Inducing behavior change with the help of information systems

Cross-domain applications, mechanisms, and real-world impact

Sebastian A. Günther

Bamberg, 2024

URN: urn:nbn:de:bvb:473-irb-935808
DOI: https://doi.org/10.20378/irb-93580

Diese Arbeit hat der Fakultät Wirtschaftsinformatik und Angewandte Informatik der Otto-Friedrich-Universität als Dissertation vorgelegen.

Erstgutachter: Prof. Dr. Thorsten Staake

Zweitgutachter: Prof. Dr. Tim Weitzel

Mitglied der Promotionskommission: Prof. Dr. Barbara Drechsel

Tag der Disputation: 02.02.2024

ORCID iD des Autors: 0000-0002-8005-7546

Dedicated to my parents,

Johannes and Petra

Acknowledgments

This dissertation is not the result of an isolated researcher, but the result of a fascinating journey that I probably could not have taken on my own. I want to express my sincere thanks to my supervisors, colleagues, friends, and family for supporting me all the way.

First, I would like to express my deepest gratitude to Prof. Dr. Thorsten Staake for having given me the opportunity to conduct this type of research. His constant guidance, support, and feedback made it possible to pursue a dissertation that studies intriguing research questions with real-world data. I am particularly grateful for the open discussions that led us to many inspiring research ideas, some of which will still materialize in the near future. Furthermore, I would like to thank Prof. Dr. Tim Weitzel for his invaluable advice on my publication efforts and for joining my PhD committee. I am also thankful to Prof. Dr. Barbara Drechsel, the third member of my PhD committee, for her valuable comments on my research plans.

Moreover, I want to thank Prof. Dr. Verena Tiefenbeck for her scientific guidance over the past years. During this time, she has become both an excellent partner for conducting field experiments and an important source of inspiration for my research. Likewise, I am thankful for the support and advice of Dr. Vlad C. Coroamă, who has substantially helped me on this journey. Special thanks go to Dr. Matthias Hauser for encouraging me, at the end of my master's, to pursue research at the doctoral level. His advice had a substantial influence on the research orientation of my dissertation. I am as well thankful for the early publications together, which greatly helped me to strengthen my subsequent academic work.

I would also like to thank my former and current colleagues Andreas Weigert, Dr. Anna Kupfer-Weinert, Carlo Stingl, Christian Lorenz, Christian Weigl, Felix Haag, Dr. Jürgen Wenig, Dr. Konstantin Hopf, Dr. Leonard Michels, Samuel Schöb, and Sarah Betz. I have had many great moments with all of them. Similarly, I would like to thank some friends that I made during this time: I am thankful to Dr. Markus Schindewolf and Stefanie Pischel, with whom I shared plenty of amazing experiences in Bamberg.

Most importantly, I owe immense gratitude to my family for their constant support and patience over the last few years. Special thanks go to Johannes and Petra, who have always believed in me and encouraged me, even in difficult times. Lastly, I would like to thank Alina for her unconditional support on this journey, as well as her steadfast belief in me. Without her, this endeavor would clearly not have been possible.

Sebastian A. Günther
(February 2024)

Zusammenfassung (German summary)

Die Bereitstellung von Feedback ist eines der gängigsten Mittel zur Verhaltensbeeinflussung, das in zahlreichen Bereichen des Lebens Anwendung findet (z. B. Fry und Neff, 2009; Karlin et al., 2015; Van der Kleij et al., 2015). Feedback, was die wissenschaftliche Literatur üblicherweise als Prozess definiert, in dem einem Individuum Informationen über sein Verhalten gegeben werden, um die nachfolgenden Aktionen zu beeinflussen, hilft Individuen, indem es bspw. die Aufmerksamkeit auf Handlungen lenkt, Fehlvorstellungen über Ursache-Wirkungsbeziehungen korrigiert und somit Feedback-Empfänger beim Erlernen und Anwenden von effektiven Strategien zur Verhaltensregulation unterstützt (Hattie und Timperley, 2007; Karlin et al., 2015; Kluger und DeNisi, 1996). Eine zentrale Voraussetzung zur Generierung von Feedback sind Verhaltensdaten, die in der Vergangenheit teilweise manuell erfasst werden mussten, heutzutage jedoch durch die Verbreitung der Informationstechnologie (IT) zunehmend automatisiert und hochaufgelöst in quantitativer Form zur Verfügung stehen. In Anbetracht dieser neu gewonnenen Datenquelle lassen sich die Effekte von Feedback-Interventionen auf viele Alltagshandlungen mit einer außergewöhnlich hohen externen Validität bewerten und dabei das Verständnis über die Wirkmechanismen von Feedback grundlegend erweitern. Das Verständnis ist einerseits erforderlich, um die gängigen Theorien zu den Wirkungsweisen von Feedback weiterzuentwickeln, und andererseits nützlich für politische Entscheidungsträger und Organisationen, die bereits großflächig auf Interventionen wie Informationskampagnen oder monetäre Anreize setzen (z. B. Bandiera et al., 2007; Brambila-Macias et al., 2011; Sierzchula et al., 2014) mit dem Ziel, wichtige Kennzahlen unserer Gesellschaft (z. B. Klimaschutz, Lebenserwartung, Produktivität) durch Verhaltensänderungen im Alltag zu verbessern.

Trotz der hohen Relevanz von IT-basierten Feedback-Interventionen für die Erforschung von Wirkmechanismen bestehen noch einige offene Fragen, die beantwortet werden müssen, um das Potenzial von Feedback nachhaltig für politische Entscheidungsträger und Organisationen zu heben. So wird in der Forschungsliteratur oftmals Feedback mit anderen Interventionen wie Geldanreizen oder moralischen Appellen kombiniert, ohne die Einzelwirkung der verwendeten Interventionen herauszuarbeiten. Dadurch ist unklar, welche Wechselwirkung zwischen den Interventionen und Feedback vorherrscht (cf. Fang et al., 2023; Liebe et al., 2018), was die Theorieentwicklung über die Auswirkungen von Feedback erschwert. Theoretisch betrachtet kann bspw. eine Intervention Feedback symbiotisch ergänzen, indem diese zusätzliche Antezedenzien eines Verhaltens so beeinflusst, dass die kumulative Wirkung beider Interventionen die Summe ihrer Einzeleffekte übersteigt. Andererseits kann eine Intervention auch die Wirkung von Feedback abschwächen, indem sie den Fokus der Feedback-Empfänger von dem Inhalt der Verhaltensrückmeldung weglenkt. Für die praktische Anwendung von Verhaltensinterventionen hat ein tieferes theoretisches Verständnis über solche Wechselwirkungen ebenfalls eine hohe Relevanz. Einerseits schafft die Kontrastierung unterschiedlicher Interventionen Wissen darüber,

wie sich die Effektgrößen verschiedener Interventionen zu der von Feedback unterscheiden. Zum anderen hilft das Verständnis über Wechselwirkungen mit Feedback bei der Implementierung von ökonomisch effizienten Verhaltenskampagnen, da ungünstige Wechselwirkungen vermieden werden können.

Vor diesem übergeordneten Forschungshintergrund wirft diese kumulative Dissertation mit insgesamt neun Forschungspapieren einen nuancierten Blick auf die Wirkung von Feedback-Interventionen, die darauf abzielen, Individuen bei Verhaltensänderungen in Alltagshandlungen durch Leistungsrückmeldungen zu unterstützen. Dabei werden Forschungsfragen zu den Auswirkungen von Feedback-Interventionen in den Bereichen Umwelt, Online-Lernen und Gesundheit untersucht, wodurch die Dissertation eine weite Bandbreite von Alltagshandlungen – von routinebasierten Handlungen wie dem Händewaschen bis zu kognitiv komplexen Handlungen wie dem Lernen – abdeckt. Zur Beantwortung der Forschungsfragen greift sie auf umfangreiche Verhaltensdaten zurück, welche die Dissertation mittels Adaption vorhandener Informationssysteme (IS) oder der Entwicklung dedizierter IT-Werkzeuge im Feld gewinnt. Damit folgt die Dissertation einer kürzlichen Aufforderung innerhalb der IS-Disziplin (Burton-Jones et al., 2021), die Vorhersagekraft etablierter Theorien unter den neuen Gegebenheiten der digitalen Transformation zu evaluieren, um über deren weitere Passung zu theoretisieren. Für die ganzheitliche Adressierung des Themenfelds beinhaltet die kumulative Dissertation ein einführendes Kapitel und neun Forschungspapiere, die wie folgt auf zwei Kapitel aufgeteilt sind.

Kapitel 1: Instanziierung und Validierung gewährt einen exemplarischen Einblick in die technische Entwicklung und Validierung von IT-Werkzeugen, die in den Forschungspapieren zur Bereitstellung von Feedback benötigt werden. Hierbei zeigt das Kapitel für zwei relevante Anwendungsfälle, dass in den zugehörigen Verbrauchsdaten Informationen enthalten sind, wodurch gewisse Personen oder Verbrauchsereignisse erkannt werden können. Diese Informationen sind damit Grundlage für nachgelagerte Anwendungen, mit denen spezifischere und damit potenziell wirkungsvollere Feedback-Interventionen ermöglicht werden sollen. Ebenso veranschaulicht das Kapitel am Beispiel eines konzipierten Feedback-Systems zur Verbesserung der Handhygiene, dass bestehende Barrieren in der Verbreitung von Feedback-Systemen durch neue Möglichkeiten der Digitalisierung überwunden werden können.

Kapitel 2: Evaluation und Empfehlungen beleuchtet schließlich die Auswirkungen von Feedback-Interventionen auf Alltagshandlungen und untersucht im Einklang mit theoretischen Erkenntnissen, inwiefern die Verhaltensreaktionen von den Eigenschaften der Feedback-Empfänger (z. B. Einstellungen, Ausgangsniveau des Verhaltens) abhängen. Dabei liefert jedes durchgeführte Experiment des Kapitels einen spezifischen Beitrag für die Theorie und bewertet zugleich die praktische Bedeutung der untersuchten Interventionen. Das Kapitel liefert zunächst anhand eines natürlichen Experiments im Umweltbereich Evidenz, dass Organisationen, die ihre CO_2-Emissionen durch Umweltprojekte kompensieren und das salient ihren Kunden gegenüber kommunizieren, den Ressourcenverbrauch der Kunden erhöhen. Während sich der im Experiment beobachtete Mehrverbrauch der Kunden durch die Folgenlosigkeit ihres Handelns auf den

Klimawandel und das daraus reduzierte moralische Verantwortungsbewusstsein erklären lässt (cf. Onwezen et al., 2013), hat die zusätzliche Bereitstellung von Echtzeit-Feedback diesem ungünstigen Verbrauchsanstieg entgegengewirkt. Mit einer anderen großangelegten randomisiert kontrollierten Studie erbringt das Kapitel zudem empirische Evidenz, dass kontingente Geldanreize, entgegen der Vorhersage der Motivation Crowding Theory (Deci et al., 1999), nicht die intrinsische Motivation zum Ressourcenschonen untergraben. Stattdessen führte der evaluierte Geldanreiz zu verhaltensbedingten Ressourceneinsparungen, was gemäß dem Standardmodell der klassischen Ökonomie auf eine Nutzensteigerung des Ressourcenschonens durch den Geldanreiz zurückzuführen ist (cf. Berg, 2010). Interessanterweise bewirkte Verbrauchsfeedback jedoch deutlich höhere Einsparungen, die nicht durch den Geldanreiz verstärkt wurden, was belegt, dass die Wirkung vom Verbrauchsfeedback dominant ist. Ergänzende Analysen, die darauf abzielen, die Effekte hinsichtlich ihrer Heterogenität zu untersuchen, unterstreichen die Robustheit der Haupteffekte in Anbetracht spezifischer Eigenschaften der Studienteilnehmer und zeigen, dass Vielverbraucher tendenziell stärker durch die Interventionen einsparten.

Das Kapitel greift das Thema der Heterogenität auch im Bereich des Online-Lernens auf. Dort leistet das Kapitel einen Beitrag, indem es eine theoretische Erklärung für die heterogenen Effekte von vergleichendem Feedback auf die Lernunterstützung durch digitale Lernplattformen liefert. Hierbei zeigen zwei randomisiert kontrollierte Studien, dass vergleichendes Feedback, das unter Berücksichtigung von theoretischen Implikationen der Social Norms Theory gestaltet wurde (Berkowitz, 2005; Schultz et al., 2007), akademisches Aufschiebeverhalten (sog. Prokrastinationsverhalten) auf digitalen Lernplattformen reduzieren kann. Das einführende Kapitel der kumulativen Dissertation, das *Kapitel 1* und *2* voransteht, liefert zudem ergänzende Informationen, welche belegen, dass dadurch der Kurserfolg der Studierenden verbessert wurde. *Kapitel 2* schließt mit zwei Beiträgen, die einen Ausblick auf weiterführende Feldexperimente geben. So zeigt ein Forschungspapier ein experimentelles Design zur Evaluation einer Feedback-Komponente auf, die durch Anwendung maschineller Lernverfahren personenspezifische Lernvorschläge bereitstellt. Damit sollen digitale Lernplattformen befähigt werden, bei der Lernunterstützung durch Feedback automatisch interpersonelle Unterschiede zu berücksichtigen (cf. Pintrich et al., 1991; Pintrich, 2004). Letztendlich berücksichtigt das Kapitel interpersonelle Unterschiede auch bei der Untersuchung von Verhaltensänderungen im Bereich der Routinehandlungen. Diesbezüglich stellt es ein Forschungsdesign am Beispiel des Händewaschens vor, mittels dessen die Auswirkungen von Verhaltensinterventionen auf relevante Antezedenzien des Händewaschens untersucht und quantifiziert werden können.

Die in der Dissertation gewonnenen Erkenntnisse tragen auf vielfältige Weise zum Stand der Forschung bei. Es wird nicht nur das Verständnis über die Wirkung von IT-basierten Feedback-Interventionen erweitert, sondern auch die Wechselwirkung von Feedback mit anderen Interventionen untersucht. Damit werden widersprüchliche theoretische Vorhersagen aus der Literatur aufgelöst und die Aussagekraft von spezifischen Theorien in Bezug auf Feedback entweder erhärtet oder infrage gestellt. Dabei lassen einige Experimente durch ihren relativ langen Beobachtungszeitraum tiefergehende Schlüsse auf die zeitliche Dynamik und die Stabilität der Verhaltensauswirkungen zu. Über verschiedene Verhaltensdomänen hinweg erweitert somit die Dissertation grundlegend das Verständnis über die Auswirkungen von Feedback, wobei dieses auch oft in direkten Vergleich zu anderen Verhaltensinterventionen gesetzt wird, die anhand von IT skalierbar umgesetzt werden können (z. B. Umweltappelle, Geldanreize, automatischer Emissionsausgleich bei Umwelthandlungen). Gerade hierdurch entstehen einige wichtige Impulse für die Theorieentwicklung. Demnach könnte zukünftige Forschung bspw. bestehende Theorien weiterentwickeln, um deren Aussagekraft zu den Interaktionseffekten von Feedback und anderen Interventionen zu steigern.

