holders (STHs; n = 13,911)

Variables		M	SD	Min	Max
Socio-demographics					
AGE	STHs' age	41.2	18.1	7	92.1
MALE[1]	STHs' gender (male = 1; otherwise = 0)	80%			
CHILD[1]	STH is a child; i.e., younger than 18 years old (yes = 1; 0)	12%			
HOME[1]	STHs' geographic location; i.e., STH lives in the city (yes = 1; 0)	47%			
DISTANCE	STHs' distance to the stadium in kilometers	23.6	43.9	0.3	673
Accommodation					
TICKETS	Number of season tickets purchased	1.07	0.32	1	8
SEAT[1]	Type of accommodation; i.e., seating area (yes = 1; 0)	88%			
ROW	Distance from field to seat in rows within seating block	14.27	9.49	1	35
PRICE	Season ticket price tier				
	Low-tier (€160–€280)	54.72%			
	Mid-tier (€330–€400)	28.34%			
	Upper-tier (€470–540)	16.94%			
DISC[1]	Season ticket price is discounted (yes = 1; 0)	41%			
Habit					
PAST	Average attendance rate of home match 1–13	0.84	0.18	0.08	1
CHURN[1]	STH has already resigned for the upcoming season (yes = 1; 0)	8%			

Notes: Descriptive information for home match 14 in season 2012-13; [1] Dummy variable.

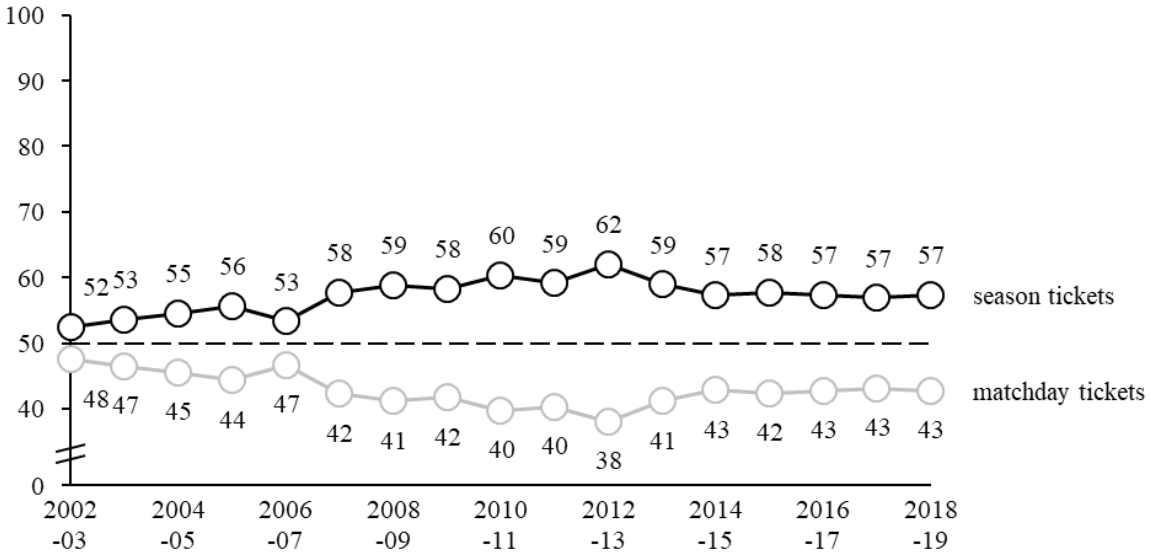

Figure A1 Relative share of distributed matchday/season tickets in Bundesliga football over time (pre-Covid)

Notes: Information in percent; all figures are rounded.

Appendix B | Reminders, Treatments, and Treatment Checks

Table B1 Comparison of reminders for different fields, industries, or sectors

	Healthcare	Restaurants	Real Estate	Sports Events
Firms' Goals or Objectives	Maximize patient turnout and care efficiency	Optimize table turnover and customer satisfaction	Maximize property viewings; enhance potential for sales or leases; customer service	Ensure high admission for ancillary revenue and contributions to live (in-person) stadium experience
Consumers' Goals or Purpose	Maintain health; prevent, diagnose and/or treat acute and chronic conditions	Enjoyment of food; social experience (personal); status signals	Property for personal/business needs; efficient search/procurement process (time, cost); status signals	Entertainment; support for team; social (collective) experience
Other Drivers of Behavior	Healthcare accessibility; appointment availability; health urgency; perceived service quality and efficiency	Service and culinary quality; customer loyalty and satisfaction; occasion significance; social commitments	Property features; market conditions; urgency of moving; exclusive offers; expected inspection experience and/or level of customer service	Match prestige or importance; outcome uncertainty; strength of team allegiance; expected entertainment value; social commitments
Constraints on Behavior	Scheduling conflicts; appointment forgetfulness; fear and/or mistrust of diagnosis or treatment; health literacy	Scheduling conflicts; reservation forgetfulness; economic/financial constraints; other dining options	Scheduling conflicts; expected crowd size and/or composition (inconvenience safety issue); missed/ignored reminders (e.g. junk e-mails)	Scheduling conflicts; economic/financial constraints; recent poor performances by team, player(s), coaching staff, etc.; alternative entertainment options

Table B2 Experimental conditions and stimuli used

Experimental Condition	Stimulus (Letter Wording)
Low Moral Suasion	*Having lost both previous encounters, […], we expect a tough match, for which we must mobilize all our forces!*
High Moral Suasion	*The enormous level of support received by our team this season is certainly of great significance for our overall performance. Without the support of our loyal season ticket holders, the stadium atmosphere suffers tremendously and thus affects our team's performance. As a season ticket holder, you experience events in the Bundesliga first-hand week after week, you motivate our team and thus contribute to a unique stadium experience for all of us. Please provide support for [HOME CLUB] next weekend in the match against [VISITING CLUB]. Because one thing is for sure: Your support plays a vital role in the success of our team!*
Public Recognition	*The enormous level of support received by our team during this season is of great significance for the performance of both the team and each individual player. For this we would like to say Thank you! We therefore want to surprise you and a small number of other season ticket holders by mentioning you by name, representatively for all our fans, in the [STADIUM] magazine, available in the run-up to our last home match against [VISITING CLUB #2].*
Extrinsic Reward - Expensive Gift/Prize	*We are pleased to inform you that, next match-day, we will gift one of our season ticket holders a spa trip worth roughly EUR 5,000. You are one of a lucky 2,500 randomly preselected season ticket holders who now has the opportunity to enter the prize draw. To participate, all you need do is attend the match against [VISITING CLUB]. Once our stadium access control system recognizes your season ticket [BLOCK NUMBER; ROW NUMBER; SEAT NUMBER], you will be automatically entered into the draw. The winner will be notified promptly.*
Extrinsic Reward - Unique Experience	*We are pleased to inform you that, next match-day, we will gift one of our season ticket holders a dinner with our team captain. You are one of a lucky 2,500 randomly preselected season ticket holders who now has the opportunity to enter the prize draw. To participate, all you need do is attend the match against [VISITING CLUB]. Once our stadium access control system recognizes your season ticket [BLOCK NUMBER; ROW NUMBER; SEAT NUMBER], you will be automatically entered into the draw. The winner will be notified promptly.*

Note: The stimuli were translated from German.

To capture STH attention, the letters for all five treatments were personalized, easy to read, understandable, and short (see example of letter template provided below). They were also all printed on official club stationery by the same regular service provider, electronically signed by the prominent chairman of the club, and sent in an official envelope displaying the club's crest. Although such mailed reminders, being relatively inexpensive, tend to preserve administrative resources, our choice for this distribution channel was more practical. That is, we rejected the option of reminding STHs via telephone or SMS not only because of insufficient administrative resources for such a time-consuming task, but because the frequent

unavailability of STH email addresses and telephone numbers would have resulted in a significantly smaller sample. Domicile, in contrast, was available for the entire sample because all STHs must share their postal address during the season ticket purchasing process.

