The Impact of Behavioral Design and Users' Choice on Smartphone App Usage and Willingness to Pay: A Framed Field Experiment

Christina Timko ()
Maja Adena (WZB Berlin, TU Berlin)

Discussion Paper No. 559

January 03, 2026

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, Ludwig Maximilian University of Munich, 80539 Munich, Germany
info@rationality-and-competition.de

The Impact of Behavioral Design and Users' Choice on Smartphone App Usage and Willingness to Pay: A Framed Field Experiment

Christina Timko and Maja Adena (WZB Berlin & TU Berlin) *

Abstract: Behavioral design in smartphone apps aims at inducing certain, monetizable behavior, mainly increased engagement, measurable by usage time. Such design is rarely transparent and often restricts users' ability to make alternative choices. In a framed field experiment, we document that behavioral design doubles app usage time compared to a version without behavioral elements. Providing users with choices—simply explained and conveniently adjustable design features—reduces usage time and increases their willingness to pay for the app. These findings suggest that offering choice could pave the way for new business models based on more responsible app design.

Keywords: smartphone app, behavioral control, filtering algorithm, transparency and choice, self-determination, corporate social responsibility, field experiment.

JEL codes: C93, O33, D83, L86, M14

* Christina Timko (dr.chris.timko@gmail.com) Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum, Germany, until 2023. Maja Adena (maja.adena@wzb.eu), WZB Berlin Social Science Center, Reichpietschufer 50, 10785 Berlin, Germany. CT thanks her partners in the GOAL project for their valuable comments (in alphabetical order): Yannik Borutta, Anja Folberth, Matthias Haag, Marc Hauer, Johannes Kevekordes, Carsten Orwat, Catharina Rudschies, Ingrid Schneider, and Katharina Zweig. CT especially thanks Kastanie Eins for the app development and Anne-Kathrine Kjær Christensen at Specifii in Denmark for her behavioral design workshop and her input on the user profile, concepts, and design ideas. The authors thank their students and testers for valuable feedback. Special thanks to Nicholas Ostrode and Malte Niederstadt for supporting the literature review, the preparation of the analysis work, and the conduction of the experiment and Katharina Dorn for supporting the analysis. CT thanks Michael Roos for entrusting her with this project. The data was collected via the MaXLab of the Otto von Guericke University in Magdeburg, Germany in November 2021. Data and programs are available at: [will be filled in upon acceptance]. The study was funded in line with the GOAL (Governance of and by Algorithms) project by the German Federal Ministry of Education and Research (BMBF) from 01/09/2019 to 31/12/2021, Funding ID: 01IS19020 (CT). Support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) and FOR 5675 (project number 518302089), and by Leibniz through P146/2022 and K595/2024 is gratefully acknowledged (MA).

1. INTRODUCTION

Smartphone apps have become a central part of daily life, generating significant revenue and consuming substantial user time. In 2023, over 6 million apps generated $550–900 billion USD in revenue, with an annual growth rate of 19.5%. Users spend an average of 4.8 hours daily on apps, primarily social and communication tools.[1] Most app business models achieve high conversion rates and viral growth relying on behavioral design (Alter, 2017; Lambrecht et al. 2014). Behavioral design is a design framework for programming (i.e., intentionally and systematically changing) human behavior by modifying the physical and digital environment (Combs & Brown, 2018). It can take various forms, including app appearance, usability, rewards, social features, progress monitoring, black-box algorithms, and personalization. It includes behavioral design elements from persuasive technology, nudging, behavioral interventions, and even aggressive and manipulative design (Michie et al., 2013; Oinas-Kukkonen & Harjumaa, 2008; Thaler, 2018).[2]

App providers commonly apply behavioral design to influence consumption decisions, nudge users to share personal data that can be monetized, or manipulate them into taking certain actions (Luguri & Strahilevitz, 2021). They personalize content and advertising to sell products (Boerman et al., 2017), make price discrimination possible (Acquisti et al., 2016), assist decision making (Kleinberg et al., 2018), nudge and influence users (Dellarocas, 2006), bind their attention and loyalty (Claussen et al., 2013), and create complex, much-tested tools with hidden business purposes (Montag et al., 2019). Putting user engagement but not user benefits at the heart of the design, behavioral design can exploit cognitive biases and induce harmful, risk-prone, or addictive behavior (Mosquera et al., 2020). It can induce users to share more data than they otherwise would

[1] https://web.archive.org/web/20241226101832/https://mindsea.com/app-stats/

[2] See also the overarching systematics that we created for this study in Table C1 in Appendix C

be willing to do (Acquisti et al., 2015). Moreover, one risk in digital media is digital addiction, where traditional metrics may overestimate consumer utility of social media apps (Allcott et al., 2020). Recent research provides evidence that social media are habit-forming and their excessive use is driven by self-control problems, with up to 31% of social media use attributable to such issues (Allcott et.al., 2022). Further studies find that news content personalization reduces knowledge (Beam, 2014) and contributes to the creation of echo chambers (Bail et al., 2018).