Gleichermaßen sind die gewonnenen Erkenntnisse der Dissertation auch für die Praxis relevant. Für verschiedene Handlungen belegen die Ergebnisse, dass Feedback-Interventionen große, wünschenswerte Effekte haben. Feedback kann bspw. nicht nur ungünstige Nebeneffekte von anderen Interventionen verhindern (d. h. Mehrverbrauch durch Klimakompensationsprogramme), sondern auch über längere Zeit verhaltensbedingte Umwelteinsparungen induzieren – und das in Situationen, in denen die Feedback-Empfänger dadurch keinen finanziellen Nutzen haben. Hierbei zeigt insbesondere der Vergleich verschiedener Interventionen, dass Feedback einige nennenswerte Vorteile gegenüber Umweltappellen oder Geldanreizen besitzt. Ebenfalls demonstrieren die Studien im Bereich des digitalen Lernens, dass eine leicht umsetzbare Feedback-Intervention einen deutlichen Effekt auf die Onlineaktivitäten und die Verminderung der Prokrastination von Studierenden haben kann. Somit können die Ergebnisse der Dissertation in IS und IT-Systeme einfließen, bei denen das Ziel besteht, entsprechende Verhaltensänderungen zu bewirken.

Insgesamt veranschaulicht diese Dissertation, welches Potenzial in IT-basiertem Feedback steckt, um bedeutsame Kennzahlen unserer Gesellschaft (z. B. Klimaschutz, Produktivität) zu verbessern. Dabei verdeutlicht die Dissertation insbesondere, dass die damit verbundenen Hard- und Softwarelösungen oftmals noch adaptiert oder gar implementiert werden müssen, um möglichst wirkungsvolle und gleichzeitig verhaltenstheoretisch interessante Feedback-Interventionen zu erzeugen. Denn gerade bei Alltagshandlungen, bei denen kognitive Entscheidungsprozesse primär instinktiv sind und unbewusst stattfinden (d. h. auf System 1-Prozessen basieren nach Kahneman, 2012), erwirken Feedback-Interventionen, die zunehmend durch IT skalierbar ermöglicht werden, verhältnismäßig große Effekte auf das Handeln (cf. Kluger und DeNisi, 1996).

Literaturverzeichnis

Bandiera, O., Barankay, I. und Rasul, I. (2007). “Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment”. *The Quarterly Journal of Economics* 122 (2), S. 729–773. DOI: 10.1162/qjec.122.2.729.

Berg, N. (2010). “Behavioral Economics”. In: *21st Century Economics: A Reference Handbook.* Hrsg. von R. C. Free. Bd. 2. Thousand Oaks, CA, USA: SAGE Publications, S. 861–872. ISBN: 978-1-4129-6142-4.

Berkowitz, A. D. (2005). “An Overview of the Social Norms Approach”. In: *Changing The Culture Of College Drinking: A Socially Situated Health Communication Campaign.* Hrsg. von L. C. Lederman und L. P. Stewart. Cresskill, NJ, USA: Hampton Press, S. 193–214. ISBN: 978-1-57273-593-4.

Brambila-Macias, J., Shankar, B., Capacci, S., Mazzocchi, M., Perez-Cueto, F. J. A., Verbeke, W. und Traill, W. B. (2011). “Policy Interventions to Promote Healthy Eating: A Review of What Works, What Does Not, and What Is Promising”. *Food and Nutrition Bulletin* 32 (4), S. 365–375. DOI: 10.1177/156482651103200408.

Burton-Jones, A., Butler, B. S., Scott, S. V. und Xu, S. X. (2021). “Next-Generation Information Systems Theorizing: A Call to Action”. *MIS Quarterly* 45 (1), S. 301–314. DOI: 10.25300/MISQ/2021/15434.

Deci, E. L., Koestner, R. und Ryan, R. M. (1999). “A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation”. *Psychological Bulletin* 125 (6), S. 627–668. DOI: 10.1037/0033-2909.125.6.627.

Fang, X., Goette, L., Rockenbach, B., Sutter, M., Tiefenbeck, V., Schoeb, S. und Staake, T. (2023). “Complementarities in Behavioral Interventions: Evidence from a Field Experiment on Energy Conservation”. *Journal of Public Economics* 228, Artikelnr. 105028. DOI: 10.1016/j.jpubeco.2023.105028.

Fry, J. P. und Neff, R. A. (2009). “Periodic Prompts and Reminders in Health Promotion and Health Behavior Interventions: Systematic Review”. *Journal of Medical Internet Research* 11 (2), Artikelnr. e16. DOI: 10.2196/jmir.1138.

Hattie, J. und Timperley, H. (2007). “The Power of Feedback”. *Review of Educational Research* 77 (1), S. 81–112. DOI: 10.3102/003465430298487.

Kahneman, D. (2012). *Thinking, Fast and Slow.* London, UK: Penguin Books. ISBN: 978-0-14-103357-0.

Karlin, B., Zinger, J. F. und Ford, R. (2015). “The Effects of Feedback on Energy Conservation: A Meta-Analysis.” *Psychological Bulletin* 141 (6), S. 1205–1227. DOI: 10.1037/a0039650.

Kluger, A. N. und DeNisi, A. (1996). “The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory”. *Psychological Bulletin* 119 (2), S. 254–284. DOI: 10.1037/0033-2909.119.2.254.

Liebe, U., Gewinner, J. und Diekmann, A. (2018). “What Is Missing in Research on Non-Monetary Incentives in the Household Energy Sector?” *Energy Policy* 123, S. 180–183. DOI: 10.1016/j.enpol.2018.08.036.

Onwezen, M. C., Antonides, G. und Bartels, J. (2013). “The Norm Activation Model: An Exploration of the Functions of Anticipated Pride and Guilt in Pro-Environmental Behaviour”. *Journal of Economic Psychology* 39, S. 141–153. DOI: 10.1016/j.joep.2013.07.005.

Pintrich, P. R., Smith, D. A. F., Garcia, T. und McKeachie, W. J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Report. Ann Arbor, MI, USA: National Center for Research to Improve Postsecondary Teaching and Learning. URL: https://eric.ed.gov/?id=ED338122.

Pintrich, P. R. (2004). "A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students". *Educational Psychology Review* 16 (4), S. 385–407. DOI: 10.1007/s10648-004-0006-x.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J. und Griskevicius, V. (2007). "The Constructive, Destructive, and Reconstructive Power of Social Norms". *Psychological Science* 18 (5), S. 429–434. DOI: 10.1111/j.1467-9280.2007.01917.x.

Sierzchula, W., Bakker, S., Maat, K. und van Wee, B. (2014). "The Influence of Financial Incentives and Other Socio-Economic Factors on Electric Vehicle Adoption". *Energy Policy* 68, S. 183–194. DOI: 10.1016/j.enpol.2014.01.043.

Van der Kleij, F. M., Feskens, R. C. W. und Eggen, T. J. H. M. (2015). "Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis". *Review of Educational Research* 85 (4), S. 475–511. DOI: 10.3102/0034654314564881.

Table of contents

Introductory paper

Table of contents

1 Introduction

While everyday activities like commuting to work, being physically active, or helping coworkers do not seem to fundamentally change the world, they can have large effects over time on important macro-level outcomes of our society (e.g., climate protection, life expectancy, or economic productivity). Knowing the relevance of everyday activities, policymakers and organizations share a long tradition of implementing interventions to improve associated behavioral outcomes. Classical examples of such interventions include financial incentives to reward "desirable" behaviors (e.g., bonus payments in organizations, subsidies in policy; see Bandiera et al., 2007; Sierzchula et al., 2014) and information campaigns to make individuals aware of behavioral consequences (see, e.g., Dolnicar et al., 2017). With the ubiquity of information technology (IT), numerous streams of data are now becoming available to policymakers and organizations that quantify behavioral outcomes, empowering more specific interventions. A very promising development in this context is feedback interventions, which can be defined as the provision of person-specific information on one's behavior (e.g., on performance). Research has indicated that feedback interventions can substantially decrease personal and societal issues in everyday contexts (e.g., to improve health behavior, conserve resources, or support self-regulated learning; see Bravata et al., 2007; Sanguinetti et al., 2020; Van der Kleij et al., 2015).

Despite the demonstrated potential of feedback in improving associated behavioral outcomes, there are still many unresolved questions around feedback interventions that are relevant for both theory and practice. First, in many contexts, it is not clear how feedback compares to other behavioral interventions such as financial incentives or information campaigns in inducing behavior change. Yet deeper knowledge about the effect sizes of different interventions is clearly relevant as this allows scholars to contrast the explanatory power of different theoretical models on potential behavioral outcomes. Such a comparison of interventions is also insightful for practitioners who are in the position to implement behavioral programs, aiming to maximize the effectiveness of the programs (see, e.g., Milkman et al., 2022). Next, the interplay between feedback and other behavioral interventions is not well understood as related theories often provide conflicting predictions with several conceivable behavioral outcomes. For instance, the combination of interventions might lead to higher "desirable" effects by simultaneously tackling different barriers to behavior change (see, e.g., Huis et al., 2012; Michie et al., 2011; van Valkengoed et al., 2022). Conversely, an intervention could also attenuate the effects of feedback by, for example, directing individuals' attention away from the feedback information (Kluger and DeNisi, 1996). Although many studies combine feedback information with other interventions, their experimental designs often do not shed light on the individual effect of each intervention and thus neglect the presence and nature of interaction effects (see Fang et al., 2023; Liebe et al., 2018). The consequences of such experimental designs are twofold.

On the one hand, they limit the development of the theoretical understanding of feedback interventions, whose effects on individuals are already complex as they substantially depend on task- and situation-specific variables (Kluger and DeNisi, 1996). On the other hand, they might also lead to suboptimal behavioral programs when practitioners combine feedback with other interventions that, in the end, decrease rather than increase associated behavioral effects (as indicated by, e.g., Asensio and Delmas, 2015; Manthei et al., 2023; Pellerano et al., 2017; Sudarshan, 2017).

In comparing the effects of feedback information to those of other interventions in the field, studies could respond to a recent call for research in information systems (IS) research (Burton-Jones et al., 2021). Many theories and models related to behavioral interventions have been tested, evaluated, and refined under relatively strict assumptions (i.e., in highly controlled environments), making it difficult to project their findings onto real-world situations. For example, theoretical phenomena are often studied over relatively short periods (e.g., a few hours to days, with few data points per individual). As individuals might become used to an intervention, the impact of the intervention might wear off over time in real-world applications (e.g., because of satiation or habituation effects; see McSweeney, 2004). Another implication of controlled environments is that individuals might feel more observed in shorter experiments, decreasing the external validity of the findings. In fact, research has demonstrated that individuals systematically behave differently when observed by third parties (Schwartz et al., 2013). A plausible reason for this behavior is that individuals tend to have the desire to convey a positive self-image to others (Ariely et al., 2009) and therefore respond to an observation by third parties by behaving in a socially acceptable way (e.g., overstating one's efforts for altruistic or pro-environmental behavior). Since such effects are difficult to capture in controlled environments, both negative and positive effects may be overestimated. As field research can mitigate these biases, the overall implication is that field experiments are valuable in informing (i) practitioners' creation of (cost-)effective behavioral programs and (ii) scholars' knowledge of theoretical phenomena (see, e.g., Harrison and List, 2004; Levitt and List, 2009). The good news is that many IT systems are already in the field and can serve as a basis for IS-related research, often with modest associated costs to study behavior across various domains (e.g., in the form of smartphone sensors, smart electricity meters, digital learning platforms, shopping platforms, etc.). In addition, such systems often provide the possibility to explore the behavioral responses of individuals with different socio-demographic backgrounds, yielding macro-level insights for both organizations and policymakers.

Against this backdrop, this dissertation seeks to expand understanding of how technical opportunities for behavior change, namely the provision of feedback, can have a meaningful real-world impact. To this end, the dissertation assesses the effects of specific feedback interventions across several behavioral domains, with the overarching goals of (i) evaluating theoretical predictions that are in conflict and (ii) revisiting theoretical findings in other contexts. As a theoretical lens, the dissertation uses an adaptation of the belief–action–outcome framework

(Melville, 2010). This framework is based on Coleman's micro-macro scheme (Coleman, 1986; Coleman, 1990) to rigorously explain how social phenomena, opportunities, or events translate to macro-level outcomes. According to Coleman, macro-macro links, such as the impact of the ubiquity of IS on environmental sustainability, cannot be studied at the macro level due to data inadequacies (e.g., confounding variables, too little variation to explain the respective outcome). Instead, researchers must measure the impact of a phenomenon on the individual level before returning to the macro-level variable in question (Coleman, 1986; Coleman, 1990). It is important to note that research does not have to address all three links to make a meaningful contribution but can also focus on a single link (Coleman, 1986). Figure 1 depicts the overall research framework and structure of the dissertation.

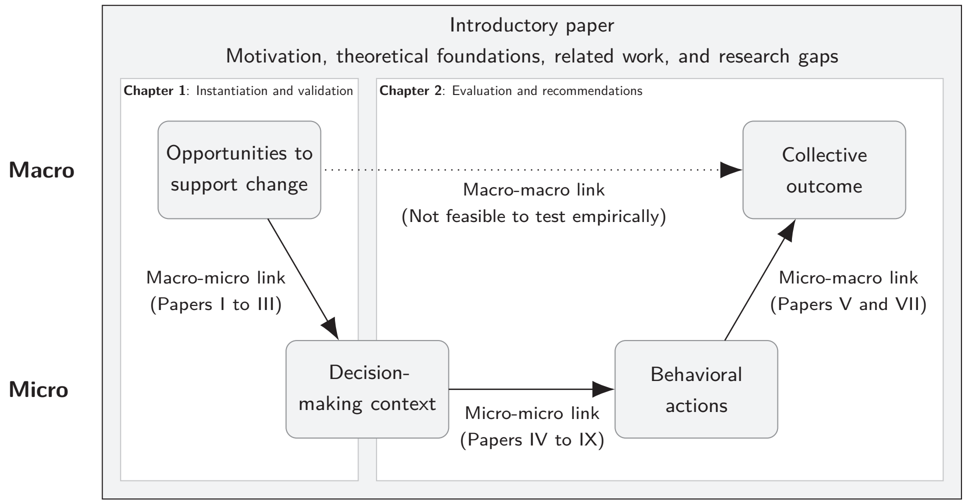


Figure 1: Overall research framework of the dissertation

The dissertation addresses each of the elements in Figure 1 (i.e., the macro-micro, micro-micro, and micro-macro links) by studying specific questions associated with supporting individuals to make behavioral changes. Consequently, the dissertation contributes not only by reviewing existing literature but also by programming, adapting, and evaluating IT tools to generate insights into how technical opportunities for behavior change could influence micro- and macro-level outcomes if implemented. The resulting IT tools represent a core research instrument for this dissertation. Based on the data the IT tools captured, the dissertation primarily applies statistical methods to generate its evidence. Hence, the dissertation takes a positivist perspective that assumes that "data truly measures reality" (Weber, 2004, p. iv) and which is independent of a researcher and their instruments (Myers, 2013). This perspective is reflected in the nine research papers in this dissertation that follow the introductory paper. All the IT tools presented in the research papers are primarily centered on feedback, namely the provision of person-specific information on one's behavior (i.e., on performance), as an investigated mechanism for behavior change. Beyond its focus on feedback, this dissertation encompasses experiments that also test alternative interventions (e.g., performance-contingent financial incentives for behavior change) to draw broader macro-level recommendations for both organizations and policymakers. Such a macro-level recommendation could be that financial incentives for resource conservation work strongly for individuals with low intrinsic motivation for resource conservation and that for individuals with high intrinsic motivation, feedback interventions are more (cost-)efficient; this would imply that IT systems for behavior change could maximize their overall impact by approaching subgroups of individuals differently (i.e., different interventions per subgroup). Before continuing with outlining the relationships between the research framework and the papers of this dissertation, Table 1 presents the papers and provides an overview of their aims.