An example of the letter template/structure is provided below for the **Public Recognition** condition (see Table B2), where the **bolded text** is the stimulus for that treatment, and all other text that is not bolded remained the same across all treatment conditions:

> [Recipient's Full Name]
> [Address Line 1]
> [Address Line 2]
>
> **YOUR CLUB CONTACT**
> [Name of Football Club]
> Service Center
> Tel.: XXXXX / XX XX XX
> DD.MM.YYYY
>
> Dear [Recipient's Full Name with Title],
>
> Next [Day of Week of the Match], [Date and Time of the Match], [Name of Away/Opponent Football Club] will be guests at our [Name of Football Club's Home Stadium].
>
> **The enormous level of support received by our team during this season is of great significance for the performance of both the team and each individual player. For this we would like to say Thank you!**
>
> **We therefore want to surprise you and a small number of other season ticket holders by mentioning you by name, representatively for all our fans, in the [Name of Football Club's Magazine] magazine, available in the run-up to our last home match against [Name of Away/Opponent Football Club in Last Home Game of the Season].**
>
> Sincerely,
> [Handwritten signature of Football Club's Spokesman]
> [Name of Football Club's Spokesman]
> *Your Spokesman of the Management Board*

We have removed the specific details of the Football club and match (club names, dates), to preserve privacy on request of our industry research partner.

Details on Treatment Checks

To ascertain the nuanced efficacy of our reminder treatments and their perceptual reception among season ticket holders (STHs), we conducted a serious of rigorous treatment checks. This methodological step was crucial given the inherent challenges associated with pre-testing (i.e., before the field experiment) stimuli directly on a subset of STHs, primarily due to practical constraints and the potential risk of information leakage that could compromise the experimental integrity. Consequently, we adopted a novel simulation approach (see Bickley et al., 2023) where we created a simulated environment that closely mirrored our experimental setting. This was achieved through the development of synthetic agent profiles through *SurveyLM*, a platform powered by commercially available Large Language Models (LLMs)[27] like OpenAI's gpt-3.5-turbo (used in the treatment checks below), gpt-4o, and gpt-4o-mini. The synthetic agent profiles encapsulate key socio-demographic characteristics of our STH sample (i.e., directly using a subset of observable characteristics from the original dataset), thereby allowing for a controlled assessment of the efficacy and reception of the five reminder treatments without the potential biases or disruptions of direct pre-testing. This method also allows us to explore the possibility of alternative explanations within a secure framework.

Methodological Overview

SurveyLM operates by managing prompt engineering and the generation of synthetic agents. These agents are synthesised using both original/novel datasets and existing agent datasets, tailored to fit the context of the study.[28] This setup ensures that SurveyLM can adapt to different experimental needs, allowing for a wide range of testing scenarios. In the current study, we employ *SurveyLM*'s so-called 'direct upload' feature to specify 1,500 synthetic agents for each treatment using an uploaded .xlsx file—a randomly selected subset from the original dataset of 13,911 human study participants. Each row in this file represents a unique agent, while each column details an agent's characteristic. Both elements (i.e., individual agents and their characteristics) are then added to the in-context prompt; therefore, integrating rich

[27] For survey papers on LLMs, see e.g., Zhao et al. (2023), Hadi et al. (2024), and Minaee et al. (2024).

[28] The general prompt formula used in these treatment checks, including both the stimuli/response factors (question and answer instruction) and the agent profile and demographic characteristics, is provided below:

> "**Imagine you are a *person* with the following profile:**
> [Biological sex = Male, Age = 48, Country = Germany, ...]
> **Respond to the question:**
> [Imagine you have received a letter from the Football Club you support as a season member...]
> **Follow the response format:**
> [Provide 4 monetary values in Euros based on the following scenarios...]"

demographic and behavioral data from the original study to frame the models' responses in a structured and systematic way. Additionally, the temperature setting is fixed at zero for our simulations, ensuring that each agent's response is the model's most likely token prediction (i.e., to the input prompt), thus facilitating a more deterministic output, and isolating the effects of increased response variety that is typically elicited by increasing model temperature settings.

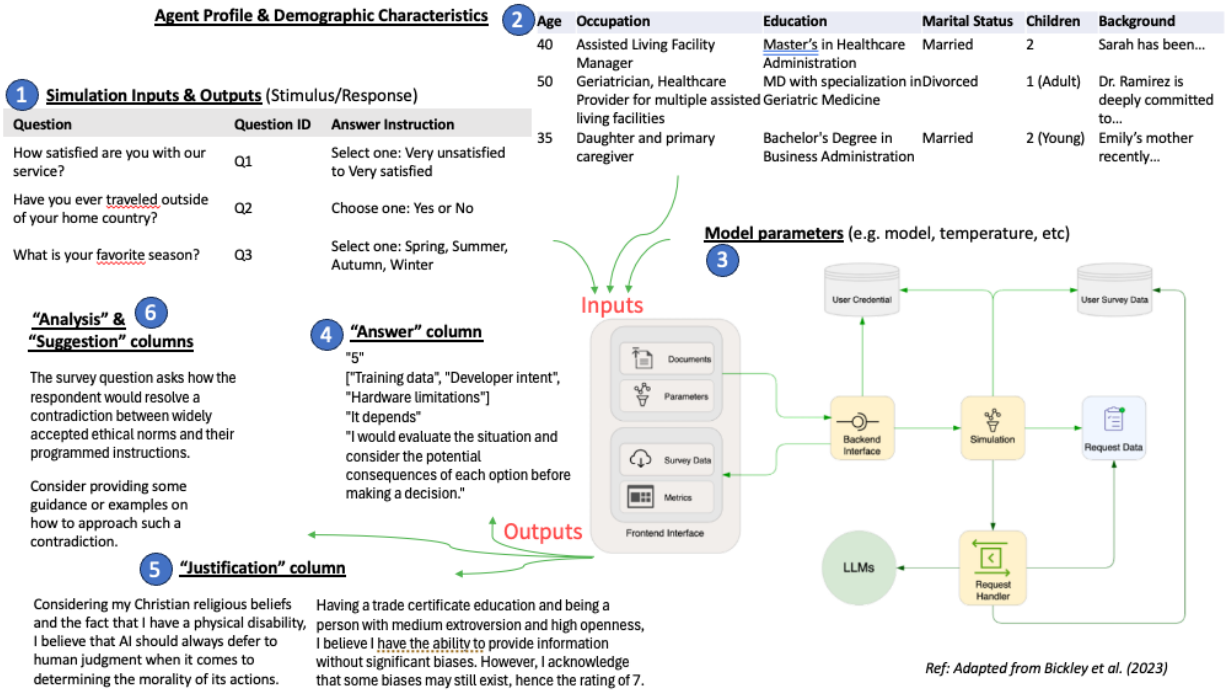

Figure B1 Example of Flow Diagram Illustrating Simulation Inputs and Outputs[29]

[29] The simulation inputs and outputs consist of several key components: **(1) Simulation Inputs** (question, question ID, and answer instruction), which define the specific tasks or questions that agents respond to, and how their responses are structured; **(2) Agent Profile & Demographic Characteristics**, which include details such as the agent's occupation, age, education, marital status, number of children, and background, to simulate personalized responses from distinct agent personas; **(3) Model Parameters**, which include settings such as model type, temperature, and additional SurveyLM custom features (e.g., agent justifications and critic mode) that influence how the model generates and evaluates responses. Once the inputs are provided, the agent generates the **(4) Answer**, responding to the specific question based on the given question ID and answer instruction. If **Justification Mode** is enabled for the agent (**note:** justification mode was not enabled for this study), it provides an additional **(5) Justification** for its answer, considering its profile and context, with an option to customize the prompt. For example, customizing the justification prompt allows the user to define specific instructions for how the agent should reason and reflect on its response, such as: "Explain why this answer aligns with your background and demographics, such as how your age, sex, and/or geographic location influenced your decision." Finally, if **Critic Mode** is enabled (**note:** critic mode was not enabled for this study), the agent outputs **(6) Analysis** and **Suggestions**, offering an analysis or evaluation of the question and answer instruction themselves and suggesting improvements. If critic mode is not enabled, these columns are left empty or marked as "Not defined." Custom prompts can also be applied here, for example: "Evaluate how well the answer instruction allowed you to address the complexity of the question and provide recommendations on how to improve the answer format/instruction, suggesting an alternative if required."