Behavioral design becomes problematic when it deliberately disregards user preferences, compromises freedom of choice, and undermines individual autonomy and the basic right of self-determination (Susser et al. 2018). In other words, if it lacks free and informed choice. Such designs are termed sludge, deceptive design, or dark patterns (Mathur et al. 2019; Thaler, 2018). Since users are not informed about, not asked to consent to, or unable to avoid such practices, they are addressed by new legislation that strengthens consumer protection, such as the newly enacted Digital Services Act (DSA) by the European Parliament (European Commission, 2022), which requires more transparency and user-adjustability in apps. Digital nudges, as reviewed by Bergram et al. (2022), while aiming to guide people's choices, they also allow users to retain autonomy by providing options and provide an avenue towards in-app self-regulation. Boosts, on the other hand, foster people's competence to make their own choices—that is, to exercise their own agency (Hertwig & Grüne-Yanoff, 2017).[3]

Business goals achieved through behavioral design in apps seem to contradict responsible design principles. Thus, our research question examines whether app design can reconcile business objectives with corporate responsibility. In our experiment, we explore an app design that includes typical behavioral design elements—aimed at fulfilling primary business goals—while also

[3] See also Fogg (2002) who pioneered a discussion on requirements for responsible behavioral design. For more details, see Table B1 in Appendix B.

providing users with choices—clearly explained and easily adjustable features—to satisfy corporate responsibility aims. We compare this new *Choice* design to a *Baseline* version with minimal functional design and a version with full *Behavioral Design*.

A related study by Mariotti et al. (2022) explores the optimal design of information nudges to influence present-biased decision-making under uncertainty. Their model predicts that targeted nudges, such as traffic-light labels, guide consumers by making the consequences of their choices more immediate and tangible. The authors emphasize the potential benefits of reduced transparency in managing impulsive behavior. In the digital context of our study, we evaluate the need for convenient tools to help users self-regulate their potentially misguided behavior and explore how much value they would place on these tools.

Our study aims to show the implications of in-app user choice, filling critical gaps in the literature. First, while previous research predominantly focuses on the effects of traditional behavioral nudges, our investigation specifically examines the holistic effects of behavioral design systematically aligned to meet the goal of increasing engagement (usage time). Second, we evaluate the potential financial implications for app providers by measuring hypothetical willingness to pay (WTP), an aspect that has received limited attention in existing studies on digital nudging. WTP reflects users' valuation based on their app experience, including aspects of empowerment and convenience that depend on the design. Additionally, our study empirically tests the effects of user choice in a practical setting—a framed field experiment with a news app designed by the authors. This innovative method not only contributes to the understanding of how customizability influences user behavior but also aids in identifying potential pathways for responsible app design and enhances the discourse around responsible app provision.

The experimental details are as follows: 141 student participants had access to the experimental news app for two weeks. There were three experimental conditions: *Baseline*, an app version featuring a minimal functional design; *Behavioral Design*, incorporating behavioral design elements; and *Choice*, which included the same behavioral design elements with additional explanatory information and adjustable design features. We observed participants' usage behavior and collected additional information through a post-study survey. We found that *Behavioral Design* resulted in an increase in usage time by 0.51 of standard deviation compared to the *Baseline* and that the usage time in *Choice* was in-between the two other treatments. We also found that the WTP for the app was lowest in the *Behavioral Design* treatment. Thus, we delivered first evidence that behavioral design combined with choice has a potential to improve business practices and enhance users' agency. However, due to the limitations of our study, those findings need further validation.