Table 1: Overview of the papers in this dissertation

Paper	Context	The aim of the paper
I	Environmental behavior	Empower feedback on different end uses (i.e., appliances and fixtures) through measurements of a single sensor attached to a household's main water pipe.
II	Environmental behavior	Enable personalized feedback on a shared fixture by reliably differentiating between different end users based on consumption data.
III	Health behavior	Identify the adoption issues of hand hygiene monitoring systems and propose a novel feedback system that overcomes the identified issues.
IV	Environmental behavior	Studies indicate that feelings of guilt and responsibility for the environmental consequences associated with one's actions are antecedents to performing pro-environmental behaviors. A growing number of organizations are compensating for the greenhouse gas emissions associated with consumers' demand through green initiatives without passing on the costs of offsetting to consumers. However, the compensation for emissions might reduce consumers' feelings of guilt and responsibility and thereby dampen their curtailment efforts. In this paper, a natural field experiment tests the impact of carbon offset programs on conservation behavior and evaluates feedback as a countermeasure against unintended effects (i.e., higher consumption levels).
V	Environmental behavior	In environmental research, there is the omnipresent fear that financial incentives to reward resource conservation could backfire by crowding out individuals' intrinsic motivation for resource conservation. To better understand the implications of motivation crowding theory, this paper aims to (i) investigate the individual effects of financial incentives and consumption feedback and (ii) assess their interplay through a randomized controlled trial. In doing so, the paper sheds light on whether financial incentives weaken or amplify the conservation effects induced by feedback. As the experiment includes treatment groups that do not receive such incentives, the paper can rule out several alternative explanations that could have masked crowding-out effects in previous research.
VI	Learning behavior	Comparative feedback is a popular intervention to support learning on digital learning platforms, yet, it has led to mixed effects in the past, meaning that it can even harm learning by discouraging online activity among platform users. This paper puts forth the argument that mixed effects can be explained from the perspective of social norms theory. More specifically, some learners might have the erroneous normative belief that others in their cohort are quite active on a learning platform even when they are not. Descriptive comparative feedback could resolve such beliefs and thus reduce learners' extrinsic motivation for using the platform. Interestingly, research on social norms has shown that injunctive norms can counteract such adverse effects in other behavioral domains. Against this backdrop, this paper tests the effects of a descriptive comparative feedback intervention to reduce procrastination in higher education students of an elective master's level course. By featuring an injunctive norm and conducting subgroup analyses, it tests whether there are adverse effects in relevant subgroups. In doing so, the paper aims to generate evidence that the injunctive norm has counteracted adverse effects.

(To be continued on the next page)

Table 1: Overview of the papers in this dissertation (continued)

Paper	Context	The aim of the paper
VII	Learning behavior	While Paper VI reports desirable effects of the feedback intervention on students' online learning behavior without any evidence of heterogeneous treatment effects, Paper VII investigates the robustness of these findings. To this end, it presents the results of a follow-up experiment conducted with a compulsory bachelor's level course. By conducting subgroup analyses with a much larger sample size, the paper seeks to generate further evidence that the injunctive norm has counteracted the potential adverse effects of descriptive comparative feedback.
VIII	Learning behavior	According to self-regulated learning theory, learners substantially differ in their self-regulation strategies and personal characteristics. Previous studies have commonly neglected this aspect in their feedback design, even though theory suggests that it is critical to account for such differences in order to provide meaningful feedback to learners. This paper seeks to address the heterogeneity of learners by presenting a feedback intervention that aims to provide personalized feedback at scale. Moreover, the paper sheds light on the planned experimental evaluation with students in higher education and describes expected results from the perspective of self-regulated learning theory.
IX	Health behavior	According to the literature, behavioral interventions such as feedback affect multiple antecedents relevant to inducing behavior change. However, understanding of the exact mechanisms is limited as there is little experimental research quantifying these influences. This paper proposes an experiment to investigate the cause-effect mechanisms of feedback and salient social norms on handwashing behavior. In addition, the paper proposes a mathematical model for quantifying the effects of these interventions on related antecedents of handwashing behavior.

Returning to the research framework, all nine papers in this dissertation are related to the first link (i.e., macro-micro) by demonstrating how IT systems could alter individuals' decision-making contexts to induce behavior change. In doing so, the papers make use of recently emerged sensors and IT hardware with the primary aim of providing feedback on different behaviors (i.e., resource conservation, learning, and handwashing behavior). It is important to note that the development and adaptation of IT tools take only a subordinate role in the text of the papers, with many focusing on the associated behavioral effects (the micro-micro link). However, three papers in this dissertation (i.e., Paper I, II, and III) present more details on the underlying technical implementation by investigating the technical or organizational feasibility of IT tools for feedback provision, thereby directly addressing the first link.

Four of the remaining six papers address the second link (i.e., micro-micro) by highlighting how changes to individuals' decision-making context influence their subsequent behavioral actions. The resulting insights subsequently (i) inform (new) IT tools that potentially support behavior change at scale without eliciting negative reactions (i.e., reactance, frustration) from their users and (ii) advance the understanding of theoretical phenomena. Beyond summarizing the results and contributions of these papers, the dissertation therefore also provides more in-depth insights into the development, adaptation, and monitoring of the associated IT tools. Lastly, two papers in this dissertation present preliminary work concerning the micro-micro link.

To address the third link of the research framework (i.e., the micro-macro link), one needs to account for differences in individuals (e.g., attitudes, goals, etc.) that could influence the macro-level outcomes of the interventions. The dissertation considers this in two papers (namely Paper V and Paper VII), with one from the environmental domain and one from the education domain.

The content of the nine papers is structured into two different chapters that follow the introductory paper. More specifically, Chapter 1 focuses on the presentation of the developed and adapted IT tools for facilitating feedback interventions. Chapter 2 is primarily focused on testing the effects of feedback interventions on individuals' behavior (the micro-micro link). By taking differences among individuals into account, this chapter also sheds light on the potential effects of feedback interventions at the macro level (i.e., the micro-macro link).

By focusing on these three overarching links of the research framework, the dissertation demonstrates that such opportunities for feedback provision can have large effects on improving personal and societal outcomes while still being relatively easy to implement. Especially noteworthy are two papers from the environmental domain. The first one provides substantial evidence that feedback can mitigate the higher levels of consumption that ensue when motives for resource conservation are eliminated through compensation of consumption-related CO_2 emissions. More specifically, even though the seeming inconsequentiality of one's actions for climate protection led to higher consumption levels, consumption feedback counteracted this increased resource usage. The second paper provides robust evidence that small, performance-contingent financial incentives—contrary to the predictions of motivation crowding theory (e.g., Deci et al., 1999)—do not crowd out intrinsic motivation for resource conservation but rather induce resource conservation. The good news is that practitioners do not have to provide financial incentives as the conservation effects of feedback were substantially higher and dominant over financial incentives. In other words, while financial incentives were effective alone, they did not increase the effects of feedback. Given the scalability of feedback interventions through IT systems, the dissertation documents that IT systems can indeed fundamentally tackle environmental issues with low marginal costs. With these insights, the dissertation provides a blueprint for practitioners to create further IT tools that may fundamentally change personal or societal outcomes. At the same time, it informs scholars by advancing research on theoretical phenomena in the field. For instance, it evaluates theoretical predictions on the interplay of feedback information and other interventions that are in conflict. In addition, it revisits theoretical predictions that have been produced mostly in artificial settings (e.g., motivation crowding theory; see Esteves-Sorenson and Broce, 2022) or that result from correlational evidence (e.g., the first and second proposition of feedback intervention theory; see Kluger and DeNisi, 1996). Based on the resulting findings, the dissertation identifies numerous future research avenues.

This introductory paper is structured as follows. Section 2 begins by providing the theoretical foundations of feedback interventions. This includes an overview of their characteristics, the associated behavioral mechanisms, and the potential interplay of feedback with other interventions. The subsequent section presents the research gaps of this dissertation, along with a description of domain-specific theories. As a consequence, this section presents a theoretical lens on how feedback interventions may operate within specific domains, providing a basis for the research questions that will subsequently be presented for each paper that focuses on behavioral effects. Section 4 explains the methodology of the papers, from the development and adaptation of affiliated IT tools to a description of the employed research methods. Section 5 outlines the data analysis methods used in the papers. The sixth section summarizes the motivation and results of each paper. Based on these results, Section 7 outlines the contributions to the literature and the implications for practice. After discussing the limitations and future research avenues in Section 8, the introductory paper ends in Section 9 with a brief conclusion.

2 Theoretical foundations of feedback

This section presents the theoretical foundations of feedback interventions. It first provides an overview of feedback intervention characteristics before describing feedback intervention theory in detail, which presents a theoretical lens on the general cause-effect relationships between feedback interventions and behavior change. Ultimately, the section ends by shedding light on potential interaction effects between behavioral interventions, which are necessary for understanding the interplay of feedback information with other interventions. In doing so, this section presents a general overview of the theoretical foundations of this dissertation. A deep dive into the theoretical foundations as related to the research questions is provided in Section 3 and, more detailed, in the corresponding papers.

2.1 Overview of feedback intervention characteristics

Research on the effects of feedback interventions has been published for over 100 years in psychology and associated behavioral domains (Kluger and DeNisi, 1996). An underlying reason is that feedback information plays a fundamental role in skill learning and improving behavioral outcomes. From learning first words as a child to developing complex skills as an adult (see, e.g., Clark, 2016; El Boghdady and Alijani, 2017), feedback helps individuals recognize discrepancies between their performance and their goals and activate cognitive processes to reach those goals. As feedback is ubiquitous in our lives, it is important to generate an overview of the specific characteristics of feedback interventions. These characteristics are not only helpful in illustrating the scope of this dissertation, they also have substantial influence in correctly anticipating and understanding the effects of feedback interventions.

To begin, the dissertation defines feedback as a process of providing an individual with information on their behavior to modify their subsequent actions, with the overarching goal of improving behavioral outcomes (similar to, e.g., Hattie and Timperley, 2007; Kluger and DeNisi, 1996). Hence, presenting an individual with information on the general consequences of a behavior (e.g., of environmental or health issues) is excluded from this definition of feedback since the individual does not receive any personal information on their behavior. The definition does not make any further specifications regarding the content of feedback. For instance, it includes interventions that provide information on the effectiveness of an individual's actions for improving associated outcomes (e.g., Boehler et al., 2006; Kelley and Miltenberger, 2016) and interventions that primarily make behavioral outcomes salient (e.g., Choudhary et al., 2021; Tiefenbeck et al., 2018).

In practice, the content of feedback information depends inherently on the target behavior, and there are substantial differences in acquiring the target behavior's underlying data. For some behaviors, such as job performance or public speaking, it might be challenging to quantify behavioral outcomes numerically to provide meaningful feedback. Here, the content of feedback information could be primarily based on the opinions or experiences of others (for exemplary interventions, see King, 2016; Rivera et al., 2021). For other behaviors, such as resource consumption or online learning, there are an increasing number of data streams that quantify associated outcomes. In this area, the strengths of IT come into play. Fully automatic processing of data streams enables feedback interventions at scale that provide individuals with timely information on their behavioral actions. This dissertation positions itself within this area by answering research questions on feedback interventions that are intended to present behavioral outcomes measured by IT in a tangible form (e.g., displaying liters of water used or time spent in minutes in a learning environment to feedback recipients). In this context, it is crucial to also speak more about the target behaviors that feedback interventions address as they have implications for the characteristics of feedback interventions. More specifically, from relatively habitual behaviors, such as recycling, to complex tasks, such as acquiring surgical skills, the target behaviors that feedback interventions seek to change differ substantially in their complexity.

One very intuitive way of categorizing behaviors in relation to their complexity comes from cognitive and social psychology, which differentiates between two modes of processing information (see, e.g., Evans, 2008; Smith and DeCoster, 2000). The underlying processes are usually mapped onto two cognitive systems (Evans, 2008). System 1 processes operate "automatically and quickly, with little or no effort" (Kahneman, 2012, p. 20). Hence, System 1 usually represents the instinctive, gut-level decision-making in individuals, which is responsible for our ability to detect emotions in others' voices or to understand simple sentences. In contrast, System 2 processes are responsible for dealing with more complex activities and can be seen as the reasoning self that "decides what to think about and what to do" (Kahneman, 2012, p. 21). In doing so, System 2 allocates attention, which is limited in individuals, to activities that demand it (Kahneman, 2012). Such activities comprise, for example, learning the vocabulary of a foreign language, filling out a tax form, and formulating a paragraph. While it is relatively slow in processing, System 2 is usually activated when System 1 faces an unexpected issue that it cannot solve (Kahneman, 2012). In doing so, both systems complement each other: System 1 uses mental resources sparingly for processing routine activities, while System 2 is active when activities require more mental processing.

Depending on the corresponding target behavior, this perspective on cognitive processing has different implications for the behavioral mechanisms induced by feedback interventions. For behaviors with low mental involvement, feedback interventions could just induce behavior change by making behavioral outcomes salient and thereby overriding automatic processes that otherwise lead to suboptimal outcomes (i.e., those of System 1). For behaviors that already require more mental processing (i.e., System 2 processing), feedback interventions could support individuals by providing them with tangible information on how to improve the associated behavioral outcomes (e.g., by highlighting why certain actions do not lead to the desired outcomes). Consequently, the complexity of the target behavior might play a fundamental role in the choice of content in feedback interventions.

Apart from the target behavior, there are further crucial differences in the immediacy of feedback interventions. Specifically, the notion of immediacy refers to the timing of feedback information, describing the delay between a behavioral action and the corresponding provision of feedback. Here, large differences in feedback interventions exist. While some interventions provide individuals with feedback information during an activity (see, e.g., Dahlinger et al., 2018; Jung et al., 2010; Tiefenbeck et al., 2018), meaning without any sort of delay, other interventions ensue at some point after the completion of the respective activity (see, e.g., Choudhary et al., 2021; Handgraaf et al., 2013; Kandul et al., 2020; Rivera et al., 2021). As there are conflicting terms in the literature (Stingl, 2023), this dissertation refers to the former type of immediacy as *real-time feedback* and the latter as *outcome feedback*. Many studies suggest that immediacy plays an important role in anticipating the effects of feedback. More specifically, these studies indicate that feedback should be provided in a timely manner to be (most) effective (Burton-Jones and Grange, 2013; Kulik and Kulik, 1988). A plausible reason is that timely feedback information reduces individuals' efforts in linking the outcomes made salient by the feedback to their actions and ultimately corrects perception errors related to the effectiveness of their actions (Metcalfe et al., 2009). As a consequence, individuals might find it easier to understand whether their efforts are paying off. However, this is not true for all target behaviors. There is evidence that for behaviors requiring more mental processing, a delay in feedback can be beneficial (Hattie and Timperley, 2007). This may be because the delay in feedback allows individuals to mentally process the corresponding activity deeper, as more immediate feedback (i.e., real-time feedback or immediate outcome feedback) might distract individuals from doing so by increasing their cognitive load.

Another important characteristic of feedback interventions is related to the specificity of their content. More precisely, in many contexts, feedback information can be given at different levels of detail to induce behavior change. For instance, consider an environmental campaign that seeks to reduce household electricity consumption. Feedback information could be given on the total consumption level of a household or on specific activities or appliances contributing to the total consumption level. In the former case, residents might have to determine on their own which measures they should take, while in the latter case, the intervention could point them to specific activities or appliances with large potential for electricity conservation. In general, there is evidence that greater specificity of feedback information is associated with higher immediate effects (Goodman and Wood, 2004; Goodman et al., 2004). Greater feedback specificity might have a guiding function by, for example, showing recipients which behaviors are appropriate or inappropriate (Adams, 1987; Anderson, 1982), thereby resolving perception errors in the relationships between actions and the resulting outcomes. Such a perception error in the environmental domain could be the belief that some behavioral actions lead to substantial electricity savings while, in reality, the savings are miniscule (see, e.g., Gerster et al., 2020). Greater specificity of feedback information can, however, also have undesirable effects. For example, Goodman and Wood (2004) demonstrated that greater specificity can be detrimental to individuals' learning of how to respond to poor performance levels. An underlying reason is that higher feedback specificity can discourage individuals' exploration processes for identifying responses to improve associated outcomes as the feedback might already contain this information. Following this reasoning, individuals might be worse off when they are later more independent, not receiving feedback, and things are going poorly (Goodman and Wood, 2004).