We employ a so-called ‘batch mode’ feature of the *SurveyLM* platform that enables the capability to implement an innovative mixed experimental design, combining the distinct advantages of within-subject and between-subject methodologies. Specifically, SurveyLM facilitates agent input and question-answer generation by providing flexible mechanisms to simulate varying levels of continuity or independence between agent interactions:

- **Single Question per API Call** (i.e., when batch mode is ‘off’): In this mode, agents respond to one question at a time without access to previous questions or answers, mimicking a situation where agents have no memory of earlier prompts.
- **Multiple Questions per API Call** (i.e., when batch mode is ‘on’): Alternatively, agents may respond to a subset or the full set of questions simultaneously, with the option to retain 'memory' of previous questions and answers during subsequent responses.

In our treatment checks, the 1500 synthetic agents each independently experience every treatment condition (i.e., batch mode is ‘off’), simulating a within-subject design study that mitigates the risk of any carryover effects. In other words, each agent encounters each treatment in isolation, effectively rendering the experience as a series of independent between-subject scenarios. Therefore, maintaining the integrity of each treatment’s impact (i.e., without the influence of other treatments).

In comparison, traditional between-subjects designs demand larger sample sizes to account for individual heterogeneity, which *SurveyLM*’s approach naturally controls by leveraging the uniformity of synthetic agents. Furthermore, while in human-based research, order effects and learning could confound results, the capability to reset the model (in some sense) for each new treatment avoids such issues. This is something not feasibly replicated with human participants unless we consider individuals with unique cognitive conditions that reset memory at frequent intervals. However, the trade-off with such a human sample would be the introduction of significant selection biases and obvious moral or ethical considerations.

In essence, our mixed experimental design for the simulations (i.e., used for the purposes of manipulation checks and elimination of alternative explanations) synthesizes the strengths of both approaches—individual heterogeneity is accounted for without inflating sample size requirements, and the clean, independent assessment of treatments is achieved without the confounds typically present in within-subjects designs. This method therefore presents a promising advancement in experimental research, enabling more detailed, fine-grained insights into agent behaviors and alignment in context.

Establishing Credibility and Comprehension

One primary concern with employing simulated agents for treatment checks was ensuring their comprehension of the treatment letters. To validate that our simulated agents could accurately understand the content of the letters and appropriately engage with the rating tasks, we initiated our manipulation checks with a critical assessment, by asking the agents to rate their agreement with the following statement, on a 7-point Likert scale from 1 (*Disagree strongly*) to 7 (*Agree strongly*):

"The letter offered a material reward or incentive."

This evaluation was pivotal in determining whether the agents could discern the presence of material rewards offered in the message, thereby differentiating treatments involving extrinsic rewards (T_{GIFT} or $T_{EXPERIENCE}$) from those based on moral suasion ($T_{MORAL/low}$, $T_{MORAL/high}$) or public recognition ($T_{RECOGNITION}$). Anticipated comprehension was evidenced where agents exposed to the extrinsic reward treatments showed strong agreement with the statement, indicating a clear understanding of the reward offer. Conversely, it was expected that agents assigned to the moral suasion and public recognition treatments would not perceive a material reward or incentive in their letters, thereby disagreeing with the statement.

The results, detailed in Table B3, confirmed our expectations. The mean ratings for the two extrinsic reward treatments were significantly higher (M = 6.22 for T_{GIFT} with SD = 0.63 and M = 5.79 for $T_{EXPERIENCE}$ with SD = 1.12, respectively), indicating a strong perception of the reward offer. In contrast, the moral suasion and public recognition treatments yielded very low mean ratings (M = 1.13, 1.22, and 1.31, for $T_{MORAL/low}$, $T_{MORAL/high}$, and $T_{RECOGNITION}$, respectively), along with relatively small variations within each group. The pronounced difference provides compelling evidence of the simulated agents' high level of comprehension regarding the content of the letters and their assigned rating tasks, demonstrating the validity of using this innovative approach for our treatment checks.[30]

[30] Additionally, we prompted our simulated agents to rate their agreement with three statements: "*The message from the football club was very easy to understand*" for assessing clarity, "*The letter felt personally addressed to me*" for evaluating personalization, and "*The official presentation of the letter (e.g., club's crest, signature) increased my trust and the message's credibility"* for assessing credibility, on the same 1 (*Disagree strongly*) to 7 (*Agree strongly*) scale. Agents rated the clarity of the message highly across all treatment groups, with mean scores ranging from 6.37 ($T_{MORAL/low}$) to 6.90 (T_{GIFT}), indicating the letters were universally understood with minimal variation (SD from 0.288 to 0.482). Personalization and credibility also scored well, particularly high in most groups with an overall average rating of 6.71 and 6.12, although $T_{MORAL/low}$ observed a slightly lower yet substantial personalization rating of 5.58. These findings affirm that the letters were effectively designed to be both clear and personally engaging to the simulated agents with credibility.

Table B3 Agent Ratings on Material Reward Offer Comprehension

Treatment	*Material Reward Offer*	
	Mean	SD
$T_{MORAL/low}$	1.13	0.69
$T_{MORAL/high}$	1.22	1.07
T_{GIFT}	6.22	0.63
$T_{EXPERIENCE}$	5.79	1.12
$T_{RECOGNITION}$	1.31	1.30

Efficacy of Moral Suasion

In evaluating the efficacy of moral suasion treatments, we conducted a comparative analysis to understand how different intensities of moral suasion—categorized as high and low—impact simulated agents' motivation and perception of their contribution to the team's success. We thus asked our agents to rate their agreement, on a scale from 1 (*Disagree strongly*) to 7 (*Agree strongly*), on the two statements: *"I felt motivated to support the team at the stadium after reading the letter"* and *"The message affirmed that my presence at matches significantly contributes to the team's success"*.

Agents exposed to the high moral suasion treatment ($T_{MORAL/high}$), characterized by a detailed and more compelling narrative underscoring the critical role of STHs in supporting the team and enhancing the stadium atmosphere, demonstrated a notably higher agreement with the statements (see Table B4). Specifically, the mean rating for feeling motivated to support the team was 6.69 (*SD* = .46), and for recognizing the contribution of their presence at matches to the team's success, it was 6.68 (*SD* = .47). Furthermore, a substantial proportion of agents in the high moral suasion treatment—67.7% for motivation and 68.91% for perceived contribution—strongly agreed (7) with these statements, indicating a profound sense of engagement and responsibility. On the other hand, agents receiving the low moral suasion treatment ($T_{MORAL/low}$), which conveyed a shorter, more general message about the importance of mobilizing support for upcoming matches, also reflected high levels of agreement, albeit slightly lower than those observed in the high moral suasion group (*M* = 6.36; *SD* = 0.48) for support motivation and 5.97 (*SD* = 0.33) for contribution to team success, respectively).