While our study's two-week timeframe limits our ability to capture digital addiction, the lower WTP observed in the *Behavioral Design* treatment may suggest that users are becoming aware of the addictive aspects associated with this design. Additionally, we acknowledge that the welfare effects of increased usage time in the context of reading news, as studied here, are not entirely clear.[4] Other limitations include the modest sample size and the use of student participants. Nonetheless, we demonstrate how dramatically consumption patterns of the same news content can change when presented in different app versions, as well as how various choice features can further modify these patterns.

[4] For example, while increased news consumption can enhance information access, concerns remain regarding exposure to misinformation, the reinforcement of echo chambers, reduced attention spans, and potential increases in digital addiction.

2. EXPERIMENTAL DESIGN

The study app was built by a professional app developer based on our concept. Specifically, the app was a news app, because they are commonly used, news feeds are an easy way to fill a study app with rich real-world and up-to-date content, and readers worldwide predominantly obtain their news content online (Flaxman et al., 2016). News came from 15 different German language sources, comprising the major news providers. However, our study does not specifically focus on news content. Importantly, while our study app is 'neutral'—not monetizing usage time (no ads) and not aiming to influence individual word perceptions (no fake or biased news)—its tempting behavioral design and the choice version could equally well be adapted to other environments, making our results generalizable. For more details on the experimental design and its mechanisms and risks see Table A1 and for essential pre-studies Tables B2 and B3 in the Appendix.

Three app versions were designed, one for each treatment group: (1) *Baseline* with a minimal functional design, but no additional behavioral design elements aiming to raise user engagement and no choice features. (2) *Behavioral Design* with behavioral design elements intending to prolong usage time but no choice features. (3) *Choice* with the same elements as those in the *Behavioral Design* version but including upfront information on pros and cons of behavioral design and offering adjustable design features. Those features included several separate options to monitor, adjust, or deactivate news filtering, and to adjust or deactivate push notifications. See Table 1, and for more details of the app versions and screenshots Table A3 and C2 in the Appendix.

We recruited 141 participants from the student pool of the Laboratory for Economic Research at Otto von Guericke University in Magdeburg for a 14-day online study conducted in November 2021. Table 2 presents participant demographics. The study was conducted anonymously. Participants were informed that the aim of the study was to learn about their user experience with the app. All participants signed an informed consent declaration.[5] We randomly assigned the participants to three treatment groups, stratifying on gender.[6] The post-study survey was filled in completely by 136 participants, with 45-46 participants in each group, which forms our analysis sample.

Participants who completed all parts of the study received 50 euros. The app usage time was not incentivized, ensuring the natural observation of individual choices regarding time spent with the app. Beyond the direct financial incentives, the app offered free access to daily news content that is either not available without subscription or only in limited quantities. We communicated with participants via the lab and on every fourth day sent reminder emails to those who had not opened the app for the previous three days.

Our primary outcome variable is usage time, measuring user engagement and reflecting the common monetization goal of app providers. We also measured non-incentivized willingness to pay (WTP) as an indicator of users' experience of app usage. In the post-study survey, we asked each participant to state the amount of money they would be willing to pay for a comparable market-ready app in four most common business models that rely on direct payments: individual (i.e. not shared) monthly subscription, monthly subscription shared with family and friends, donation (pay-as-you-want), and paying per article (pay-as-you-use).

[5] While our university had no ethics committee at the time, we were pointed to the precedential decision on ethical approvals by the German Association for Experimental Economic Research (GfeW) which states that by German federal law and by the ethics code of the German National Science Foundation (Deutsche Forschungsgemeinschaft DFG), human subject experiments are exempt from the IRB review as long as standard experimental protocols are used. Throughout the app development and experiment design, we adhered to ethical standards and data protection requirements, also obtaining the necessary permissions from the news providers.

[6] Table 2 confirms that our randomization achieved a good balance between treatment groups. None of the 18 p-values from the test of proportions is significant at conventional levels thus confirming a good balance. However, given a small sample, some important differences might go undetected. Therefore, in the robustness checks, we include controls for those individual characteristics.