The last characteristic of this overview, which is not exhaustive in nature, is related to the frequency of feedback. The notion of frequency refers to how often an individual receives feedback. For instance, outcome feedback could be provided after every completion of an activity or less frequently (e.g., only for half the completions). One motive for different feedback frequencies is related to potential trade-offs in effects (see, e.g., Lam et al., 2011; Lurie and Swaminathan, 2009; Wulf et al., 2010; Wulf et al., 1998). A higher feedback frequency might be more helpful for individuals in identifying effective strategies to improve associated outcomes (Lam et al., 2011). At some threshold, a higher frequency might, however, also have adverse effects. Individuals might experience more stress in responding to and processing the feedback intervention (Lam et al., 2011). As a consequence, individuals' attention could be more likely to shift away from the feedback information, which would reduce its effects. Indeed, Lam et al. (2011) provided evidence in an experiment of a curvilinear impact of feedback frequency, meaning that a higher frequency is only beneficial up to a certain threshold. Interestingly, the authors also indicate that the threshold depends on the familiarity of the feedback recipients with the corresponding activity. In other words, the more individuals are familiar with the activity, the less they could be overwhelmed with processing highly frequent feedback.

To summarize, feedback interventions differ in several characteristics that are relevant to correctly anticipating and understanding the effects of feedback. The following section complements these characteristics by outlining a general theory of the behavioral mechanisms associated with feedback interventions.

2.2 Behavioral mechanisms behind feedback interventions

Feedback can improve performance across multiple behaviors (e.g., test performance, attendance, memory tasks, physical tasks, etc.), but its negative effects have been long overlooked. To remedy this, Kluger and DeNisi (1996) conducted an extensive review of the literature and subsequently developed a theory that explains both the negative and positive effects of feedback interventions on performance. This theory incorporates many other behavioral theories, such as control theory (Carver and Scheier, 1981), goal-setting theory (Locke and Latham, 1990), and social cognitive theory (Bandura, 1997), and interweaves them in a process-oriented way by describing how individuals evaluate feedback interventions.

At its core, the overarching mechanism by which feedback operates is the manipulation of attention. Feedback directs attention to potential discrepancies between one's performance and their standards (namely goals, aspirations, etc.), leading to a set of subsequent decision paths that take place at different attention levels (Kluger and DeNisi, 1996). Before outlining the attention levels, it is important to note that Kluger and DeNisi (1996) do not provide probabilities regarding which path will be activated. Rather, such path-associated probabilities have to be determined through empirical research that specifically tests the theoretical model of Kluger and DeNisi (1996).

According to feedback intervention theory (Kluger and DeNisi, 1996), feedback usually directs attention to the task level. Here, individuals compare the outcomes of their actions to their standards. The resulting feedback sign can be (i) positive (if the outcome surpasses their standards), (ii) neutral (if the outcome conforms to their standards), or (iii) negative (if the outcome deviates from their standards). Based on the feedback sign, individuals regulate their efforts differently. For instance, when the sign is positive, individuals tend to increase their efforts if they perceive the received feedback to be an opportunity to reach other self-goals. Otherwise, if the efforts do not serve other self-goals, efforts might be reduced, and, thus, performance might drop.

In particular interesting is the description of feedback intervention theory when the feedback sign is negative and individuals cannot resolve the negative feedback sign by increasing their efforts. In this case, other attention levels besides the task level might come into play. More specifically, attention might shift to the task detail level. On this level—when the task is not well-known—recipients may generate and test new hypotheses to improve their performance. This can foster learning (i.e., individuals may recognize which strategies are effective in reaching their standards) and subsequently improve their performance. At the same time, the task detail level can also have adverse effects on performance when the generated hypotheses are unrealistic or incorrect, causing recipients to stop trying and quit the task (due to, e.g., frustration). Similarly, when the task is already well-known, recipients may experimentally test new strategies to improve their performance, which might temporarily harm their performance.

Instead of the task detail level, attention may also shift away from the task level to the self level. Here, feedback recipients may wonder whether it is in their behavioral control to resolve the standard gap (i.e., the discrepancy between one's performance and their standards) and whether the task is important for achieving their self-goals. As a consequence, they may quit the behavior or reduce their efforts when the task is not deemed relevant for themselves. Otherwise, only if the task is deemed easy performance will improve. This is because, that for complex tasks, the shift to the self level requires cognitive resources to evaluate the feedback intervention, distracting the individual. This could lead to performance loss, especially when the task is performed in parallel, as feedback could reduce the available cognitive resources for performing the task (consistent with cognitive load theory; see Sweller, 2011). As indicated, feedback intervention theory provides a complex viewpoint on the effects of feedback interventions, which are summarized in Table 2.

Table 2: Overview of the main mechanisms described by feedback intervention theory

Level	Overarching role and underlying cause-effect relationships
Self	This level is primarily associated with managing the attentional effects of feedback interventions on the meta-task processes of the self. – If the task is not important for achieving self-goals, individuals might diminish their efforts for the task. Instead, individuals might focus on other things. This should ultimately reduce task performance. – If the task is important for one's self-goals and easy, performance should improve. – If the task is important for one's self-goals and not easy, performance should deteriorate.
Task	This level is primarily responsible for responding to the feedback-standard discrepancy by regulating one's efforts toward the task. – If there is no discrepancy between individuals' performance and their standards, they tend to maintain their efforts. – Individuals tend to respond to a negative feedback standard discrepancy by increasing their efforts. If that does not help, individuals might shift their attention either to the self level (if they do not believe in success) or the task detail level (if they believe that learning helps). – If individuals surpass their standards, they tend to decrease their efforts. Only in cases where it helps them to reach other self-goals, they increase their efforts.
Task detail	This level is primarily responsible for learning processes regarding the task. – For well-known tasks, paying attention to this level might deteriorate performance, as individuals might be distracted from the task through the feedback intervention. – For other tasks, individuals may generate hypotheses on how to improve their performance. If the hypotheses are objectively correct, individuals could have a positive learning effect, which should improve performance. Otherwise, individuals might have no or a negative learning effect, which can adversely affect performance. If the hypotheses do not match reality and their task motivation is still high enough, individuals may continue developing and testing hypotheses.

Apart from these effects of feedback, Kluger and DeNisi (1996) provide behavioral predictions (i.e., propositions) on feedback interventions. They suggest that specific cues in feedback interventions could additionally moderate the effects of feedback information. For instance, cues that draw attention to the self (e.g., to meta-task goals) should reduce performance as individuals could subsequently question the relevance or integrity of the feedback information. In contrast, cues can also have a positive effect when they direct additional attention to the task by enhancing task motivation or when they correct erroneous hypotheses (e.g., on the effectiveness of certain strategies) so that individuals can more effectively achieve their standards. Interestingly, the authors also supposed that feedback should be most effective for non-complex tasks, as feedback might not allocate attentional resources that are required for task completion. In the associated paper, they found evidence for most of these predictions through a meta-analysis. In addition, Kluger and DeNisi (1996) suppose that personality variables have a large influence on the evaluation of feedback information. For instance, individuals who tend to avoid negative stimuli might be more likely to shift their attention to the self level after receiving feedback, which might reduce subsequent task performance. Individuals without this attitude might pay more attention to the other levels of the attention hierarchy, in which they are more likely to maintain or improve their task performance level.

Taken together, feedback intervention theory provides a coherent understanding of the effects of feedback that is backed by decades of empirical research; this understanding substantially informs research that seeks to reduce personal or societal issues through feedback interventions. Feedback intervention theory abstracts from concrete behaviors but rather provides a generic description of the cause-effect relationships related to such interventions. Consequently, it provides important information on the design of effective feedback interventions. However, it is also crucial to further understand domain-specific theories of individual behavior. Depending on the individual behavioral domain, such theories could fundamentally shed light on the moderating effects of feedback interventions by uncovering which factors shape individuals' standards. Moreover, such theories might indicate domain-specific barriers to behavior change that feedback could tackle. For instance, in the environmental domain, individuals may often have erroneous perceptions of the environmental consequences of their actions (e.g., which household activities are most energy-intensive). Feedback could resolve such erroneous hypotheses and foster learning at the task level (i.e., resource use at the appliance and fixture level), potentially leading to substantially increased conservation behavior.

2.3 Potential interaction effects of behavioral interventions

As outlined, feedback information is often combined with other behavioral interventions. An underlying reason for this is that a combination of interventions might simultaneously tackle multiple barriers to behavior change, potentially inducing higher effect sizes compared to the provision of a single intervention (van Valkengoed et al., 2022). Yet the question of how interventions—or, in a broader sense, different treatments—might complement each other is of a complex nature and relevant in many other domains (e.g., toxicology or epidemiology; see Roell et al., 2017). For instance, medical research has observed that a combination of therapies might be beneficial to patients. Multiple therapies can be used at certain doses to minimize side effects while potentially increasing the desired therapeutic effects compared to a strategy with a single therapy (see, e.g., Gradman et al., 2010; Mokhtari et al., 2017). Before describing the specific types of interaction effects, it is important to note that there are many synonyms for interaction effects (Roell et al., 2017). This section therefore refers to terms that have been used in both behavioral and biomedical studies.

Turning to the potential interaction effects, the related literature discusses three different types, which are illustrated by Figure 2. First, the concept of additivity describes that the estimated effect of combined interventions results from the individual effect of each intervention (Roell et al., 2017; Welton et al., 2009). More precisely, when intervention A has an effect of 15% and intervention B has an effect of 30%, additivity implies that both result in a combined effect of 45% (see Figure 2a). The assumption that interventions work independently so that their effects can be simply summed up is, however, often not realistic. For instance, feedback intervention theory describes that cues (i.e., other elements or interventions) might fundamentally change the interpretation and perception of feedback information and its effects on behavior.

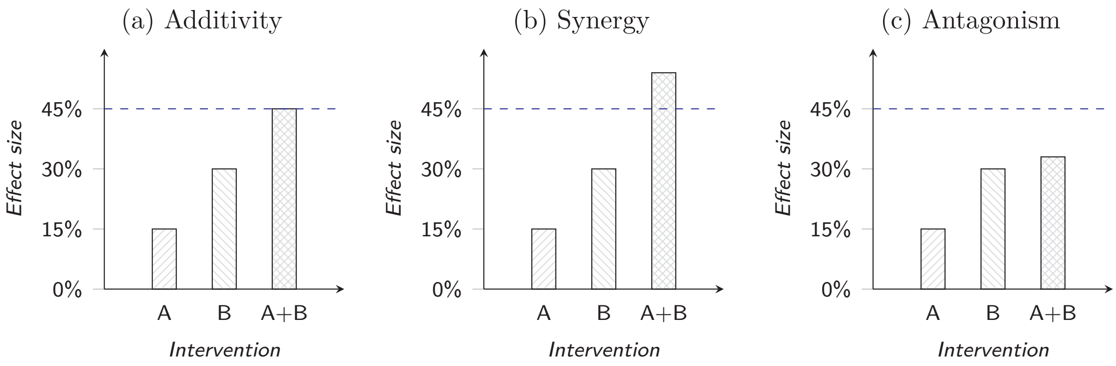


Figure 2: Potential interaction effects of two interventions

In general, there are two further hypothetical outcomes apart from additivity. On the one hand, a combination of two interventions might act synergistically so that the overall effect is greater than the sum of the individual effects, as displayed in Figure 2b (Roell et al., 2017; Welton et al., 2009). This might be the result when interventions tackle different barriers to behavior change (see, e.g., van Valkengoed et al., 2022). For example, Fang et al. (2023) revealed that an intervention that primarily provided information on the general consequences of an environmental behavior and that was ineffective alone substantially enhanced the effect of a feedback intervention that made the associated resource use more salient. A plausible reason for this interplay is that the first intervention raised individuals' awareness of the relevance of their resource use so that they subsequently paid more attention to the corresponding feedback information. On the other hand, it could also be that the combined effect of interventions is less than the sum of the individual effects, which is also known as antagonism, as displayed in Figure 2c (Roell et al., 2017; Welton et al., 2009). A potential reason for such outcomes is that both interventions primarily operate through similar mechanisms (e.g., by directing attention toward resource conservation) so that the interventions can be seen as substitutes for behavior change (Fang et al., 2023). As long as the combined effect exceeds the individual effects of the interventions, a combination can still be favorable for organizations and policymakers to implement.

It is, however, important to illustrate two further potential outcomes related to antagonism. Another scenario could be that two interventions are pure substitutes, such that the combined effect equals the intervention with the highest effect. The intervention with the highest effect could then be seen as the dominant factor facilitating behavior change, since the second intervention has no marginal effect on behavior. Lastly, the combined effect could also be less than the individual effect of each intervention. Such an adverse interplay could result from a situation in which one intervention fundamentally changes the interpretation of the second intervention. A related example is given by feedback intervention theory. If a cue (i.e., another element or intervention) changes the processing of feedback information so that individuals' attention shifts to the self level rather than to the task level, reduced effects can be expected (see Section 2.2).

To summarize, anticipating the interplay of behavioral interventions is a complex task since several outcomes might be conceivable from a theoretical perspective. However, experimental studies often test combinations of behavioral interventions without shedding light on the individual effects of interventions (see, e.g., Fang et al., 2023; Liebe et al., 2018). As a consequence, it is unclear what the interplay of the underlying behavioral interventions is, with consequences for theory and practice.* In particular, the absence of experimental studies that assess interaction effects limits our theoretical understanding of how interventions could complement each other. To put this into perspective, behavioral interventions might influence multiple determinants of behavior (e.g., by making social norms salient, correcting perception errors, etc.). Without an adequate understanding of how interventions operate individually, theorizing on their interplay might lead to spurious results. From a practical perspective, this lack of understanding might result in suboptimal decisions for organizations and policymakers when they combine interventions in their behavioral programs. Consider the case where a policymaker combines financial incentives with consumption feedback but is unaware of their individual effects. It could be that the additional provision of financial incentives for behavior change has no marginal effect on resource conservation when consumption feedback is already in place. Following this reasoning, the policymaker might end up with a smaller campaign and thus produce less environmental impact than what would be possible, as the policymaker could have used the budget spent for financial incentives to reach more people. Against this backdrop, the papers of this dissertation that seek to answer questions on the interplay of interventions deliberately assess the individual effects of the associated interventions.

* For a more profound understanding, it should be clarified what is considered to be a combination of behavioral interventions: even when a small element, such as a normative statement, is added to feedback information, the dissertation refers to this as a combination of interventions (i.e., feedback + normative statement). This is because seemingly small design choices might have a substantial impact on behavioral responses through interaction effects (see, e.g., Ableitner et al., 2017; Schultz et al., 2007).

3 Complementary foundations and their relation to the research questions

After this rather general information on feedback and its effects on human decision-making, this section describes the research questions related to behavior change that the dissertation and its papers address. As the dissertation encompasses research questions on different behavioral domains (environmentally sustainable behavior, learning behavior, etc.), this section is organized into five subsections that follow a specific structure. Each of the subsections first presents the theoretical foundations that are necessary to understand the specific context in which feedback operates. Thereafter, the subsections outline the research gaps that motivated the papers.