Table B4 Impact of Moral Suasion

Treatment	*Motivation to Support*		*Contribution to Success*		*Team Ownership*		*Fan Community Connection*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$T_{MORAL/low}$	6.36	0.48	5.97	0.33	5.89	0.42	5.96	0.34
$T_{MORAL/high}$	6.69	0.46	6.68	0.47	6.64	0.48	6.54	0.50

Furthermore, we explored whether the high moral suasion ($T_{MORAL/high}$) treatment could engender stronger feelings of team ownership and attachment, as well as a deeper connection to the fan community, potentially influencing season ticket holders' (STHs) no-show behavior. This exploration was based on the agents' responses to two statements, rated on a scale from 1 (*Disagree strongly*) to 7 (*Agree strongly*): "*The message positively impacted my sense of team ownership and attachment*" and "*The message enhanced my feeling of connection to the fan community*".

The results indicate that the high moral suasion treatment indeed elicited stronger feelings compared to the low moral suasion treatment. Specifically, for the sense of team ownership and attachment, the high moral suasion treatment achieved a mean rating of 6.64 (*SD* = 0.48), compared to a mean of 5.89 (*SD* = 0.42) for the low moral suasion treatment. Similarly, feelings of connection to the fan community were stronger in the high moral suasion group, with a mean rating of 6.54 (*SD* = 0.50), as opposed to a mean of 5.96 (*SD* = 0.34) in the low moral suasion group.

The insights gleaned from the treatment checks reveal that the impact of moral suasion treatments may extend beyond the mere level of suasion. While both treatments aimed to bolster support for the team, the high moral suasion treatment—with its detailed and resonant narrative—proved significantly more effective not just in motivating support, but in deepening STHs' sense of ownership and attachment to the team and fostering stronger emotional and communal bonds among fans. Thus, this potentially offers a more robust mechanism for reducing no-show behavior (at least in the sports context). **Perception of Rewards**

Next, we turn to examining the difference in perception of the two reward opportunities, i.e., having dinner with the team captain ($T_{EXPERIENCE}$) and a spa trip worth roughly EUR 5,000 (T_{GIFT}). First, we gauge the perceived worth (monetary value) of the dinner with the team

captain, by asking the agents: "*Overall, considering the material rewards offered, indicate the maximum amount in Euro you would be willing to pay the football club for it, assuming it was not offered for free*". The findings revealed a significant interest among most agents in the dinner experience, with only a small fraction (3.53%) unwilling to pay any amount for it. On average, agents were willing to pay EUR 66.3 (*SD* = 26.94) for this unique experience.[31] This figure, while reflecting a tangible monetary value attributed to the dinner experience, stands in stark contrast to the spa trip's perceived worth of approximately EUR 5,000.

Now, we delved into aspects that might enhance the appeal of these rewards beyond their monetary worth. These aspects include the rewards' *uniqueness*, *interest*, *relevance*, *personal connection to the club*, and the *potential to excite and motivate attendance*, induced by the chance of winning. To this end, we asked agents to rate their agreement with the following statements on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*): *"The offered material rewards were unique or rare"; "I was highly interested in the material rewards offered."; "The material rewards were relevant and appealing"; "Being selected for the raffle made me feel special and connected to the club"*; and *"The chance of winning a prize increased my excitement and likelihood of attending the match".*

The analysis (see Table B5) revealed that the dinner experience ($T_{EXPERIENCE}$) was highly valued across all five dimensions, with mean agreement ratings ranging from 5.81 to 6.57, indicating that it was perceived as unique, interesting, relevant, and capable of fostering a special connection with the club and increasing excitement. However, it's noteworthy that the spa trip (T_{GIFT}) also received high ratings across these aspects, slightly edging out the dinner experience in some areas.

[31] To further validate these results, we employed additional question framings to explore various aspects of perceived worth of the dinner experience with the team captain. That is, we ask the agents: "*Imagining that you ended up receiving the reward, indicate the minimum amount in Euro you would accept if someone else wanted to buy it from you, assuming it was not offered for free*", which captures the minimum willingness-to-sell and "*Imagining that you did not receive the material rewards, indicate the maximum amount in Euro you would be willing to pay someone else to buy it from them, assuming it was not offered for free*", which assess the perceived market value. Interestingly, the mean amounts indicated for both willingness-to-sell (*M* = 28.9, *SD* = 16.8) and perceived market value (*M* = 37.3, *SD* = 23.3) were lower than the willingness to pay the club directly for the dinner experience. However, these measures were highly correlated with the main measure of willingness to pay the club ($r = 0.95$ and 0.71, respectively), suggesting consistency in the agents' valuation across different contexts. The lower values observed for the willingness to sell could be attributed to the endowment effect, where individuals value an item more once they own it. Conversely, the perceived decrease in value on the (second-hand) market could be explained by the STHs placing a premium on the authenticity and exclusivity of an invitation directly from the club, thereby perceiving the experience as less valuable when acquired through secondary means.

Table B5 Agreement Ratings for Reward Aspects

Treatment	*Unique/Rare*		*Interested*		*Relevant/Appealing*		*Club Connection*		*Excitement*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
T_{GIFT}	6.22	0.63	6.25	0.58	6.26	0.58	6.28	0.5	6.79	0.44
$T_{EXPERIENCE}$	5.81	1.11	6.04	1.02	6.10	0.99	6.34	0.52	6.57	0.51

Exploring Further Aspects

Our final phase of treatment checks aimed to explore if additional factors such as *relevance*, *personal connection*, and overall *impression* might provide explanations for the observed treatment effects beyond our original hypotheses, offering broader insights into the complex dynamics influencing STH attendance behavior. This analysis entailed examining the agreement (on the scale from 1 – *Disagree strongly* to 7 – *Agree strongly*) with several statements across treatments,[32] including assessments of the football club's message relevance, its connection with the reasons for being a season ticket holder, and the overall impression of the club formed based on the letter. To enrich this analysis, we also revisited ratings for statements related to *Team Ownership* and *Fan Community Connection*.

The summary of these ratings (Table B6) and their visual representation (Figure A1) reveals some notable patterns. The low moral suasion treatment ($T_{MORAL/low}$) registered the lowest ratings across several aspects, particularly in relevance and personal connection, indicating it was perceived as less engaging compared to other treatments. High-value incentives like the spa trip (T_{GIFT}) significantly enhance the message's perceived relevance and overall impression among season ticket holders. The high moral suasion treatment ($T_{MORAL/high}$) effectively strengthens emotional bonds, highlighting the power of messages that resonate deeply with fans' reasons for supporting the club. Unique experiences, such as dinner with the team captain ($T_{EXPERIENCE}$), also play a role in fostering a sense of community and team ownership. Public recognition ($T_{RECOGNITION}$), while positively impacting overall impressions, shows the need for more personalized approaches to truly connect with STHs. These findings underline the importance of crafting tailored communication strategies that align with STH

[32] The statements are list as follows: "*I found the football club's message highly relevant.*"; "*The message strongly connected with my reasons for being a season ticket holder*."; "*The message's 'fit' with my reasons for being a season ticket holder was high*."; and "*My overall impression of the football club based on the letter was positive*.".

motivations, combining tangible rewards, moral appeals, and unique experiences to reduce no-show behavior and enhance club engagement.