We expect the app usage time to be higher in the Behavioral Design group compared to the Baseline and Choice groups. Without choice, behavioral design exploits behavioral biases by serving users' immediate pleasures through gamification and social comparison with other readers, or by keeping them trapped in echo chambers. The usage time in the Choice group is expected to lie between that of the other two groups as it allows for self-determined active choices. This leads us to:

Hypothesis 1: *Usage time: Baseline < Choice < Behavioral Design*

We expect WTP for the app to be higher in the Choice group, compared to the Behavioral Design group. The educational boosts and adjustable design features offer a larger choice set that is more likely to meet user preferences, and thus should increase users' utility on average. In addition, there is also value added to the presentation of the choice set per se, which should lead users to make better informed choices, increasing user experience, and thus their valuation. Moreover, transparency has already been shown to increase WTP in an online privacy context (e.g. Tsai et al. 2011). This leads us to:

Hypothesis 2: *Willingness to pay: Behavioral Design < Choice*[7]

3. RESULTS

Over the 14 days of the experiment, 7,839 articles were available in the news feed, of which 2,301 (29%) were read by at least one participant. We classified an article as read if the participant opened the article for at least five seconds. The average app usage time was 6.6 minutes per day. This compares to eleven minutes spent daily by a representative German person on reading print newspapers and to three minutes spent daily on reading newspaper content online in 2021, the year of the study. However, the usage time differs by group, see Table 3. Participants in the *Baseline*

[7] We have no prior on the WTP for *Baseline* in comparison to other groups.

group spent an average of 67 minutes over the two weeks. This corresponds to slightly less than five minutes per day. In the *Behavioral Design* group, the average usage time was 132 minutes, which is almost ten minutes per day. In the *Choice* group, it was 88 minutes, which is over six minutes per day. The time spent on reading articles alone was 39, 86, and 54 minutes in the respective treatments, which is between close to three minutes per day in the *Baseline* group to more than six minutes per day in the *Behavioral Design* group. Thus, the observed usage time in the *Behavioral Design* group is double that of the *Baseline* group.

To reduce the number of hypothesis tests, we combined the measures of total time usage and time spent reading articles using principal component analysis. The resulting standardized variable facilitates the interpretation of effect sizes in the regressions. Table 4, Column I, presents results of an OLS regression controlling for the stratification variable (gender).[8] It suggests a 0.17 of standard deviation higher usage time in the *Choice* treatment than in the *Baseline*, significant at $p<0.1$ (two-sided test). *Behavioral design* increases usage time by as much as 0.51 of standard deviation ($p<0.05$). Both differences turn insignificant once we account for multiple hypothesis testing (MHT) using the Romano-Wolf method (Clarke 2021) but, on the other hand, randomization p-value for Westfall-Young multiple testing of treatment significance (Young 2019), rejects the hypothesis of no overall treatment effects. When comparing *Choice* and *Behavioral Design*, a Wald test in Column I yields insignificant differences between those treatments. Applying a nonparametric Wilcoxon rank-sum (Mann-Whitney) test to pairwise comparisons rejects equality between *Choice* and *Baseline* ($p=0.079$), *Baseline* and *Behavioral Design* ($p=0.032$) but not between *Behavioral Design* and *Choice* ($p=0.686$). Overall, while the

[8] Table A4 in the Appendix shows that the coefficients remain stable when including further control variables.

time use averages are ordered in line with our hypothesis 1, the differences between the *Behavioral Design* and *Choice* groups are not statistically different.

Since the averages may mask the dynamics of the usage time, we also shortly report usage patterns over time. In Figure 1, we present selected outcome variables by treatment group and over time. The graphs for additional variables can be found in Figure A1 in Appendix A. The first graph in Figure 1 shows the development of the article reading time over the study period for the three groups separately. We see that, at the beginning of the study, the reading time per day in the *Choice* group was more similar to the *Behavioral Design* group than to the *Baseline* group and from day seven on it was the other way around. We observe a more negative trend in usage over time in the *Choice* group than in the two other groups.[9]

To better understand user engagement in the *Choice* group, we briefly report on the actual usage of choice features in the *Choice* group—the only treatment group with such features (see also Figure A3 in the Appendix). We first note that the majority of the participants (66%) changed the offered filtering settings (e.g. deactivation, displaying popular articles first, etc.) including the adjustment of the filtering algorithm's outcome: 30% did so once and 36% did so multiple times. Overall, the participants adjusted filtering options throughout the study, including on the last day. Altogether, 74% of the participants adjusted at least one of the default settings. The high usage rate of choice features, when offered so, confirms their value for users, and we might observe a learning or habituation effect over time.