3.1 Environmental behavior, carbon offsetting, and real-time feedback

As Paper IV of this dissertation is concerned with investigating the effects of so-called carbon offsetting programs on resource conservation, this subsection outlines several psychological models explaining environmental behavior. These models provide a theoretical perspective not only on how individuals might respond to programs that make one's consumption carbon-neutral but also on their potential interaction effects with consumption feedback.

3.1.1 Theoretical foundations

There is a long history of theoretical models that try to explain individuals' environmental behaviors. One very prominent example is Schwartz's (1977) norm activation model, which was initially developed to understand altruistic behaviors. Specifically, Schwartz (1977) conjectures that individuals' altruistic behaviors depend on the activation of personal norms, namely feelings of moral obligation to engage in a specific behavior. According to Schwartz, personal norms substantially depend on awareness of the consequences of performing or not performing a given behavior as well as one's associated feelings of responsibility or efficacy in changing the associated outcome. Consequently, the model explains that individuals may not behave altruistically if they are not aware of the outcomes or perceive someone else as being responsible for the outcomes (e.g., policymakers).

As pro-environmental behavior can be seen as a special form of altruism (i.e., others benefit from one's pro-environmental actions), this model has been widely adopted and empirically tested in the environmental domain (see, e.g., Onwezen et al., 2013). For instance, it has been explored whether feelings of responsibility and problem awareness independently activate personal norms or whether one's feeling of responsibility mediates the relationship between problem awareness and personal norms. Taken together, there is strong evidence that supports the latter relationship (de Groot and Steg, 2009; Steg and de Groot, 2010), which is displayed in Figure 3.

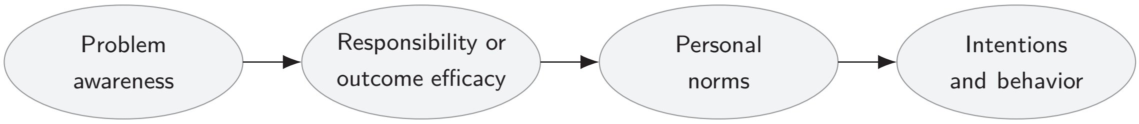


Figure 3: Norm activation model (adapted from Steg and de Groot, 2010)

From a theoretical viewpoint, the outlined relationship of Figure 3 is also plausible: awareness of an environmental problem (e.g., climate change, environmental degradation) might (i) create a sense of responsibility for one's actions and (ii) lead to a judgment of whether one can change the associated outcomes; this in turn might evoke feelings of moral obligation and ultimately shape intentions for pro-environmental behavior.

Apart from the norm activation model, Ajzen's (1991) theory of planned behavior has been widely used in the environmental domain to explain behaviors (see, e.g., Greaves et al., 2013; Kaiser and Gutscher, 2003; Scalco et al., 2017). This theory assumes that a behavior results from individuals' intentions toward it. The stronger one's intentions toward resource conservation, the higher the conservation behavior should be. Importantly, the model describes three relationships on how such intentions form: individuals' attitudes, subjective norms, and behavioral control all directly influence the intention toward a behavior. Attitude refers to an individual's view toward the associated behavior (e.g., favorable or unfavorable view). Knowing about the environmental and societal consequences of a specific behavior (e.g., speeding, excessive air-conditioning use) might contribute to a higher intention toward resource conservation. A subjective norm is the perceived pressure of performing or not performing the related behavior. Such pressure results from expectations of what others would approve or disapprove of. For example, knowing that society disapproves of such behavior, most people might avoid littering. Perceived behavioral control refers to the ability to perform a behavior, which is based on individuals' experiences and anticipated obstacles regarding that behavior. To put this into context, individuals might have high levels of behavioral control to commute via public transportation or to carpool in urban areas but low levels in rural areas where such environmentally friendly transportation modes are less developed. It is important to note that perceived behavioral control is—beyond intention—the only antecedent that is also hypothesized to have a direct influence on behavior (Ajzen, 1991). Figure 4 summarizes the relationships from the theory of planned behavior.

Notably, Ajzen's (1991) theory of planned behavior assumes that differences in personal or sociodemographic variables (e.g., value orientations) are reflected in the outlined antecedents

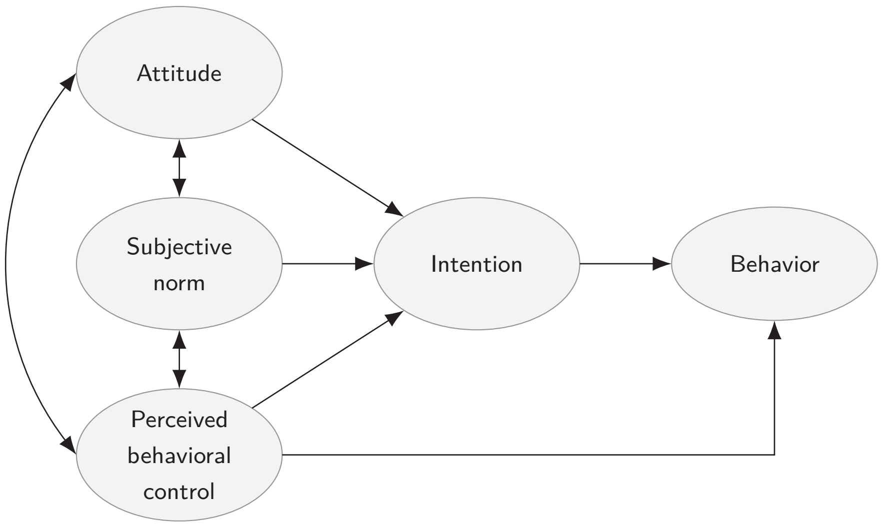


Figure 4: Theory of planned behavior (adapted from Ajzen, 1991)

(Steg and Nordlund, 2019). Following this reasoning, individuals with strong environmental attitudes should have, ceteris paribus, a more favorable attitude toward pro-environmental behavior (i.e., resource conservation). Yet explicitly incorporating such antecedents tends to further increase the prediction power of behavior, as several studies have demonstrated (see, e.g., Bamberg and Möser, 2007; Roos and Hahn, 2019).

The theory of planned behavior provides a plausible explanation of why attitudes alone do not necessarily predict the behavior in question—a phenomenon widely known as the attitude–behavior gap (see, e.g., Boulstridge and Carrigan, 2000; Juvan and Dolnicar, 2014). Especially in the environmental domain, this gap is often present. Many individuals are seemingly concerned about environmental degradation but fail to translate their attitudes or intentions into corresponding actions. A different viewpoint on this topic is provided by Kaiser et al. (2010), who ground their idea in the Campbell paradigm (Campbell, 1963). According to their viewpoint, environmental behavior could be driven by two distinct factors: environmental attitude and behavioral costs. In this context, behavioral costs are related to effort and other personal resources (e.g., time, money) that individuals anticipate or face when they engage in a specific behavior (Kaiser et al., 2010). In contrast to the previously outlined models (e.g., norm activation model), it is important to note that environmental attitude and behavioral costs influence one's environmental behavior independently and complementarily. This means that, for pro-environmental actions with high levels of costs (e.g., avoiding travel by plane when planning to go on a holiday in a distant country), an individual requires higher levels of environmental attitude to sacrifice their personal comfort for actually performing these actions. Vice versa, an individual with low levels of environmental attitude is more likely to show pro-environmental actions when the associated behavioral costs are relatively lower (e.g., reusing shopping bags, turning lights off, recycling). Figure 5 outlines this relationship between attitude and costs for environmental behavior.

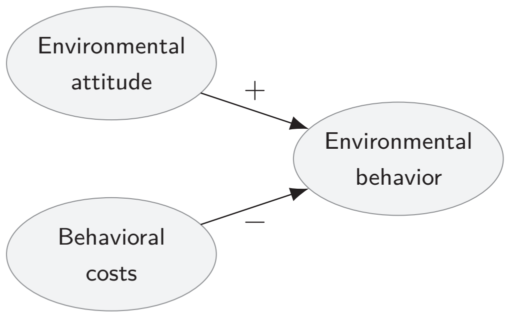


Figure 5: Campbell paradigm (adapted from Kaiser, 2021)

To summarize, the outlined models provide theoretical perspectives on how important factors influence environmental behaviors. All the models are relevant as it is not clear under which circumstances which model is most useful in explaining behavior (for more information on this topic, see Steg and Nordlund, 2019). The associated psychometric factors also provide multiple suggestions on how feedback interventions could empower individuals by facilitating pro-environmental behaviors. Feedback interventions might create a sense of problem awareness as many individuals are not aware of the environmental consequences of their behaviors. By directing individuals' attention toward these consequences and stressing their relevance, individuals might change their attitudes and start conserving resources. Moreover, feedback could increase individuals' perceived behavioral control by supporting them in setting goals and monitoring their goal achievement. Feedback might facilitate this by periodically confronting individuals with the measured outcomes of their last behavioral actions, which they can use over time to test the effectiveness of different behavioral strategies for goal achievement. In doing so, feedback information could also decrease the costs of showing pro-environmental behaviors, since individuals might consequently require fewer cognitive resources to implement behavior change. Following the reasoning of the Campbell paradigm, the reduction of behavioral costs could lead to a situation in which lower levels of environmental attitude are necessary for individuals to ultimately show pro-environmental behaviors (Taube and Vetter, 2019). Lastly, feedback interventions that make subjective norms salient might correct perception errors. For instance, individuals could notice that their resource use substantially exceeds the stated social norm and thus be motivated to subsequently engage in resource conservation.

3.1.2 Research gap

The first research gap, which the dissertation addresses in Paper IV, is concerned with programs that make products and services climate-neutral by offsetting the associated negative externalities through the implementation of green projects elsewhere (e.g., reforestation initia-

tives, energy-efficiency measures; see Gössling et al., 2007). Previous research has documented that such programs can lead to unintended consequences of increasing consumers' demand when consumers decide to pay for the offsetting (Harding and Rapson, 2019; Jacobsen et al., 2012; Kotchen and Moore, 2008). Hence, these programs might reduce individuals' feelings of responsibility or individuals' sense of guilt because the individuals' actions are then seemingly inconsequential for the environment (e.g., see the norm activation model in Figure 3). The dissertation extends prior research by exploring the effects of an offset program when organizations offset, on behalf of consumers, the negative externalities of the consumers' demand (without passing the costs of offsetting to consumers). In addition, the dissertation explores whether consumption feedback can mitigate the resulting adverse effects of the carbon offsetting (i.e., higher consumption levels). These aspects are summarized in the following research question:
Research Question 1: *To what extent does consumption feedback counteract adverse behavioral effects (i.e., higher consumption levels) that might ensue when the CO_2 emissions of products and services are offset by sustainability initiatives?* (Paper IV)

3.2 Consumption feedback in the context of financial incentives and moral appeals

As Paper V of this dissertation investigates the interplay of financial incentives with consumption feedback, this subsection first outlines theoretical foundations regarding the effects associated with financial incentives. Subsequently, it describes the research gaps that Paper V addresses.

3.2.1 Theoretical foundations

3.2.1.1 Standard economic model

In numerous instances, an alternative strategy to feedback information is to provide financial incentives for behavior change. The underlying reasoning is that individuals should respond to an increase in utility associated with the provision of financial incentives by engaging in behavior change. This perspective primarily comes from behavioral economics, a subfield of economics that studies behavioral decisions by using insights from both psychology and economics, with the overarching aim of providing realistic theories describing human behavior (Berg, 2010). In this domain, an important reference point against which individuals' behavior is typically compared is the standard economic model (Dobusch and Kapeller, 2009). This model has a long tradition in research in behavioral economics as it originated from neoclassical economics in the 19th century (Dobusch and Kapeller, 2009; Screpanti and Zamagni, 2005). The standard economic model posits that individuals—also referred to as homo economicus in this context—behave fully rationally and aim to maximize their expected utility (Mullainathan and Thaler, 2000). It is important to note that the notion of expected utility is not bound to economic decisions (such as utility associated with financial incentives) but can relate to any utility that results from individuals' preferences (see, e.g., Doepke and Zilibotti, 2017).

In maximizing expected utility, the standard economic model assumes at least three traits: the homo economicus has unlimited self-interest, has unlimited processing power in optimally solving problems, and possesses unlimited willpower in executing plans (Berg, 2010; Mullainathan and Thaler, 2000). Mathematically, the idea of maximizing expected utility has been formalized in axioms that describe rational behavior (for an overview, see von Neumann and Morgenstern, 2007). For example, the homo economicus is always able to rank different decision options according to their expected utility. Thus, a rational actor is able to discriminate among two decision options so that the actor prefers one option over the other or is indifferent to both. Likewise, if an actor ranks option A over B and B over C, rational behavior implies that the actor also prefers option A over C (von Neumann and Morgenstern, 2007).

Although the standard economic model can be ascribed a sense of elegance by providing a rather generally applicable perspective on behavior, it is clear that individuals often do not behave accordingly. This has spurred a series of criticisms of the standard economic model. In particular, Simon (1955) argues that this discrepancy stems from the fact that individuals are subject to "bounded rationality," which implies that they have limited self-interest, limited processing power in solving problems, and limited willpower in executing plans. As a consequence, a fruitful research stream has emerged in behavioral economics that has extensively tested the limitations of the standard economic model, with far-reaching insights into human decision-making. For instance, individuals tend to excessively discount the utility of future gains (called hyperbolic discounting; see Frederick et al., 2002), rate potential losses higher than equally sized potential gains (as pointed out in prospect theory; see Kahneman and Tversky, 1979), and react differently to an option depending on how it is presented (i.e., the framing effect; see Tversky and Kahneman, 1981).

In the context of this dissertation, these two perspectives on human decision-making are quite relevant to the interaction effects of feedback and financial incentives. Specifically, the standard economic model could predict an increase in the utility of resource conservation when consumption feedback is combined with financial incentives. Following this reasoning, the increase in utility would imply higher conservation effects. Conversely, it could also be that the additional provision of financial incentives does not increase conservation effects, which would provide evidence against the standard economic model. According to bounded rationality, individuals are limited in cognitive resources, and thus consumption feedback might be so salient that the additional provision of incentives does not create more attention for resource conservation.

3.2.1.2 Motivation crowding theory

In addition to the standard economic model and the concept of bounded rationality, there is also a view that financial incentives could lead to unintended effects. Backed by over 50 years of research, many studies have demonstrated that financial incentives could actually decrease behavioral performance. The reasoning is that financial incentives could undermine intrinsic motivation; the motivation to perform a behavior for one's inherent enjoyment rather than for external factors such as monetary rewards (e.g., Deci, 1971). This phenomenon is now referred to as motivation crowding theory, and there are many theoretical explanations for it.

First, individuals might feel controlled by rewards. This could reduce their perceived autonomy level and their intrinsic motivation according to self-determination theory (Deci and Ryan, 2012). Similarly, they may feel that the principal (the entity which gives the reward), does not trust the incentivized individuals to engage in the given behavior without additional incentives. According to reciprocity theory, this could dampen intrinsic motivation (Falk and Fischbacher, 2006). Interestingly, there is also a mechanism related to self-perception theory, which posits that individuals are often not aware of the underlying motives for their behaviors (Bem, 1972). Instead, they deduce the causes of their behavior through an ex-post interpretation of the situation. In the presence of rewards, this interpretation process might change: individuals may develop the feeling that they are doing the incentivized behavior just for the incentives, which could decrease their intrinsic motivation (Tang and Hall, 1995). Likewise, the provision of incentives could direct individuals' attention from the moral to the economical dimension. As a consequence, individuals might wonder whether the incentives are worth the efforts, which could decrease motivation (Ariely et al., 2009). The last argument in this brief overview comes from the image motivation for prosocial behaviors. Individuals tend to seek a positive image from others about themselves. If a behavior is suddenly rewarded, this may cast doubt on individuals' motives for engaging in the incentivized behavior. Subsequently, individuals might change their behavior as they want to avoid appearing greedy (Bénabou and Tirole, 2006).