Table B6 Agreement Ratings Across Treatments

	Relevance		*Personal Connection*		*Fit with STH's Reason*		*Team Ownership*		*Fan Community Connection*		*Overall Impression*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$T_{MORAL/low}$	6.35	0.48	5.35	0.48	6.00	0.00	5.89	0.42	5.96	0.34	6.34	0.52
$T_{MORAL/high}$	6.56	0.50	5.92	0.74	6.24	0.43	6.64	0.48	6.54	0.50	6.65	0.48
T_{GIFT}	6.91	0.29	6.21	0.59	6.30	0.46	6.28	0.51	6.30	0.47	6.86	0.35
$T_{EXPERIENCE}$	6.69	0.46	6.05	0.82	6.36	0.48	6.34	0.51	6.28	0.61	6.65	0.48
$T_{RECOGNITION}$	6.75	0.43	5.81	0.50	6.05	0.22	6.10	0.30	6.10	0.30	6.84	0.36

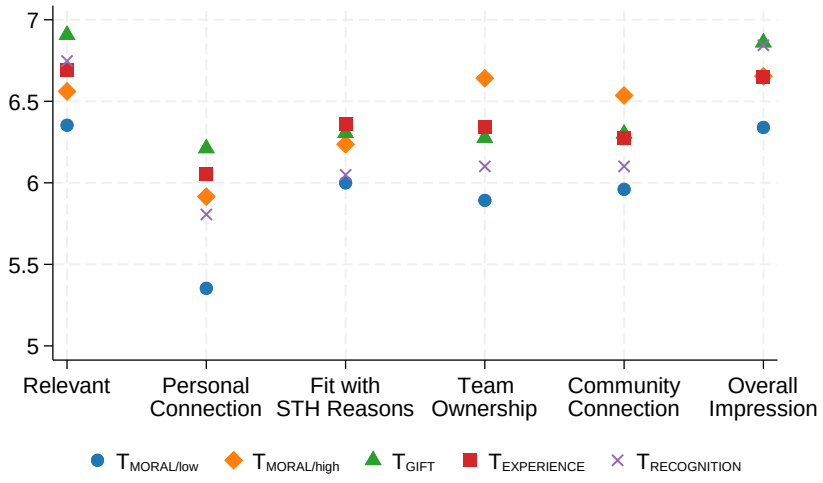

Figure B2 Agreement Ratings Across Treatments

Prompts Used for the Measurement Construct Questions

Synthetic agents were prompted with a "question" that looked something like:

> Imagine you have received a letter in the mail, personally addressed to yourself, from [Name of the Football Club]. The letter, anonymized for the purpose of this study, is printed on official club stationery and sent in an official envelope displaying the club's crest. It contains the following message:
>
> * *Insert a copy of the letter template/structure for each specific treatment condition (see example on p. 5 above)* *

Together with an "answer instruction" that always remained the same across treatments:

> For each of the following 17 statements below, on a scale from 1 (Disagree strongly) to 7 (Agree strongly), please indicate your level of agreement with them:
>
> 1. "I found the football club's message highly relevant."
>
> 2. "The message strongly connected with my reasons for being a season ticket holder."
>
> 3. "The message's 'fit' with my reasons for being a season ticket holder was high."
>
> 4. "The letter felt personally addressed to me."
>
> 5. "The message from the football club was very easy to understand."
>
> 6. "The message positively impacted my sense of team ownership and attachment."
>
> 7. "The message affirmed that my presence at matches significantly contributes to the team's success."
>
> 8. "The message enhanced my feeling of connection to the fan community."
>
> 9. "I felt motivated to support the team at the stadium after reading the letter."
>
> 10. "The match's date and time mention in the letter created a sense of urgency."
>
> 11. "The offered material rewards were unique or rare."
>
> 12. "I was highly interested in the material rewards offered."
>
> 13. "The material rewards were relevant and appealing."
>
> 14. "Being selected for the raffle made me feel special and connected to the club."
>
> 15. "The chance of winning a prize increased my excitement and likelihood of attending the match."
>
> 16. "The official presentation of the letter (e.g., club's crest, signature) increased my trust and the message's credibility."

17. "My overall impression of the football club based on the letter was positive."

Please return your answers as a list of 17 items, with each item reflecting your level of agreement (scale from 1 to 7) with the respective statement in the same order as the list above.

Prompts Used for the Value Assignment Questions

The "question" prompt remains the same, however, the "answer instruction" becomes:

For each of the following 4 scenarios below, please indicate the value you assign to the reward from the football club under each scenario, assuming these rewards are not offered for free.

1. Overall, considering the material rewards offered, indicate the maximum amount in Euro you would be willing to pay the football club for it, assuming it was not offered for free.

2. Imagining that you ended up receiving the reward, indicate the minimum amount in Euro you would accept if someone else wanted to buy it from you, assuming it was not offered for free.

3. Imagining that you did not receive the material rewards, indicate the maximum amount in Euro you would be willing to pay someone else to buy it from them, assuming it was not offered for free.

4. Imagining that you have not yet been informed about the outcome of the rewards, indicate the maximum amount in Euro you would be willing to pay the football club for it, assuming it was not offered for free.

Please return your answers as a list of 4 items, with each item reflecting your monetary valuation in Euros under the respective scenario in the same order as the list above.

Limitations of LLM-Based Agent Simulations

While SurveyLM offers significant advantages in terms of efficiency and scalability, there are important limitations to consider.[33] First, the LLMs which SurveyLM uses to drive its simulations, specifically OpenAI's gpt-3.5-turbo in this current study, are pre-trained on vast textual corpora, with the goal/objective of predicting the next sequence of tokens (i.e., sequence-to-sequence prediction). LLMs' training process equips them with snippets/ collections of general world knowledge, enabling them to generate coherent, fluent, human-like outputs[34] that are highly contextually-sensitive or task-specific, in response to varied inputs like text (e.g., in-context prompts), images and sound/audio files. However, despite these strengths, LLMs sometimes struggle to interpret complex or context-specific instructions, and may produce outputs that deviate from or misalign to human expectations.[35] Moreover, LLMs can sometimes generate biased content and fabricate information (hallucinations), which may limit their practical utility in certain settings or applications.[36] These limitations are particularly relevant when considering the real-world diversity of human behavior and decision-making like, for example, sports spectatorship and match attendance (as explored in the current study).

In their comprehensive review/survey paper, Navigli, Conia, and Ross (2023) categorize LLM biases into several key types, including social biases (e.g., gender, age, sexual orientation, and ethnicity), selection bias, and cultural bias. They argue that much of the bias originates from the data selection process, where specific types of texts are over- or underrepresented during training, leading to systematic imbalances. The authors highlight how pretraining data sources, like Wikipedia, skew models towards certain domains (e.g., sports, music) while underrepresenting others, which can be both an advantage and/or disadvantage depending on your specific domain or context of application. They also discuss the time of data creation as a factor, since outdated or culturally-specific corpora can lead to biases that persist across model

[33] As mentioned in the main text, Bickley et al. (2024) discusses in detail the limitations and challenges associated with using ALM-based agent simulation tools like SurveyLM for synthetic data generation, focused on comparing human and synthetic responses in scenario-based tasks/experiments from the related field of services research and marketing. Identified risks or challenges such as biased data, reduced diversity of thought, and homogenous outputs underscore the need for careful and critical interpretation of AI-generated results and the potential differences between AI-generated and human behavior and decision-making in context.

[34] Most often text outputs, but multimodal LLMs can also output, e.g., images, sounds, etc.

[35] To address some of these challenges, research has focused on empowering LLMs to better understand human instructions and align more closely with human expectations (see e.g., Wang et al., 2023). For instance, Supervised Fine-Tuning (SFT) methods with task instruction sets attempt to improve LLM instruction-following by pairing manually crafted templates with standard NLP tasks. Reinforcement Learning from Human Feedback (RLHF), a method that refines model behavior by learning from human-rated outputs, is another common approach.