Next, we asked the participants for their WTP for an individual monthly subscription, shared subscription, the amount they would be willing to donate, or how much they would be

[9] While our treatment does not allow us to distinguish between the effects of transparency and choice, the observed pattern suggests a stronger impact of choice—individuals appeared to learn over time—whereas transparency, presented upfront, seems to have had no effect.

willing to pay per article. Table 3 below shows the averages of the respective outcome variables. The WTP is lowest in the *Behavioral Design* across all measures. Once again, to streamline hypothesis testing, we combined all measures into a single variable using principal component analysis. The resulting standardized variable simplifies the interpretation of effect sizes in the regressions presented in Table 4, Column II. The *Choice* and *Baseline* groups have similar WTP, while the WTP in the *Behavioral Design* is significantly lower. While accounting for the MHT does not reject the similarity of treatments, the Westfall-Young multiple testing procedure does reject the hypothesis of no overall treatment effect. Comparing the *Choice* and *Behavioral Design* groups, a Wald test reveals a significant difference between the treatments ($p < 0.05$). Additionally, a nonparametric Wilcoxon rank-sum (Mann-Whitney) test for pairwise comparisons rejects the equality between the *Behavioral Design* and *Choice* groups ($p = 0.047$), confirming our Hypothesis 2. The same is true for the comparison between the *Baseline* and *Behavioral Design* groups ($p = 0.045$).

4. DISCUSSION AND CONCLUSIONS

Odysseus' story about avoiding the Sirens' behavioral control over his free choice illustrates the problem of tempting apps and the solution in form of prior awareness raising and, based on that, informed decision-making and self-regulation, as it is offered in the choice version of the study app. Nevertheless, the choice version does not kill pleasure, since it is convenient to use and still offers the entire range of behavioral design elements. Just like Odysseus still enjoyed listening to the Sirens while having taken care to protect himself.

Our two-week-long field experiment with a news app provides causal evidence that behavioral design significantly increases app use and lowers WTP compared to the baseline with a minimal functional design. Offering consumers choice in the version with explanatory information and adjustable design features leads to ‘intermediate’ usage time and significantly higher WTP than in the version with behavioral design. Specifically, we measured engagement through usage time and verified that behavioral design doubles average usage time compared to the baseline, aligning with Luguri’s & Strahilevitz’s (2021) findings on behavioural design effects in real-world apps. Choice empowers users to make better decisions about how much to engage with a tempting app. On average, users value having alternative choices in apps more than having a full behavioral design. Thus, this study contributes to the understanding of the time-increasing effect of concerted behavioral design in apps and the valuation of transparency and choice measures (Cemiloglu et al., 2023; European Commission, 2022; Mariotti et al. 2022). However, these findings should be interpreted with caution due to the limitations of this study, including a small student sample size and a limited scope of measurements. Further research is necessary to replicate and expand upon these results.

The welfare implications of reduced app usage in the context of a news app remain ambiguous. While increased reading time can benefit users when the news is personally relevant and meaningful (Schrøder, 2019), it poses risks if users are merely driven to compete with others, read too quickly without adequate attention, or become trapped in echo chambers. Moreover, real apps may pose additional risks, such as digital addition or spread of fake news. There is also the potential misuse of behavioral data collected during usage, all which we could not replicate in our study for ethical reasons. In this regard, the trustworthy academic integrity of our study represents a limitation, and the measured WTP may be inflated.

Regarding external validity of our results, we believe that they are generalizable to any app, because behavioral design that intends to increase usage time and succeeds even for a rather serious and common everyday activity like news reading, is also likely to succeed in other serious or more playful, mundane or extraordinary environments too. Whether behavioral design elements are "beneficial" or "dark", depends on the targeted user behavior and the intention of the designer (Thaler, 2018). Whether the design is tempting, will depend on what individual users fall for. Our study merely showcases what systematically concerted behavioural design and a range of choice features can achieve, and what effect size they can have, independently from any business model or further context. Thus, responsible design does not necessarily implement making any normative decisions on behalf of the users.

By the design and use of our authentic study app—a pioneering research method—we opened up the way for future complex app studies in controlled environments. For example, future studies might investigate more details of the underlying mechanisms of single behavioral design elements or social network effects. For this to succeed, a large user pool is required, either by an own testbed of a commercially run study app or jointly with providers of real apps.