Though these reasons explain why extrinsic rewards (e.g., incentives) might lead to unintended effects, it is important to note that this effect strongly depends on the types of extrinsic rewards and the incentivized activity. A highly cited meta-analysis from the psychology domain (i.e., Deci et al., 1999) indicates that only tangible contingent rewards (i.e., rewards that are not unconditionally given) should crowd out intrinsic motivation. This should particularly happen in activities that individuals find interesting and have ex-ante a strong intrinsic motivation for (Deci et al., 1999). Gneezy and Rustichini (2000b) provided further interesting insight in this regard: financial incentives that are too low in value are likely to reduce performance, while higher values appear to increase performance. One plausible reason for this is that, depending on the magnitude of the financial incentives, the price effects could outweigh the damage to intrinsic motivation and thereby increase performance.

Hence, the existence of the crowding-out effect is supported by a wide range of experimental studies (Deci et al., 1999; Gneezy et al., 2011; Gneezy and Rustichini, 2000a; Mellström and Johannesson, 2008). Despite this, there is still a debate over whether this effect can be explained by alternative behavioral mechanisms (see Esteves-Sorenson and Broce, 2022). For example, most empirical evidence from psychological studies that initially coined the theory used two experimental phases: a phase with reward and a phase without reward, while the adverse effects (e.g., drop in performance) have often been observed in the second phase (Esteves-Sorenson and Broce, 2022). Instead of damage to intrinsic motivation, an adverse effect on performance could also be attributed to fatigue after showing increased performance levels in the first phase. More precisely, it is quite common that after a performance period (e.g., induced by rewards), a period of fatigue occurs in which performance levels may drop below the baseline before recovery (see, e.g., Sievertsen et al., 2016). Regarding the relevance of motivation crowding theory for this dissertation, Paper V sheds more light on this phenomenon by considering the effects of performance-contingent financial incentives in the presence and absence of consumption feedback.

3.2.2 Research gap

Prior research has indicated that feedback can induce large, cost-effective resource conservation effects (Tiefenbeck et al., 2018), yet it is not clear how the effects of feedback compare to appeals (which pronounce the environmental consequences of one's actions) or to financial incentives (which reward behavior change). Likewise, existing research does not convincingly show to what extent environmental appeals or financial incentives moderate the effects of consumption feedback. From a theoretical perspective, this is also interesting as the theories outlined above provide several contradicting behavioral predictions. For instance, motivation crowding theory predicts that financial incentives might undermine individuals' motivation for resource conservation, which would reduce the effects of feedback on resource conservation. In contrast, if financial incentives induce resource conservation (as the utility for behavior change is then given for a "rational" actor), standard economic theory would predict that financial incentives increase resource conservation effects when combined with consumption feedback.

Another important discussion in the environmental domain is related to the question of how the observed effects at the micro level might extrapolate to the macro level. Specifically, literature provides scarce insights regarding the extent to which individual differences play a role in behavioral responses to interventions. Motivation crowding theory implies that extrinsic rewards, such as financial incentives, should crowd out pro-environmental behavior only from those with an intrinsic motivation for it. In a similar vein, it might be that environmental appeals only work for those individuals with a strong environmental attitude, not for those with a weak environmental attitude.

Against this backdrop, Paper V aims to respond to the outlined questions by presenting the results of a relatively large randomized controlled trial. In doing so, the paper addresses the following research questions:

Research Question 2: *To what extent do the effects of consumption feedback, economically feasible financial incentives, and environmental appeals differ in promoting resource conservation?* (Paper V)

Research Question 3: *What is the interplay of (i) consumption feedback and financial incentives and (ii) consumption feedback and environmental appeals in promoting resource conservation?* (Paper V)

Research Question 4: *Between standard economic theory or motivation crowding theory, which most accurately explains the behavioral response to consumption feedback with financial incentives for resource conservation?* (Paper V)

Research Question 5: *In view of the different attitudes and financial situations of individuals, are non-financial interventions (e.g., feedback, appeals) as effective as financial incentives in promoting resource conservation?* (Paper V)

3.3 Social normative feedback to support self-regulated learning

Papers VI, VII, and VIII address research questions associated with online learning behavior in higher education. Before outlining the research gaps, it is important to first generate a greater understanding of how students in university contexts regulate their learning. To this end, this subsection describes self-regulated learning theory and how the theory relates to procrastination behavior. Procrastination behavior is insofar relevant for the dissertation, as two papers evaluate a feedback intervention inspired by social norms theory to mitigate procrastination behavior. Hence, the following paragraphs describe social norms theory in more detail.

3.3.1 Theoretical foundations

3.3.1.1 Self-regulated learning theory

To begin, it is important to note that learning in higher education is fundamentally different from learning in school environments (Vosniadou, 2020). While higher education students typically have much greater freedom in their choices (e.g., selecting courses based on their interests), this comes with certain costs. Being more anonymous in physical classroom sessions for their instructors, higher education students largely have to regulate their learning processes on their own (Vosniadou, 2020). It is therefore oftentimes more up to the students' initiative to identify and fill knowledge gaps or to seek help. As a consequence, students' need for self-regulation to succeed in higher education (Broadbent and Poon, 2015; Richardson et al., 2012) might be one major reason that many students drop out of study programs. This section therefore sheds more light on underlying psychological mechanisms regarding self-regulated learning and their relationship to procrastination.

Self-regulated learning is a form of learning in which students can actively monitor, control, and regulate certain aspects of themselves (i.e., cognition, motivation, behavior) as well as certain aspects of their environments (Pintrich, 2000). Such self-regulated learning processes take place only when students try to reach or maintain a standard (e.g., a goal). For instance, upon identifying a discrepancy with their goal, students might monitor and regulate their cognition, motivation, and behavior until reaching the goal (Pintrich, 2004). In this context, the regulation of cognition refers to the students' selection and use of rehearsal, elaboration, or organizational strategies to understand the corresponding course materials (Pintrich, 2004). Similarly, the regulation of motivation can consist of various strategies that students might apply to increase their motivational beliefs for learning (e.g., updating task value or task interest beliefs, see Pintrich, 2004). The regulation of behavior refers to the use of time management strategies (i.e., making schedules for learning), effort strategies (i.e., controlling efforts for the respective course), or help-seeking behavior (i.e., seeking help from others) (Pintrich, 2004). Beyond these areas of regulation, students may also adapt their environment to reach their goals. This includes removing potential distractions from their study environment (e.g., turning the TV or the radio off) or structuring the learning environment so that it supports task completion (Pintrich, 2004). It is important to note that this is just a brief overview of potential regulation strategies, which Table 3 summarizes. Pintrich (2004) provides more extensive details on this topic.

Table 3: Brief overview of self-regulated learning areas (adapted from Pintrich, 2004)

Areas for self-regulation	Exemplary activities
Cognition	Setting goals, monitoring of cognition, selection and adaptation of strategies for learning and thinking, and making judgements about one's progress
Motivation/Affect	Making judgements on one's efficacy, assessing the difficulty of tasks, and selecting and adapting strategies to stay motivated
Behavior	Planning effort for a task, deciding to increase or decrease effort depending on one's progress, seeking help from others
Context	Monitoring of the study environment, removal of potential distractions that might interfere with learning

Based on the concept of self-regulated learning, the literature has developed several instruments for identifying psychometric factors related to self-regulation. One of the most influential instruments is the Motivated Strategies for Learning Questionnaire (MSLQ), which was developed by Pintrich et al. (1991) to understand the self-regulated learning behavior of college students. This is now an internationally used instrument to explain self-regulated learning across various contexts (e.g., for online learning or blended learning, which is a combination of remote computer-mediated content with physical classroom sessions; see Broadbent, 2017; Cho and Summers, 2012) and samples (e.g., undergraduate students, graduate students; see Duncan and McKeachie, 2005). Notably, this instrument has also sparked further developments for slightly different contexts (e.g., scales to determine self-regulated learning adapted to higher education in Germany; see Wild and Schiefele, 1994).

Before presenting the fundamental factors of the MSLQ to generate a more profound understanding of self-regulated learning, it is important to note how underlying psychometric factors relate to behavior (i.e., regulation of cognition, motivation, behavior, and environment). In general, self-regulated learning behavior is seen as a mediator between personal or contextual characteristics and the associated learning outcomes. In other words, while personal or contextual characteristics have direct effects on learning outcomes, self-regulation mediates these relationships (Pintrich, 2004). This can explain why individuals with relatively high self-regulated skills may fail to succeed in learning if they do not activate their self-regulation strongly enough.

In the MSLQ, psychometric factors are divided into two overarching categories that measure a learner's traits and perceptions of a specific course. The first category covers motivation scales that capture value orientations (e.g., intrinsic or extrinsic value), control and self-efficacy beliefs, and test anxiety (Pintrich et al., 1991). The second category is concerned with cognitive and metacognitive strategies as well as resource management strategies. Cognitive and metacognitive strategies comprise scales that cover self-regulated activities such as rehearsal, elaboration, or organization (Pintrich et al., 1991). All these scales therefore aim to measure how students' interact with course materials. In contrast, resource management strategies aim to measure how well students' regulate their study and time management and how they make use of their environments (e.g., seeking help, learning with peers) (Pintrich et al., 1991). For an extensive overview of the individual scales, refer to Pintrich et al. (1991).

While the MSLQ is widely used, a meta-analysis with data from 19,900 students (Credé and Phillips, 2011) provided evidence that most of its 15 associated constructs have almost no predictive power over students' performance. More precisely, only three scales had a moderate to strong relationship with grades (i.e., the scales of self-efficacy, effort regulation, and time and study environment). The literature has identified several aspects as underlying reasons for this. One reason is that self-regulation scales are based on potentially subjective self-reports that are not immediately captured while learning. Instead, students rely on their experiences (i.e., memories) to answer the items of the questionnaire, which could be more inaccurate due to the temporal distance to their learning (Klingsieck, 2018). In addition, the application of specific learning strategies also depends on the individual task (Credé and Phillips, 2011). These conditional mechanisms, however, are not captured by the MSLQ.

Before considering the phenomenon of procrastination in students, it is important to note that the literature on self-regulation provides several process-oriented models that describe how students temporally engage in self-regulation. Pintrich's (2004) process model is described below, as it also sheds light on relationships to learner characteristics, which are quantified by questionnaires such as the MSLQ (for an overview of different models, see Panadero, 2017). Figure 6 displays the different phases of self-regulated learning.

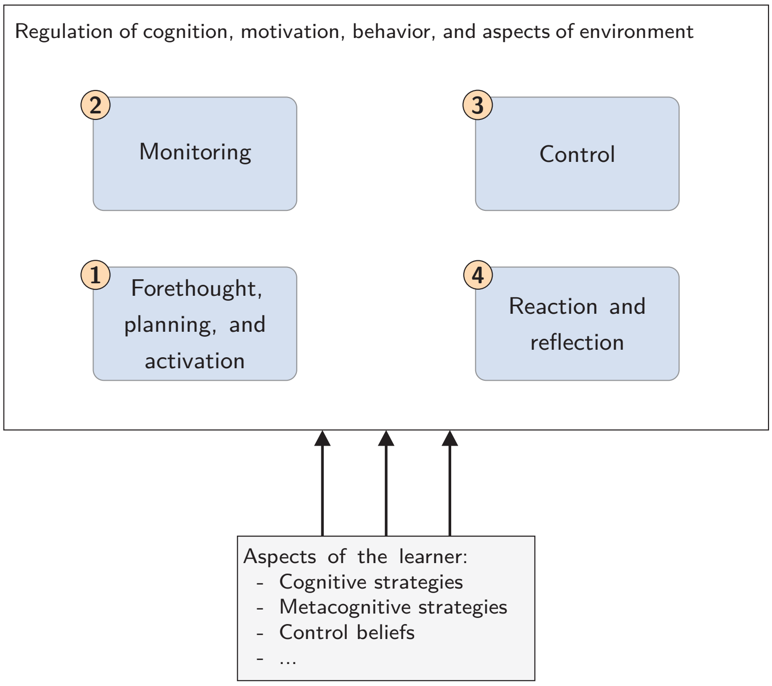


Figure 6: Phases of self-regulated learning and the role of learner characteristics

As depicted in Figure 6, the self-regulated learning process can be separated into four distinct phases in which the regulation of cognition, motivation, behavior, and aspects of the environment take place. These phases are chronologically ordered; however, in practice, students do not have to successively follow all these steps (Pintrich, 2004). In the first phase (1), students usually set learning goals and activate prior knowledge on a topic. This goes along with motivational judgements and perceptions of the task (e.g., task difficulty or value). In the second phase (2), students use various metacognitive processes to monitor themselves, the task, or the context (Pintrich, 2004). For example, students monitor their efforts and determine whether they may need help. The third phase (3) is concerned with controlling one's learning process; this covers the selection and adaptation of strategies that support their learning. Apart from that, students might increase their efforts or seek help from peers. In the fourth phase (4), students make judgements about their learning (Pintrich, 2004). More precisely, students might judge how well they did and attribute this to aspects of their self-regulation (e.g., specific behaviors or cognitive strategies). As self-regulation can be seen as an iterative process, the outcomes of the fourth phase might substantially influence the next forethought phase (i.e., the first phase). Ultimately, it is important to note that, in all four phases, learner characteristics (e.g., from MSLQ) might have a substantial influence on self-regulation (Pintrich, 2004). For example, students with higher levels of metacognition might regulate their cognition (e.g., the selection of learning strategies) more effectively in the third phase, which could in turn improve associated learning outcomes.

In summary, self-regulated learning theory provides an overarching understanding of which factors are important for students to succeed in higher education. A related topic, that this dissertation with Papers VI and VII addresses, is procrastination. Procrastination, which can be defined as voluntarily delaying one's tasks despite the negative consequences of the delay (Steel, 2007), is considered a great problem today, especially in education. In fact, studies have indicated that almost half of students procrastinate to a problematic degree (see, e.g., Day et al., 2000; Özer et al., 2009; Solomon and Rothblum, 1984), often causing severe negative effects on their academic performance, as related meta-analyses demonstrate (Kim and Seo, 2015; Steel, 2007). This can have further negative consequences not only for themselves (e.g., on students' well-being or financial situation in case of a longer enrollment in study programs; see Patrzek et al., 2012; Sirois et al., 2003) but also for society (e.g., delayed entry into the job market).

There are multiple underlying reasons why students procrastinate as well as multiple factors that predict procrastination. A common mechanism in human decision-making is the tendency for individuals to be motivated by smaller, immediate rewards than by more temporally distant, larger rewards (Frederick et al., 2002; Steel and König, 2006). This can explain why, for example, students delay coursework and instead enjoy leisure time. The closer a deadline comes, the higher the perceived value of previously procrastinated coursework might become.