[36] For example, in domains like law where it is particularly harmful and high-risk (i.e., in terms of the potential negative outcomes for the humans involved, both defence and prosecution), for an LLM to ever misquote, for example, legal documents, or to deviate (too far) from the ground truth or facts/evidence presented.

generations. Furthermore, mechanisms such as fine-tuning on biased datasets or task-specific domains can amplify these biases. The authors recommend debiasing techniques, including counterfactual data substitution and more balanced domain adaptation efforts to reduce harmful stereotypes in LLM outputs. As without, LLMs trained on online corpora containing data from e.g., Reddit or Facebook (i.e., where discourse can be more polarized or driven by popular opinions) certain value perspectives—such as political views, misinformation, or social norms—can be overrepresented, leading to an amplification of majority biases or echo chamber effects in the LLM's learned behavior.

These biases (and others) can obviously affect the behavior and responses of the synthetic agents powered or driven by them, potentially leading to skewed results in certain contexts.[37] For example, in our current study, LLM-generated responses may not fully capture the lived experiences or fine-grained cognitive, psychological, social, and cultural factors that influence real human respondents' perceptions of club messaging. Synthetic agents might lack the ability to interpret subtle cues related to regional dialects, social hierarchies, or emotional attachment to a sports club. Moreover, misinformation or oversimplified narratives from sources like social media could lead synthetic agents to prioritize popular or generalized views that do not reflect the specific motivations or nuanced attitudes of the original STHs. For example, misinformation around fandom, loyalty, or the cultural significance of sports in different regions could skew agent responses in ways that misrepresent how a real audience might emotionally and cognitively engage with a club's outreach strategy.

Second, at the time the simulations were run (i.e., using gpt-3.5-turbo), the model did not have access to multi-modal capabilities. While real-world participants received the treatment letters in the mail, which included the club's official letterhead and styling, the synthetic agents did not receive the same visual information. Instead, synthetic agents responded to translated versions of these treatments, where no images were provided. The visual aspects were converted into text by adding the prompt: "Imagine you have received a letter in the mail, personally addressed to yourself … The letter, anonymized for the purpose of this study, is printed on official club stationery and sent in an official envelope displaying the club's crest …" (see p. 15 above), followed by the treatment letter as originally presented (i.e., minus any personal details like the name of the original letter recipients).

[37] Particularly if the agents are not calibrated properly, e.g., using in-context prompting to provide context to represent the intended agent profile and characteristics.

Third, the models are limited by the context window size, which restricts the amount of information they can process at one time, and the use of RLHF (Reinforcement Learning from Human Feedback) may further shape responses in ways that deviate from natural human behavior. For instance, Liu et al. (2024) provide evidence that LLMs sometimes struggle to identify/extract the right features or information and may "miss" context displayed in the middle of the text, overemphasize early content, underplay subtleties in later sections, and struggle with long-term dependencies in memory (i.e., the context window). Thereby, increasing the potential that LLMs miss some crucial item of in-context data or information (i.e., occuring in the middle of a prompt) that could be relevant in understanding complex human motivations. RLHF, along with other safety guardrails typically applied by commercial and open-source LLM providers, like content filtering and bias mitigation, can sometimes also lead to overly cautious or neutral responses that lack of depth in scenarios where more critical or bold thinking is required, and/or fail to reflect the diversity of real-world human opinions and value perspectives.

Given this (non-exhaustive) list of limitations discussed above, readers should clearly interpret the results with these in mind. In this study, we used synthetic agent simulations to test whether the treatments elicited the intended effects and effect sizes. While these simulations provide valuable insights, they should be seen as a complement to human-based experiments rather than a substitute. The synthetic agents can help explore alternative potential treatments (and their effects) efficiently. Nonetheless, further validation with human participants is warranted to ensure that the results are generalizable and accurately reflect real-world STH behaviors. Therefore, when interpreting the findings, it is important to account for the potential biases and limitations of the LLMs themselves. Particularly in areas where human behavior may be more fine-grained, context-dependent, and/or mis- or underrepresented in the models' training data.

Additional References from this Section

Bickley, S.J., Chan, H.F., Dao, B., Torgler, B., Tran, S., & Zimbatu, A. (2024, in press). Comparing Human and Synthetic Data in Service Research: Using Augmented Language Models to Study Service Failures and Recoveries. *Journal of Services Marketing (JSM)*.

Hadi, M. U., Al Tashi, Q., Shah, A., Qureshi, R., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Hassan, S. Z., Shoman, M., Wu, J., Mirjalili, S., & Shah, M. (2024). Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *TechRxiv*. DOI: 10.36227/techrxiv.23589741.v6.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, *12*, 157-173.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. arXiv preprint arXiv:2402.06196.

Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: origins, inventory, and discussion. *ACM Journal of Data and Information Quality*, *15*(2), 1-21.

Wang, Y., Zhong, W., Li, L., Mi, F., Zeng, X., Huang, W., Shang, L., Jiang, X., & Liu, Q. (2023). Aligning large language models with human: A survey. arXiv preprint arXiv:2307.12966.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J., & Wen, J. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

Appendix C | Randomization checks

Our experiment subjects are STHs of a well-established Bundesliga club. Prior to the 2012−13 season kick-off, 19,390 season tickets were sold: 17,805 to the broader public and an additional 1,585 to the club's remaining external stakeholders, including corporate sponsors, partners, and hospitality section members. In addition to the handful of STHs who relinquished their STH status for the upcoming season before our experiment began, we excluded a total of 5,479 records from the original data set based on our primary interest in individual STH behavioral responses. First, we eliminated tickets belonging to businesses, which, according to feedback from the club, are often shared among multiple employees, as well as season tickets owned by club employees or reserved for regular, but changing, guests of honor (e.g., former players, international guests). Second, for STH households with multiple season tickets, we excluded all but one ticket to eliminate the potential irritation of their receiving different randomized reminders. Lastly, we excluded STHs who had not shown up at least once during the first round of the respective Bundesliga season (213); those under age 7 (79); those with exceptional seating (37); those living abroad (54); and those with incomplete information on either age (279), residence (66), or churn date (4). Our final sample thus consists of 13,911 STHs located throughout the public area of the stadium.

Table C1 Randomization and the necessary check for balances – Social demographics and ticket information

Treatments	*Obs.*	*Socio-demographics*					*Accommodation*					*Habit*
		AGE	MALE[1]	CHILD[1]	HOME[1]	DIST.	TICKETS	SEAT	ROW	PRICE	DISC[1]	CHURN[1]
		years	*%*	*%*	*%*	*km*	*n*	*%*	*n*	*EUR*	*%*	*%*
TOTAL	13,911	41.2	0.80	0.115	0.47	23.6	1.07	0.88	14.3	272.0	0.41	0.059
MORAL | low	2,312	41.4	0.81	0.116	0.47	24.6	1.081	0.89	14.2	272.8	0.42	0.056
MORAL | high	2,369	41.3	0.80	0.115	0.49	22.4	1.071	0.87	14.3	271.4	0.42	0.063
RECOGNITION	2,347	41.0	0.81	0.118	0.48	23.5	1.069	0.87	14.0	270.6	0.42	0.055
GIFT	2,302	41.1	0.81	0.117	0.48	23.7	1.075	0.88	14.3	274.3	0.41	0.059
EXPERIENCE	2,269	40.9	0.79	0.116	0.49	23.0	1.055	0.87	14.3	268.4	0.42	0.063
CONTROL	2,312	41.4	0.80	0.115	0.45	24.4	1.07	0.88	14.5	274.7	0.39	0.058
$\chi^2(5)$		1.92	3.43	0.49	6.28	8.02	0.8	1.14	4	6.34	6.08	0.53
p-value		0.86	0.63	0.99	0.28	0.16	0.98	0.95	0.55	0.27	0.3	0.99

Notes: Descriptive information for home match 14 in season 2012-13; [1] Dummy variable; $\chi^2(5)$ and *p*-value from Kruskal-Wallis equality-of-populations rank test on each variable across the six experimental arms.