Our findings have important implications for app providers and policymakers. From a corporate social responsibility (CSR) perspective, offering users transparent and adjustable design features could represent a shift toward more ethical app development practices. Such practices not only enhance user satisfaction but may also foster trust and long-term customer loyalty. For practitioners in the field of digital media, our practical contribution is also to show an example and provide a first documented, comprehensible, feasible, and reproducible proof of concept for the design of a responsible app with regard to its behavioral design and its risks. The monitoring and correction of algorithmic news filtering results even provides an example for transparent interaction with algorithms, which is deemed a valuable insight for software engineers, as we derive from their feedback in interviews. The reproducibility of our design may be applied in any app, scaling the extent of our practical impact. Therefore, our research insights have the potential to lead to groundbreaking future trends.

The managerial implications of our research are important too. The changing regulatory conditions for app design open up new possibilities for new business models and market positioning, as well as new industry standards, for which we provided a practical tool, verified by methods of behavioral economics.

Table 1: Three app versions

BASELINE

No behavioral design that raises user engagement

- In-app consent to privacy policy and terms & conditions, optional media competency quiz
- **Minimal functional design:**
    - Article rating and display of average article ratings
    - Display of number of readers
    - Same grey design for all news categories in the news feed
- No additional behavioral design elements to raise user engagement
- No choice features

BEHAVIORAL DESIGN

Full behavioral design that raises user engagement

- In-app consent to privacy policy and terms & conditions, optional media competency quiz
- **Full behavioral design** that would reasonably be added in a real app:
    - Personalized algorithmic news filtering based on news categories
    - Choice architecture: distinct colors marking news categories and highlighted news
    - Gamified elements: achievement badges, rankings, progress history
    - Tailored push notifications on news updates, rewards, and ranking
- No choice features

CHOICE

Adjustable design features & information on chances & risks

- In-app consent to privacy policy and terms & conditions, optional media competency quiz
- **In-app consent to behavioral design** (with information on chances & risks)
- Same behavioral design elements as the Behavioral Design version
- **Adjustable design features** with short local explanations:
    - Monitor, adjust, or deactivate news filtering (several options available)
    - Adjust or deactivate push notifications

Table 2: Balancing table

	Baseline (A)	*Behavioral Design (B)*	*Choice (C)*	Test of proportions p-value		
		Share		A=B	A=C	B=C
Female	0.533	0.556	0.522	0.832	0.912	0.746
Native speaker	0.956	1.000	0.978	0.153	0.544	0.320
With no reading problems (impaired vision, reading difficulties)	0.622	0.578	0.457	0.667	0.113	0.247
Aged 18-25	0.556	0.511	0.500	0.673	0.596	0.916
With a BA degree	0.356	0.378	0.500	0.827	0.164	0.240
With a MA degree	0.244	0.244	0.130	1.000	0.163	0.163
Computer science major	0.133	0.111	0.109	0.748	0.718	0.971
N	45	45	46			

Table 3: Means of the usage time and willingness to pay (WTP) variables

			Usage time		WTP			
	N		Total usage time	Reading time	Monthly subscription	Monthly shared subscription	Donation	Per article
Baseline	45	Mean	66.784	39.709	5.167	9.856	5.633	0.867
		Std. error	(8.855)	(5.546)	(0.616)	(1.110)	(0.982)	(0.442)
Behavioral Design	45	Mean	132.009	85.677	3.811	7.233	4.478	0.328
		Std. error	(30.308)	(20.638)	(0.480)	(0.898)	(0.853)	(0.075)
Choice	46	Mean	87.526	54.472	4.793	9.554	6.054	0.835
		Std. error	(9.553)	(6.637)	(0.462)	(1.029)	(0.981)	(0.339)
Baseline= Behavioral		Two-sided t-test p-value	0.044	0.036	0.086	0.070	0.377	0.236
Baseline= Choice		Two-sided t-test p-value	0.167	0.156	0.144	0.093	0.228	0.150
Behavioral =Choice		Two-sided t-test p-value	0.115	0.091	0.629	0.843	0.762	0.955