Procrastination is also clearly linked to several factors of self-regulated learning (see, e.g., Credé and Phillips, 2011; Pintrich et al., 1991; Steel, 2007). First, procrastination is associated with feelings of self-efficacy (Steel, 2007), which are based on one's beliefs of being able to successfully complete a task. To put this into perspective, individuals may not activate their self-regulation simply because of the feeling of having little to no control over a given situation. Similarly, students are more likely to procrastinate on a specific task if they have low levels of motivation for it. Motivation, in this context, covers both intrinsic and extrinsic elements associated with achievement (Steel, 2007); this means students with strong intrinsic motives for a task (e.g., they like performing well regardless of external rewards) are less likely to procrastinate. Interestingly, the same relationship applies to scenarios in which students have strong extrinsic motives. When students are keen to receive an external reward, they could be less likely to procrastinate (Tisocco and Liporace, 2023). There are two further predictors of procrastination which are clearly linked to self-regulated learning. The first predictor is organization, which is one's capabilities in planning or structuring their life, and it is antithetical to procrastination (Steel, 2007). A plausible reason for this is that students' organizational capabilities might also induce goal setting and reflection on one's progress (Steel, 2007). The second predictor of procrastination is distractibility: an explanation for this is that individuals might pay less attention to self-control when they have low capabilities for managing distracting cues (e.g., by ignoring or eliminating them) (Steel, 2007). In terms of similarities between these two predictors of procrastination and self-regulated learning, organization is important in self-regulated learning for selecting strategies that help to manage one's progress and to effectively work through learning materials (e.g., Pintrich et al., 1991), and distractibility is also somewhat related to one's skills in regulating aspects of their environment to learn effectively (e.g., Pintrich, 2004)

In sum, there are many factors that influence whether and to what extent individuals procrastinate. This brief overview of associated behavioral mechanisms related to procrastination is not intended to be exhaustive but nonetheless provides a tangible link to self-regulated learning. Before moving on to the next section, it is important to note that some research has also distinguished between active and passive procrastination; the former refers to the previously outlined concept of self-regulation failure and the latter to some potentially beneficial delays of action (Chun Chu and Choi, 2005; Seo, 2012; Yamada et al., 2016). Subsequent research has critically revisited the concept of active procrastination and argued that it is inaccurate (see, e.g., Chowdhury and Pychyl, 2018; Hensley, 2014). Such behavior can be seen as an arousal delay, whereby individuals may seek to attain the excitement or thrill of working on a task right before the deadline (Chowdhury and Pychyl, 2018). Conversely, such behavior can also be seen as a deliberate action in reprioritizing one's tasks, which is the main reason active procrastination has been discussed as beneficial to learners (Chowdhury and Pychyl, 2018). In light of this, the phenomenon of active procrastination for purposefully delaying actions (and not for arousal delay) could therefore be seen as a signal of strong self-regulation but not of underlying procrastination behavior. Hence, this dissertation views procrastination as a sort of harmful failure of self-regulation.

3.3.1.2 Social norms theory

To mitigate procrastination in online learning, Papers VI and VII test the effects of an intervention that makes social norms salient. Social norms are unwritten rules about what behavior or attitude is acceptable within a reference group (e.g., society or a social group) (Cialdini et al., 1990), similar to the subjective norm construct from Ajzen's theory of planned behavior. Social norms theory postulates that individuals seek to conform to these norms, which can have adverse consequences if the norms are based on erroneous beliefs (Berkowitz, 2005). Such an erroneous belief for an individual could be that others show much higher levels of problematic behavior (e.g., littering) even when this is not the case. As a consequence, the individual might feel entitled to approximate their behavior toward the believed social norm, which could increase the individual's problematic behavior. Motivated by this, numerous studies have made social norms salient in different areas to potentially correct erroneous normative beliefs and induce societally desirable behavior (e.g., for conserving electricity or reducing binge-drinking among students; see Agostinelli et al., 1995; Allcott, 2011; Loock et al., 2011).

In general, literature distinguishes between two types of social norms. A descriptive norm refers to the perception of what most others commonly do, while an injunctive norm refers to the perception of what others approve or disapprove of (Cialdini et al., 1990). There is evidence that both types of social norms can play a decisive role in inducing behavior change. Specifically, Schultz et al. (2007) report evidence from a field experiment on electricity conservation that salient descriptive norms could have heterogeneous effects. Though participating households with a high consumption level responded to the social normative feedback intervention with electricity savings, the group of low-consuming households surprisingly responded by increasing their consumption levels, which Schultz et al. (2007) coined the "boomerang effect." It is possible that, before the feedback intervention, both subgroups had different beliefs regarding the social norm: residents of high-consuming households potentially overestimated the social norm, while residents of low-consuming households underestimated the social norm. The feedback intervention might have corrected these erroneous normative beliefs, leading to different behavioral responses in both subgroups. Interestingly, Schultz et al. (2007) also reported the impact of a feedback intervention that additionally comprised an injunctive norm on a different set of comparable households. The results were similar, with the exception that no boomerang effect was observed: the subgroup with low consumption levels did not increase their consumption in response to the intervention (although the subgroup also did not conserve electricity). Thus, Schultz et al. (2007) provide a reasonable explanation for why previous experimental studies on social norms have led to mixed results (see, e.g., Agostinelli et al., 1995; Werch et al., 2000).

To summarize, the related literature indicates that feedback interventions that make social norms salient can have powerful behavioral effects. At the same time, to avoid adverse effects for some subgroups, it might be useful to combine descriptive and injunctive norms.

3.3.2 Research gap

Papers VI and VII address two research questions related to online learning behavior and feedback visualizing social norms.

Previous studies suggest that comparative feedback on online learning behavior has mixed effects (e.g., Davis et al., 2017; Guerra et al., 2016; Li and Zhang, 2016), meaning that they can even harm learning by discouraging online activity. According to social norms theory, it might be that comparative feedback corrects erroneous normative beliefs. If an intervention reveals that others are not as active as the feedback recipients previously believed (by making descriptive norms salient), the intervention could actually decrease extrinsic motivation for learning (as suggested by Li and Zhang, 2016). Thus, this dissertation tests the effects of a comparative feedback intervention that combines both descriptive and injunctive norms to provide feedback on online learning behavior. The injunctive norm is included to prevent the descriptive norm from backfiring (see Schultz et al., 2007). Similarly, the feedback intervention deliberately provides feedback on online learning time as it is a tangible metric that many students can easily change. Prior research in that area indicates that comparative feedback that visualizes the performance of high-achieving students might backfire (see, e.g., Rogers and Feller, 2016). The dissertation also explores the temporal effects of this feedback intervention to create a better understanding of whether the intervention might mitigate procrastination. In this case, the intervention would increase online learning in the middle of the course, in which procrastination behavior is typically reported (see, e.g., Gafni and Filin, 2015; Geri et al., 2014) and decrease online learning right before the exam.

A second research gap relates to comparing the effects of this feedback intervention across different university courses (a bachelor's and a master's level course) as related studies on the generalizability of effects are scarce (Sønderlund et al., 2019). Such a comparison might help to estimate to what extent the effects of the feedback intervention could be extrapolated to other scenarios (e.g., samples of students with other socio-demographic backgrounds, courses with different types of learning materials).

Hence, the aspects that Papers VI and VII address are summarized by the following research questions:

Research Question 6: *Does social normative feedback mitigate the procrastination of university students in digital learning environments?* (Papers VI and VII)

Research Question 7: *Does social normative feedback lead to heterogeneous behavioral responses among users (i.e., course participants) of a digital learning environment?* (Paper VII)

3.4 Addressing the heterogeneity of learners in feedback provision

3.4.1 Theoretical foundations

As outlined in Section 3.3, self-regulated learning theory provides an overarching perspective on how students in higher education regulate their learning. The literature describes that self-regulation inherently depends on learners' individual characteristics (Pintrich, 2004). For instance, while learners with high levels of metacognitive skills might be able to successfully plan their learning and monitor their progress toward their learning goals, learners with low levels might fail to do so. Likewise, self-efficacy beliefs can have a decisive effect on learning behavior. When learners have low beliefs about being able to understand the learning materials, they may start procrastinating (see, e.g., Steel, 2007). In contrast, learners with higher self-efficacy beliefs might rather sustain their self-regulation processes to reach their goals.

3.4.2 Research gap

In studies that test feedback interventions in digital learning environments, approaches for feedback provision often neglect the heterogeneity of learners regarding their personalities and self-regulated skills. For example, although an intervention that provides salient social norms on online learning time might be generally helpful to many learners, other learners might fail to translate the feedback of increasing one's online learning time into meaningful learning actions. To address the heterogeneity of learners, feedback interventions need to acknowledge the differences among learners when providing feedback information. However, the question remains as to how such feedback information can be generated by digital learning environments.

Paper VIII responds to this question by suggesting the use of counterfactual explanations for providing personalized feedback at scale. The paper presents a concrete feedback design using this approach and outlines ongoing experiments to test its effects on learners. The preliminary work presented in Paper VIII is directed toward answering the following research question:
Research Question 8: *To what extent does feedback that presents counterfactual explanations in digital learning environments improve the learning outcomes of university students, and does such feedback address the heterogeneity of learners in feedback provision?*

3.5 Quantifying the cause-effect relationships of interventions on handwashing

Apart from environmental sustainability and online learning in higher education, the dissertation is also concerned with feedback interventions regarding health behavior (i.e., handwashing behavior). As this topic takes a subordinate role in this dissertation (in the form of initial work to improve and understand hand hygiene behavior, covered by Paper IX), this subsection provides only a brief overview of a few related theoretical relationships before outlining the associated research gap.

3.5.1 Theoretical foundations

Several behavioral beliefs have been identified as having influence on health behaviors. For instance, the knowledge about one's vulnerability to a health issue, the severity of associated negative health outcomes, and the perceived benefits and barriers of behavioral change are important antecedents of health behavior. More precisely, the health belief model (Rosenstock, 1974) posits that upon receiving an internal or external cue on a specific behavior, an evaluation ensues (e.g., comparing risks and barriers to benefits) that determines the extent to which an individual engages in health-promoting behavior. A comparable perspective comes from the two main antecedents of social cognitive theory (Bandura, 1997). More specifically, an individual's outcome expectation must be favorable for behavior change to occur, which means that the anticipated positive outcomes of performing a behavior must outweigh the negative outcomes (Fishbein and Yzer, 2003). Likewise, the likelihood of engaging in the health-promoting behaviors is also affected by efficacy beliefs. For instance, if an individual holds weak beliefs of being able to improve their associated health behavior (despite objectively favorable outcome expectations), the individual is less likely to respond. Similar implications result from protection motivation theory (Rogers, 1975), which was developed to explain the behavioral responses to fear appeals. It is particularly noteworthy that antecedents such as vulnerability, severity, efficacy beliefs, or response costs—which are related to the theoretical health models mentioned before—are also relevant here.

Apart from these beliefs, normative beliefs also explain health behavior (see, e.g., the application of the theory of planned behavior for health behavior in Conner et al., 2002). A subjective norm might serve as a reference point (and as normative pressure) on whether a specific behavior should be performed or not. Additionally, research has provided theories that describe temporal stages that one transcends to improve health behavior. For example, smoking is typically not a behavior that is immediately stopped (see, e.g., DiClemente et al., 1991). Instead, one usually goes through multiple stages of change, such as identifying and acknowledging that smoking could be a problem, subsequently taking action against it, and then to maintaining efforts to avoid smoking. Thus, to be (most) effective in supporting behavior change, interventions should be designed for the specific stage an individual is in. Both the transtheoretical model and the precaution adoption process model are based on this principle (for a description of the models, see Prochaska and Velicer, 1997; Weinstein, 1988).

3.5.2 Research gap

As outlined, health behavior depends on various antecedents of behaviors, such as knowledge of one's vulnerability and the severity of negative outcomes. Feedback interventions typically could have an influence on several antecedents. For example, feedback could direct attention toward the respective activity, increase individuals' behavioral control, or make social norms salient. Yet the literature has not provided an experiment that quantifies the extent to which individual antecedents of low-involvement everyday behaviors are influenced by interventions. Against this backdrop, Paper IX proposes a field experiment to quantify the effects of (feedback) interventions on related antecedents of handwashing behavior and, ultimately, on behavior change. Importantly, such an experiment might not only be useful for feedback interventions on handwashing behavior but could also serve as a blueprint for testing such effects on other everyday behaviors that share similar task characteristics (e.g., routine behavior with potentially low costs of behavior change). Based on the above, the preliminary work presented in Paper IX is directed toward answering the following research question:

Research Question 9: *To what extent do feedback information and salient social norms influence the antecedents of handwashing behavior and thereby promote behavior change?* (Paper IX)

4 Methodology

The dissertation uses numerous methods to enable feedback interventions and to close the outlined research gaps. To provide a holistic insight into the methodology, this section describes the research approach. After this, it outlines the main motives of this dissertation to focus on field research. Then, it presents the development and adaptation of IT tools (i.e., the macro-micro transition) that served as the basis for the employed research methods. Subsequently, the section provides insights into the research methods, together with a description of the data collection techniques used. Table 4 provides an overview of the research methodology.

Table 4: Research methodology used in this dissertation

Paper	Research context	Research method	Data collection techniques
I	Environmental behavior	Predictive analytics	Measurement data & logbook
II	Environmental behavior	Predictive analytics	Measurement data & logbook
III	Health behavior	Literature review	Past literature
IV	Environmental behavior	Field experiment	Measurement data
V	Environmental behavior	Field experiment	Measurement data & survey data
VI	Learning behavior	Field experiment	Measurement data & survey data
VII	Learning behavior	Field experiment	Measurement data & survey data
VIII	Learning behavior	Field experiment (conceptualized)	Measurement data & survey data
IX	Health behavior	Field experiment (conceptualized)	Measurement data & survey data

4.1 Research approach

With the exception of Paper III, all papers in this dissertation use a quantitative approach to answer the underlying research questions. A quantitative approach focuses on types of data that are expressed in numerical values. Numerical values are used in corresponding research methods to (i) represent theoretical constructs and (ii) generate scientific evidence on theoretical questions (Chen and Hirschheim, 2004; Recker, 2021). In the context of this research, the papers of this dissertation capture (or aim at capturing) numerical measurements that are associated with several behaviors. A significant change to the measurement data (e.g., in response to an intervention) is seen as evidence for an associated research question. The reasoning behind the focus on quantitative research is that the dissertation aims to change behavioral outcomes (e.g., energy use, learning time) or empower IT systems to change individuals' behavior. These behavioral outcomes are relatively easily quantifiable in the form of tangible numbers (e.g., energy in kWh, learning time in minutes) due to the ubiquity of IT.

An alternative to a quantitative approach is a qualitative approach, which relies on the analysis of non-numerical, narrative data (e.g., text) to provide a description and understanding of factors behind a phenomenon with the aim of theory development (Chen and Hirschheim, 2004; Recker, 2021). However, interviewing individuals about whether a behavioral intervention has led to a desirable outcome may lead to systematic biases as individuals often fail to estimate associated behavioral outcomes correctly (see, e.g., Tiefenbeck et al., 2018; Truelove et al., 2014). In this dissertation, quantitative approaches are used as they have the clear advantage of translating the effects of interventions into tangible numbers (e.g., kWh of energy saved per day), which then serve as empirical facts to inform potential stakeholders about the cost-effectiveness of behavioral interventions and to test theoretical predictions. However, one paper applies a qualitative approach to identifying barriers that hand hygiene monitoring systems face, with the goal of informing the development of new systems.

It is also possible to combine quantitative and qualitative research methods to leverage the advantages of both approaches (see, e.g., Venkatesh et al., 2013). As the papers do not combine quantitative and qualitative methods, the mixed-methods approach is not described any further.

4.2 Reasons for field research

Before delving into the employed research methods, it is important to further clarify a core aspect of this dissertation. As indicated, most research for this dissertation takes place in the field, which is clearly in stark contrast to research in more controlled environments (e.g., laboratory environments, online experiments; see Section 4.4.3). While both types of research have their advantages, this dissertation's focus on the field is driven by four main reasons.

First, most papers in the dissertation seek to investigate how feedback interventions affect behaviors, wherein the corresponding decision-making processes often operate automatically and quickly (i.e., System 1 thinking; see Section 2.1). Here, the corresponding interventions intend to override automatic processes that otherwise could lead to suboptimal outcomes for the individuals (e.g., by not paying attention to the behavioral consequences). Exploring these phenomena in more controlled environments than in the field (e.g., in a laboratory) is extremely challenging. Individuals might feel more observed, which in turn could activate System 2 thinking, leading to deliberate actions (see, e.g., Hagel et al., 2015). Consequently, this would not yield a credible estimate of how interventions affect automated decision-making processes (i.e., System 1 thinking). To mitigate such potential sources of bias, field research is key.