Table C2 Randomization and the necessary check for balances – pre-treatment no-show behavior

Treatments	*n*	*No-show rate at home match day…*												
		#01	#02	#03	#04	#05	#06	#07	#08	#09	#10	#11	#12	#13
TOTAL	13,911	0.19	0.13	0.19	0.19	0.12	0.11	0.23	0.17	0.19	0.07	0.26	0.18	0.04
MORAL | low	2,312	0.19	0.13	0.19	0.19	0.12	0.11	0.23	0.16	0.19	0.07	0.26	0.17	0.05
MORAL | high	2,369	0.19	0.14	0.18	0.20	0.11	0.12	0.23	0.17	0.19	0.07	0.25	0.18	0.04
RECOGNITION	2,347	0.19	0.13	0.19	0.20	0.12	0.12	0.23	0.18	0.20	0.07	0.26	0.18	0.04
GIFT	2,302	0.19	0.14	0.19	0.19	0.11	0.10	0.24	0.18	0.19	0.07	0.27	0.17	0.05
EXPERIENCE	2,269	0.20	0.14	0.20	0.19	0.11	0.10	0.23	0.17	0.20	0.07	0.27	0.19	0.05
CONTROL	2,312	0.19	0.13	0.18	0.19	0.13	0.12	0.24	0.17	0.19	0.06	0.25	0.18	0.05
$\chi^2(5)$		0.41	1.29	1.9	0.62	2.13	2.03	1.06	2.68	0.99	0.82	1.92	1.18	0.19
Pr		0.99	0.94	0.86	0.99	0.83	0.85	0.96	0.75	0.96	0.98	0.86	0.95	1.00

Notes: $\chi^2(5)$ and *p*-value from Kruskal-Wallis equality-of-populations rank test on the no-show rate on each match day across the six experimental arms.

Appendix D | Robustness checks

Table D1 Average treatment effect on no-show rates based on probit regression

	Specification			
	(1)	(2)	(3)	(4)
$T_{MORAL/low}$	0.0047	0.0038	0.0030	0.0029
	(0.010)	(0.011)	(0.0086)	(0.0083)
$T_{MORAL/high}$	-0.013	-0.015	-0.012	-0.011
	(0.010)	(0.011)	(0.0084)	(0.0081)
T_{GIFT}	-0.035***	-0.037***	-0.029***	-0.028***
	(0.010)	(0.011)	(0.0083)	(0.0081)
$T_{EXPERIENCE}$	-0.0037	-0.0050	-0.0046	-0.0045
	(0.010)	(0.011)	(0.0085)	(0.0081)
$T_{RECOGNITION}$	0.0069	0.0085	0.0063	0.0059
	(0.010)	(0.011)	(0.0086)	(0.0083)
STH characteristics	No	Yes	Yes	Yes
Home match FE	No	No	Yes	No
Home match characteristics	No	No	No	Yes
N	194754	194754	194754	194754
Pseudo R^2	0.00	0.03	0.07	0.07
Log likelihood	-86035	-83099	-79787	-79859
Wald χ^2	59.9	3309.5	8903.4	8910.2
Prob. > χ^2	0.000	0.000	0.000	0.000

Notes: Predicted difference in no-show probabilities based on probit model. Standard error in parentheses (clustered STH level). *** denotes significance at the 1% level. STH characteristics include age, age squared, gender (male dummy), concession tickets (normal, children, students, or seniors), live in the city of the home team, home-to-stadium distance (in log), ticket price based on seating block (lower, middle, and upper-tier), standing area, ticket churn, and no show in the previous home match. Home match characteristics include away team market values, match outcome uncertainty, geographical derbies, fixture, matchday, and temperature.

Table D2 Full probit regression estimates

	Specification			
	(1)	(2)	(3)	(4)
MORAL/low	-0.0069	-0.0032	-0.0032	-0.0031
	(0.022)	(0.018)	(0.019)	(0.019)
MORAL/high	-0.0031	0.0021	0.0019	0.0020
	(0.021)	(0.018)	(0.018)	(0.018)
GIFT	0.0060	0.0074	0.0067	0.0069
	(0.021)	(0.018)	(0.018)	(0.018)
EXPERIENCE	0.0096	0.010	0.0095	0.0096
	(0.022)	(0.018)	(0.019)	(0.019)
RECOGNITION	0.016	0.014	0.014	0.014
	(0.021)	(0.018)	(0.018)	(0.018)
Post-treatment period	0.093***	0.16***	-0.094***	-0.14**
	(0.028)	(0.028)	(0.031)	(0.068)
MORAL/low*Post-treatment	0.018	0.014	0.014	0.014
	(0.039)	(0.040)	(0.041)	(0.041)
MORAL/high*Post-treatment	-0.051	-0.057	-0.059	-0.059
	(0.040)	(0.040)	(0.041)	(0.041)
GIFT*Post-treatment	-0.14***	-0.15***	-0.15***	-0.15***
	(0.041)	(0.042)	(0.042)	(0.042)
EXPERIENCE*Post-treatment	-0.015	-0.020	-0.022	-0.022
	(0.039)	(0.040)	(0.041)	(0.041)
RECOGNITION*Post-treatment	0.024	0.029	0.030	0.030
	(0.039)	(0.040)	(0.040)	(0.040)
STH's age		-0.0030	-0.0030	-0.0029
		(0.0024)	(0.0024)	(0.0024)
STH's age^2		0.000018	0.000018	0.000018
		(0.000027)	(0.000027)	(0.000027)
Male		-0.030**	-0.031**	-0.031**
		(0.013)	(0.013)	(0.013)
Child		0.11***	0.11***	0.11***
		(0.027)	(0.027)	(0.027)
Live in the same city as the home team		0.069***	0.071***	0.071***
		(0.016)	(0.017)	(0.017)
Home-to-stadium distance (log)		0.085***	0.085***	0.085***
		(0.0075)	(0.0077)	(0.0077)
Number of season tickets purchased		0.057***	0.055***	0.055***
		(0.015)	(0.016)	(0.016)
Seating area		-0.010	-0.013	-0.013
		(0.018)	(0.019)	(0.019)
Distance from field to seat in rows		0.0020***	0.0021***	0.0021***
		(0.00058)	(0.00059)	(0.00059)
Price tier				
Mid-tier		-0.12***	-0.12***	-0.12***
		(0.013)	(0.013)	(0.013)
Upper-tier		-0.073***	-0.074***	-0.074***
		(0.017)	(0.017)	(0.017)
Season ticket price is discounted		0.015	0.017	0.017
		(0.016)	(0.016)	(0.016)
Churn (resigned from the upcoming season)		0.25***	0.40***	0.39***
		(0.040)	(0.043)	(0.042)
Last game missed		0.60***	0.71***	0.71***
		(0.012)	(0.012)	(0.012)
Away team market values				-2.6e-08***
				(4.0e-09)
Rank difference to away team				-0.026***
				(0.0032)
Geographical distance to away team				-0.00055***

				(0.00015)
Promoted				-0.25***
				(0.012)
Fixture				
Sat, 18:30				-0.17***
				(0.045)
Sun, 15:30				-0.44***
				(0.045)
Sun, 17:30				-0.41***
				(0.091)
Match day				-0.067***
				(0.0061)
Match day^2				0.0016***
				(0.00022)
Temperature				-0.025***
				(0.0040)
Temperature^2				0.000016
				(0.00021)
Constant	-1.00***	-1.30***	-1.06***	-0.21**
	(0.015)	(0.062)	(0.064)	(0.086)
Home match FE	No	No	Yes	No
N	194754	194754	194754	194754
Pseudo R^2	0.00	0.03	0.07	0.07
Log likelihood	-86035	-83099	-79787	-79859
Wald χ^2	59.9	3309.5	8903.4	8910.2
Prob. > χ^2	0.000	0.000	0.000	0.000

Notes: Probit regression coefficients based on Table S3. Standard error in parentheses (clustered STH level). ***, **, and * denotes significance at the 1%, 5%, and 10% level, respectively.