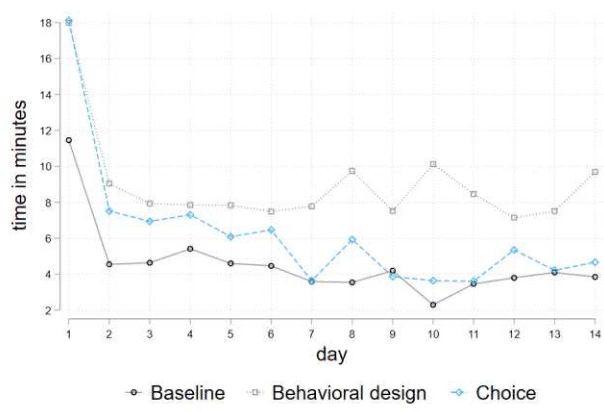
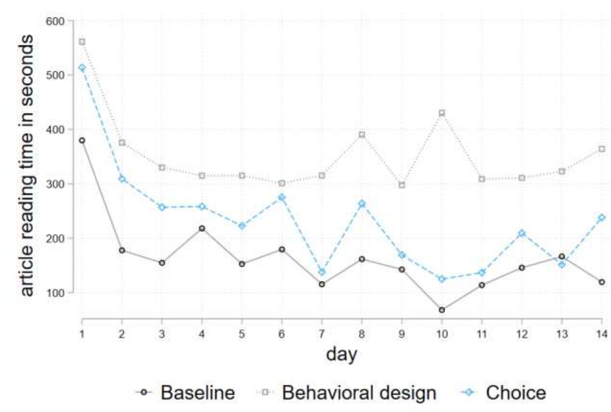
Notes: Total usage time is measured as time between opening and closing the app. Reading time is the time spent on reading articles—the article is classified as read if it is opened for at least five seconds. We also cap the maximum reading time at approximately 3-4 minutes, which is the time that an average reader would need to read the text. The WTP variables were measured in the survey at the end of the experiment. The questions read (translation): (1) *How much should a monthly subscription for a comparable market-ready newsfeed reader app cost?* (2) *How much should a monthly subscription for a comparable market-ready newsfeed reader app cost if it can be shared with family and friends?* (3) *How much should a monthly donation be for the provider of a comparable, free of cost, market-ready newsfeed reader app?* The choice was made on a slider with values up to 50 euros. The last question was: (4) *How much should one article read in a comparable, market-ready newsfeed reader app cost?* followed by a write-in answer in euros.

Table 4: The effects of treatments on usage time and willingness to pay (WTP)

	Usage	WTP	Column I+II
	I	II	III
Choice	0.172 (0.081) [0.227]	-0.029 (0.898) [0.902]	
Behavioral Design	0.510 (0.036) [0.227]	-0.418 (0.055) [0.227]	
Randomization p-value for Westfall-Young multiple testing of treatment significance	0.024[a]	0.076[a]	0.047[b]
Wald test p-value for *Choice=Behavioral Design*	0.162	0.022	
Observations	136	136	
R2	0.058	0.089	

Notes: OLS regression with strata fixed effects (gender); Outcome variables: To reduce the number of hypothesis tests, we combined the measures of time (total usage time and time spent reading articles) and the measures of WTP (monthly subscription, monthly shared subscription, donation, per article) into one dependent variable each using principal component analysis. The resulting variables are standardized, which facilitates the interpretation as effect sizes; p-values in parentheses are derived from robust standard errors; square brackets contain Romano-Wolf adjusted p-values for multiple testing (two regressions and two treatments in each). Westfall-Young p-value from: [a] the test that there is no effect of any treatment on the outcome, and [b] the test of the null of complete irrelevance (no treatment had no effect on any outcome).

Figure 1: Dynamics of average usage time per study day and by treatment groups

A: Total usage time in minutes	B: Article reading time in seconds
	

5. REFERENCES

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509–514. doi.org/10.1126/SCIENCE.AAA1465

Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. Journal of Economic Literature, 54(2), 442–492. doi.org/10.1257/jel.54.2.442

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. American Economic Review, 110(3), 629–676. doi.org/10.1257/aer.20190658

Allcott, H., Gentzkow, M., & Song, L. (2022). Digital addiction. American Economic Review, 112(7), 2424-2463. 10.1257/aer.20210867

Alter, A. (2017). Irresistible: The rise of addictive technology and the business of keeping us hooked. Penguin.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Fallin Hunzaker, M. B., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences of the United States of America, 115(37), 9216–9221. doi.org/10.1073/pnas.1804840115

Beam, M. A. (2014). Automating the news: How personalized news recommender system design choices impact news reception. Communication Research, 41(8), 1019–1041. doi.org/10.1177/0093650213497979

Bergram, K., Djokovic, M., Bezençon, V., & Holzer, A. (2022). The digital landscape of nudging: A systematic literature review of empirical research on digital nudges. CHI Conference on Human Factors in Computing Systems (CHI '22). doi.org/10.1145/3491102.3517638

Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online Behavioral Advertising: A Literature Review and Research Agenda. Journal of Advertising, 46(3), 363–376. doi.org/10.1080/00913367.2017.1339368

Cemiloglu, D., Arden-Close, E., Hodge, S. E., & Ali, R. (2023). Explainable persuasion for interactive design: The case of online gambling. Journal of Systems and Software, 195, 111517. doi.org/10.1016/j.jss.2022.111517

Clarke, D., 2021. rwolf2 Implementation and Flexible Syntax. http://www.damianclarke.net/computation/rwolf2.pdf

Claussen, J., Kretschmer, T., & Mayrhofer, P. (2013). The Effects of Rewarding User Engagement: The Case of Facebook Apps. Information Systems Research, 24(1), 186–200. doi.org/10.1287/isre.1120.0467

Combs, T. D., & Brown, R. A. (2018). Digital behavioral design. In Venice Beach: Boundless Mind.

Dellarocas, C. (2006) Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms. Management Science 52(10):1577-1593. doi.org/10.1287/mnsc.1060.0567

European Commission, Directorate-General Consumers, Lupiáñez-Villanueva, F., Boluda, A., Bogliacino, F., Liva, G., Lechardoy, L., & de las Heras Ballell, T. (2022). Behavioural study on unfair commercial practices in the digital environment : dark patterns and manipulative personalisation : final report. Publications Office of the European Union. doi.org/doi/10.2838/859030

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. Public Opinion Quarterly, 80(S1), 298–320. doi.org/10.1093/poq/nfw006

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human Decisions and Machine Predictions. The Quarterly Journal of Economics, 133(1), 237–293. doi.org/10.1093/QJE/QJX032

Lambrecht, A., Goldfarb, A., Bonatti, A., Ghose, A., Goldstein, D. G., Lewis, R., Rao, A., Sahni, N., & Yao, S. (2014). How do firms make money selling digital goods online? Marketing Letters, 25(3), 331–341. doi.org/10.1007/s11002-014-9310-5

Luguri, J., & Strahilevitz, L. J. (2021). Shining a Light on Dark Patterns. Journal of Legal Analysis, 13(1), 43–109. doi.org/10.1093/jla/laaa006

Mariotti, T., Schweizer, N., Szech, N., & von Wangenheim, J. (2022). Information Nudges and Self-Control. Management Science, 69(4), 2182-2197. doi.org/10.1287/mnsc.2022.4428

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark Patterns at Scale. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–32. doi.org/10.1145/3359183

Michie S., Richardson M., Johnston M., Abraham C., Francis J., Hardeman W., Eccles M. P., Cane J., and Wood, C. E. (2013), The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions, Annals of behavioral medicine, 46 (1), 81-95. doi: 10.1007/s12160-013-9486-6

Montag, C., Lachmann, B., Herrlich, M., & Zweig, K. (2019). Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of

Psychological and Economic Theories. International Journal of Environmental Research and Public Health, 16(14), 2612. doi.org/10.3390/IJERPH16142612

Mosquera, R., Odunowo, M., McNamara, T., Guo, X., & Petrie, R. (2020). The economic effects of Facebook. Experimental Economics, 23(2), 575–602. doi.org/10.1007/s10683-019-09625-y

Oinas-Kukkonen, H., & Harjumaa, M. (2008). A systematic framework for designing and evaluating persuasive systems. In: Oinas-Kukkonen, H., Hasle, P., Harjumaa, M., Segerståhl, K., Øhrstrøm, P. (eds) Persuasive Technology. PERSUASIVE 2008. Lecture Notes in Computer Science, vol 5033. doi.org/10.1007/978-3-540-68504-3_15

Schrøder, K. (2019). What do news readers really want to read about? How relevance works for news audiences. Digital News Project. Reuters Institute for the Study of Journalism. doi.org/10.60625/risj-n12y-az27

Susser, Daniel, Roessler, Beate and Nissenbaum, Helen F., (2019). Online Manipulation: Hidden Influences in a Digital World. 4 Georgetown Law Technology Review 1 dx.doi.org/10.2139/ssrn.3306006

Thaler, R. H. (2018). Nudge, not sludge. Science 361, 431-431. doi.org/10.1126/science.aau9241

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. Information systems research, 22(2), 254-268. doi.org/10.1287/isre.1090.0260

Young, Alwyn (2019). Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. Quarterly Journal of Economics 134 (2): 557-598. doi.org/10.1093/qje/qjy029