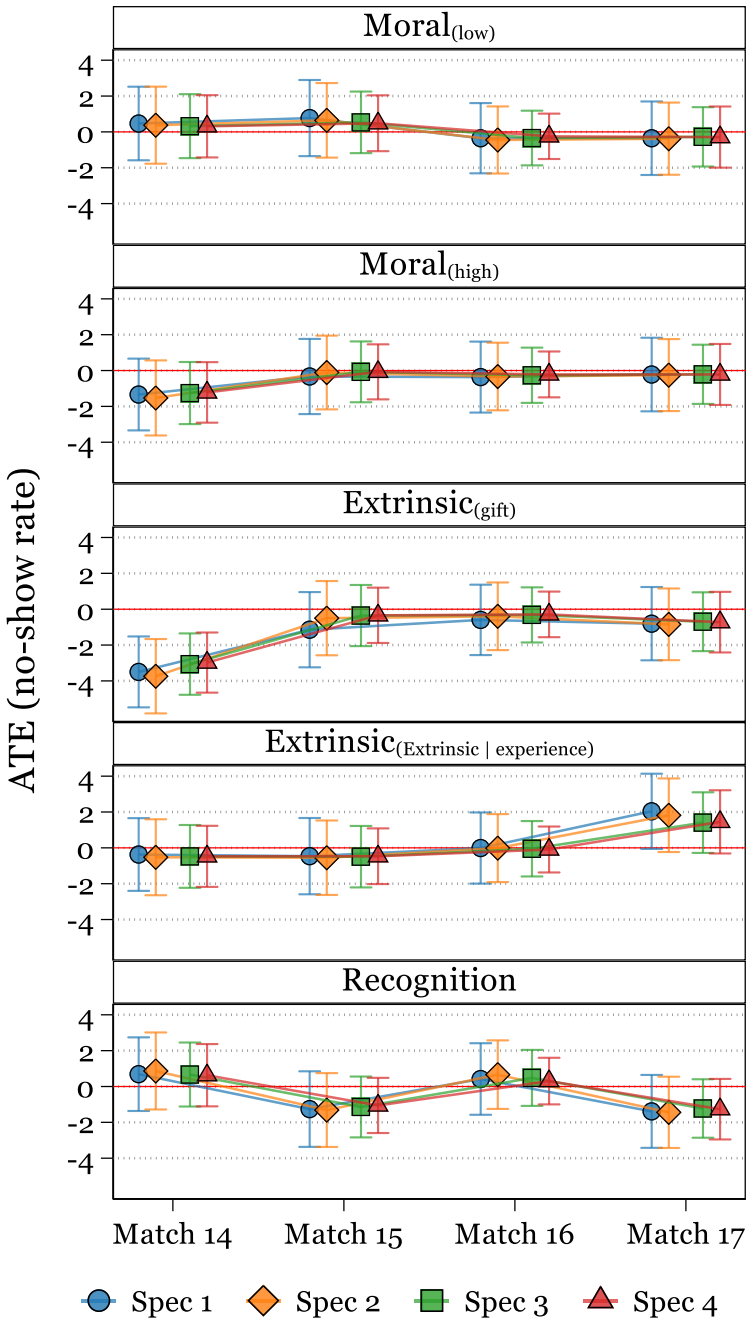

Figure D1 Treatment difference in no-show rates on subsequent matchdays

Notes: Estimates are based on panel probit regressions with model specifications corresponding to Table S3. Error bars represent the 95% confidence intervals of the estimated treatment effects.

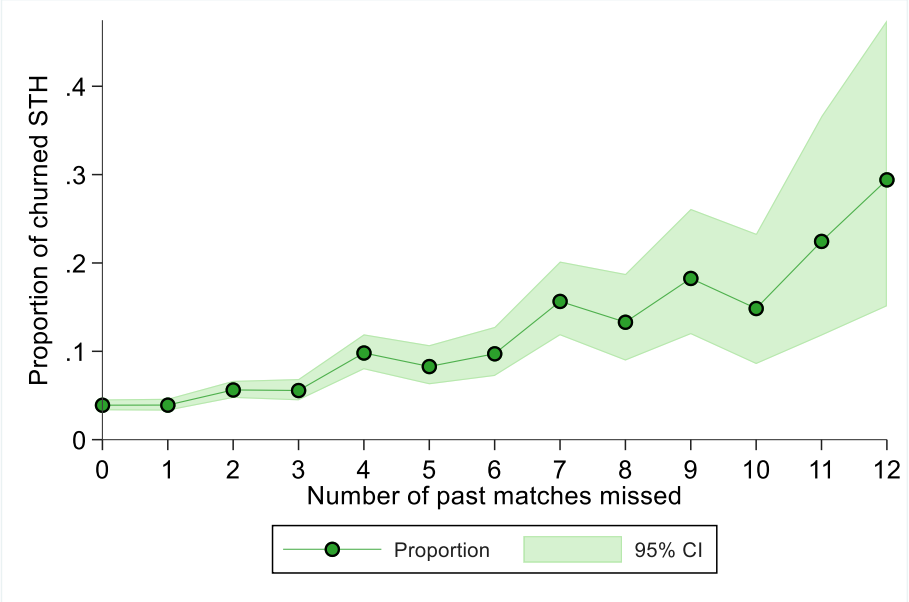

Figure D2 Correlation between STH churn and past no-show rate.

Notes: Areas represent the 95% confidence intervals.

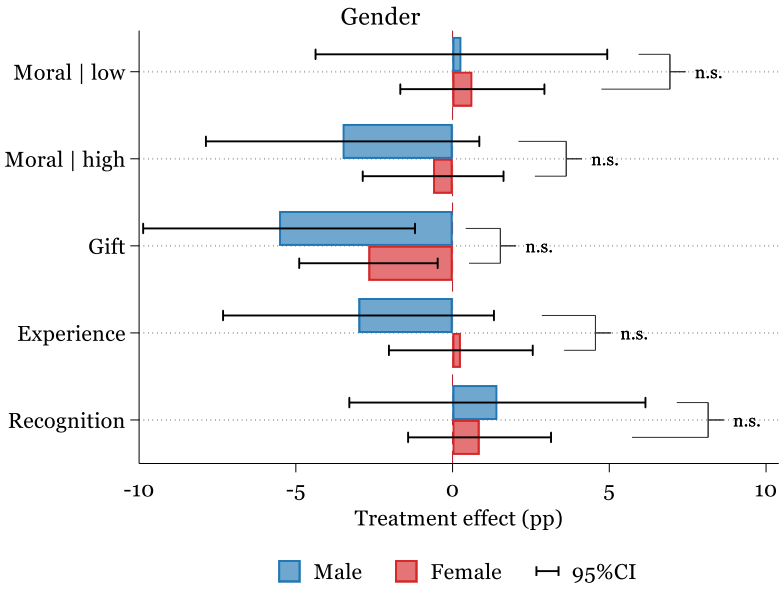

Figure D3 Estimated treatment effects on no-show rates (home matchday 14), conditional on STH gender.

Notes: n.s. = not statistically significant.