Second, as the dissertation seeks to derive meaningful knowledge of the effects of interventions on behavior, field research appears to be indispensable. An alternative way to determine the effects of interventions could be to just ask individuals about their opinions of interventions. However, studies suggest that individuals' beliefs do not appear to substantially predict how individuals will ultimately react to interventions (Nolan et al., 2008; Schultz et al., 2015). A plausible reason for this is that there are nonconscious influences on behavior that individuals do not consider relevant for their decision-making (see, e.g., Bargh, 2006; Nisbett and Wilson, 1977). Nonetheless, there is still the possibility to explore the behavioral responses to interventions in a more controlled environment than in the field. However, individuals might consequently feel more observed, which may cast doubt on the associated findings. In particular, research indicates that individuals have the desire to convey a positive self-image to others (see, e.g., Ariely et al., 2009; Vesely and Klöckner, 2020). As such, feeling more observed might result in responding in a more socially acceptable way (e.g., overstating one's efforts for pro-environmental behavior). Due to the ubiquity of IT, researchers can explore behaviors in the field without giving individuals the feeling of being observed (see, e.g., Hagel et al., 2015), because people are now used to being surrounded by IT devices that measure and quantify behavioral outcomes (e.g., resource usage, physical activity).

Third, the dissertation deliberately revisits important theories and models. Many theoretical implications related to behavioral interventions have been produced in controlled environments (e.g., motivation crowding theory; see Esteves-Sorenson and Broce, 2022) or are based on correlational evidence from meta-reviews (e.g., the first two propositions of feedback intervention theory; see Kluger and DeNisi, 1996). As a consequence, it is unclear whether the associated implications might play a role in real-world contexts. By testing theoretical implications for a specific context in the field, the dissertation can therefore increase understanding of the generalizability of the associated theoretical findings.

Fourth, as this dissertation aims to reduce personal or societal issues, it is essential to study the interventions over time. More specifically, to be feasible instruments for policymakers and organizations, interventions have to amortize. Thus, the effect sizes of the interventions have to be assessed. Regarding sustainability initiatives, amortization means that their effects on conservation behavior need to be stable over a longer period of time until the accumulated savings exceed the initial investment costs (e.g., in greenhouse gas emissions, monetary costs, etc.). Apart from amortization, observing behavioral responses over time is often an integral part of determining the success of behavioral interventions on intertemporal choices (e.g., procrastination behavior, health behavior). For instance, an intervention that is specifically designed to mitigate procrastination behavior should show its effects only when individuals actually start procrastinating. However, testing such temporal effects in more controlled environments, such as a laboratory, is a difficult task as study participants must be willing to spend a longer time in a rather artificial environment. Thus, the strengths of field research come into play as participants can stay in a setting natural to them, which in turn empowers relatively lengthy experiments without imposing much burden on the participants. At the same time, longer experiments also allow exploration of whether the technical setup that enables interventions is feasible to run correctly under realistic conditions.

Yet carrying out field research comes with a substantial drawback for researchers. To enable interventions and measure their effects, field research requires developing and setting up real-world test environments. In the context of this dissertation, this comprises the development of software solutions and the adaptation of hardware components. In the associated studies on environmental or health behavior, preparatory work involves setting up hardware devices at the place of action (e.g., in participants' showers, at faucets in public bathrooms). In the studies on online learning behavior, this covers the adaptation of a digital learning environment to enable the provision of certain feedback information. Moreover, across the field studies of this dissertation, a set of software tools were developed that monitor whether there are potential irregularities that could put the respective field study at risk.

For all papers in this dissertation, substantial preparatory work was necessary to enable field research, which is presented in the subsequent section. It is important to note that the amount of preparatory work differs among both chapters that follow the introductory paper. As Chapter 1 (Papers I to III) focuses on developing feedback interventions, and therefore the preparatory work only covers the installation of hardware devices to measure behavioral outcomes. The preparatory work of Chapter 2 (Papers IV to IX) also comprises the technical implementation of the behavioral interventions and the software tools to monitor the respective field study.

4.3 Implementation and adaptation of technical tools that enable field research

As indicated, the papers in this dissertation would not have been possible without adapting and developing IT-based tools that allow for data collection. To provide a more in-depth understanding of the research context, this section describes what has been done to empower each paper. Table 5 provides a brief overview of the different IT tools and the adjustments made.

Before outlining the details of the IT tools, it is important to mention that questions related to the design of IT systems for solving real-world problems have always been a core topic of IS research. This has manifested in research streams such as design science research (see, e.g., Hevner et al., 2004; Peffers et al., 2018). The German IS community in particular has a long history of studying design-oriented questions with the aim of achieving practical utility (Winter, 2008). What is common around these streams is that the design of IT artifacts (i.e., systems) plays a key role in underlying research, while the behaviorist approach focuses on "observing IS characteristics and user behavior" (Österle et al., 2011, p. 7). The dissertation shares methods from both streams as (i) technical tools are created and adapted (which could arguably be seen as the instantiation of IT artifacts) and (ii) the behavior of individuals is studied. Nevertheless, the dissertation follows the behaviorist approach as the guiding research paradigm because its

Table 5: Overview of the developed and adapted IT tools

IT tool	Activities to support the studies	Publication
Development of a household water monitoring system	Setup of ultrasonic sensors at the main water pipe of each recruited household to measure the water consumption.	Paper I
Extension of a shower metering device	Extension of existing hardware and software for subsequent studies on hot water use in the shower: – Development of a Wi-Fi gateway that can access the measurement data of a shower meter via Bluetooth and send the data to a server, which can then trigger subsequent interventions. – Development of a dashboard that visualizes the status of gateways across different locations and buildings. – Development of a mobile phone app that supports the installation of shower meters and gateways.	Papers II, IV, V
Extension of a touchless faucet	Development of additional functionalities: – Implementation of a gateway that transmits the measurement data of a faucet to a server. – Extension of the gateway to handle external devices (e.g., displays) for interventions.	Papers III and IX
Extension of a digital learning platform	Development of additional functionalities: – Adaptation of the registration page to include questions on personality variables and learning behavior. – Integration of a time-tracking functionality that tracks the active learning time of users. – Implementation and integration of two software components that can provide feedback to users.	Papers VI, VII, VIII

research questions are predominantly directed towards specific individual behaviors, with little focus on the underlying IT characteristics. Turning again to the associated IT tools of this dissertation, the following paragraphs shed more light on the developments made.

- **Development of a household water monitoring system:** As Paper I focuses on the conceptualization of an IT system for feedback on water consumption, it uses preliminary hardware to check the feasibility of an empirical approach. In this study, an ultrasonic sensor was attached to the main water pipe of each recruited household. This sensor measures the water consumption of a household at a resolution of 0.5 Hertz and stores each data point together with a time stamp on a persistent storage device.

- **Extension of a shower metering device:** To study environmental behavior, Papers II, IV, and V all use the shower meter "amphiro b1," which is displayed in Figure 7. The meter itself is a small device that can be installed between a shower hose and a hand-held shower head. It measures the water consumption and the temperature during each shower taken and can display this information on its small, configurable LC display. It draws the required energy for operation from the water flow and therefore requires no battery. Yet, due to the absence of a battery, the device has limitations in terms of data communication. For instance, it is only possible to access the measurements of the device via Bluetooth while it is active (up to a maximum of three minutes after an interruption to the water flow).

Figure 7: The shower meter amphiro b1 (Credit: Amphiro AG)

For all three papers, shower meters have been adapted to display different information depending on the respective study. Beyond that, Papers II and V build on substantial extensions of the shower meter. More specifically, a gateway (based on Raspberry Pi Zero W) has been developed that can capture the (live) measurements of the shower meter via Bluetooth and transmit the measurement data to a server, which then stores the measurements in its databases. This extension allows for a set of display-independent behavioral interventions that were not feasible before, such as sending text messages to study participants and offering them performance-contingent financial rewards for behavior change.

As the Bluetooth connection is only feasible for a few shower meters and, therefore, in the worst case, each shower meter requires one gateway for data transmission to the server, a set of monitoring tools was created to handle larger studies reliably. The monitoring tools consist of (i) a dashboard that displays the operational status of each gateway, (ii) a programming library that reports application errors from gateways and the server, and (iii) a reverse secure shell tunnel that allows updating of the gateways remotely. Hence, it was possible to manage large numbers of gateways (i.e., up to 200 gateways in the field) with relatively little effort. These tools enabled Papers II and V. Moreover, in Paper V, a mobile app was developed to support the installation of shower meters across different apartments and buildings. This was done to avoid experimental errors by making consistency checks when setting up a specific field experiment across different locations (i.e., by allowing only the assignment of a correctly pre-configured shower meter to a certain study participant). Figure 8 provides a visual of the dashboard and the mobile app.

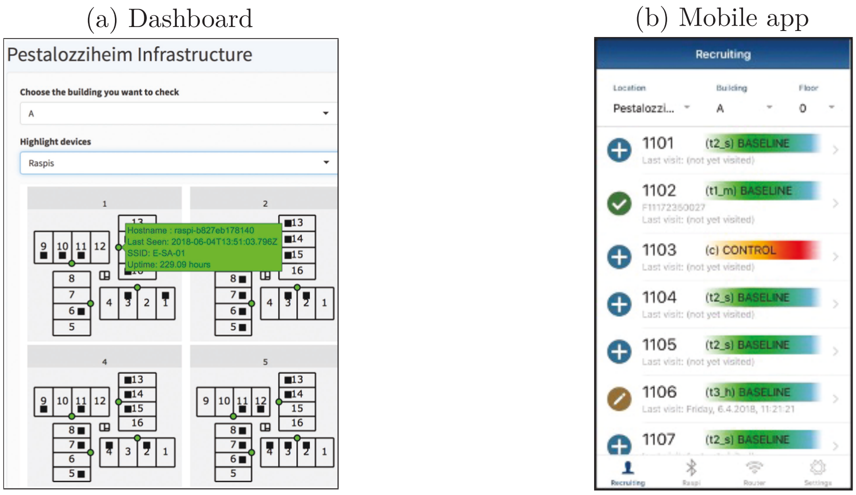


Figure 8: Exemplary tools that facilitated research with the shower meter

- **Extension of a touchless faucet:** To improve handwashing behavior, Paper III presents a recent prototype from the collaboration between Amphiro AG and Oras Oy, which is referred to in Paper IX. The prototype is an automatic touchless faucet that harvests energy for its operation from water flow and stores the energy in its battery. Beyond this, the prototype features a Bluetooth functionality to access the measurement data, which is similar to that of the already-described shower meter. To make use of the measurement data, Paper IX relies on an extension of the prototype, which is displayed in Figure 9.

Figure 9: Overview of the touchless faucet with internet connectivity (adapted from Stingl et al., 2021; Credit: Amphiro AG)

Specifically, an LTE gateway has been developed (see ① in Figure 9) that can (i) transmit the data of a faucet (②) to a server and (ii) handle study-related technical devices near the faucet. The gateway is conceptualized to show feedback on handwashing performance (i.e., appropriate duration of soap usage, etc.) by collecting data from the soap sensor (③) and handling an external display (④) that is usually mounted on the wall above a faucet. Though both papers demonstrate conceptualizations of the adapted system (and do not include statistical analyses on the effects of feedback), it is notable that the extended prototype is completely operational and used in studies (see, e.g., Dangis et al., 2023).

- **Extension of a digital learning platform:** As an experimental platform to study learning behavior, edX is used in Papers VI and VII, and VIII. edX is a free and open-source platform that enables educators to set up and run distance learning. In the context of the papers, the platform was used for two university courses. While the platform itself is completely operational, three main adjustments have been made to it to study learning behavior. First, as the platform has programming interfaces that enable modular extensions, the registration page of the platform has been adapted to include questions on personality variables and learning behavior, as well as the possibility for students to opt out of research that uses data on their learning behavior and exam performance. Second, a tracking functionality has been included in the learning platform to measure the active learning time a user spends on each page. In the event that (i) a user is inactive for a certain time (i.e., no mouse movement, clicks, or keyboard input events), (ii) a user changes the active tab to a different page, (iii) or a user minimizes the web browser, the timer of the tracking functionality stops. In case the user continues with an activity on the page, the timer continues measuring the time. As a consequence, the tracking functionality should capture the online learning activities of the users reasonably well. Third, a functionality has been implemented that allows embedding components into course overview pages that provide users with feedback on their online learning behavior. Thus, this functionality is aware of experimental parameters. This means that the functionality contains a method that checks for a specific course (i) whether it should display feedback and (ii) whether a specific user should receive feedback for a specific course. Hence, the feedback functionality supports advanced experimental designs (e.g., the difference-in-differences approach; see Section 4.4.3.2).

4.4 Research methods

4.4.1 Predictive analytics

As two papers of this dissertation (Papers I and II) investigate whether certain consumption events can be reliably predicted by related measurement data, the dissertation applies predictive analytics as one of its underlying research methods. Predictive analytics is an approach to generating empirical predictions by using statistical methods and subsequently evaluating the quality of these predictions (Shmueli and Koppius, 2011). In general, these models have no focus on explainability and causal interference but rather intend to accurately predict a dependent variable (Shmueli, 2010).

Within IS research, predictive analytics can take on a set of different roles, such as generating new theories, comparing competing theories, or testing the predictability of a phenomenon (Shmueli and Koppius, 2011). In this dissertation, this research method is used to test the predictability of certain characteristics of consumption events, with the underlying aim of fostering novel prediction-based feedback interventions. To achieve this, the research approach is divided into two main stages across both papers (Papers I and II). In the first stage, the data of interest is collected. In the papers, this comprised both measurement data and logbooks from participants. The former is used to build powerful features for prediction, while the latter is used to correctly annotate the measurement data through logs from the participants (as they indicate when a certain type of consumption event has happened). After this stage, statistical models are built that learn relationships between the features (i.e., the independent variables) and the event types (i.e., the dependent variable). In a subsequent step, the performance of the models is evaluated, which sheds light on whether the prediction power is good enough to build a corresponding feedback component. Taken together, this research approach acts as a correlational research method as there is no intended manipulation of an independent variable (e.g., through an intervention); instead, the focus of the associated papers is solely on the predictability of certain events through independent variables. Future research could empirically test prediction-based feedback systems that rely on such predicted events (e.g., via an experiment; see Section 4.4.3), and thereby evaluate the value of the predictions in terms of resulting behavioral outcomes.

4.4.2 Literature review

To summarize existing knowledge and outline further research gaps, literature reviews are an important research method (Webster and Watson, 2002). In general, such reviews can be both quantitative and qualitative. While a quantitative review is often used to extrapolate standardized effect sizes across a set of studies with similar interventions (e.g., same medical treatment) or to provide descriptive information on the progress of a field, a qualitative review is more geared towards understanding complex phenomena. What is common in all literature reviews is that authors search literature databases with at least one adequate search string. Subsequently, the retrieved documents are checked for relevance using criteria that specify whether a study should be included in the literature review or not. Then, aspects of interest are extracted from the set of relevant documents, which can be both quantitative and qualitative outcomes. In an ideal setting, this review is done systematically and is well documented so that other scholars can replicate the findings (vom Brocke et al., 2009). There are many reasons for such a systematic approach. For example, it may be worth rerunning a previous meta-analysis, since novel statistical methods might be developed, which can mitigate biases (see, e.g., Maier et al., 2023 on publication bias). Likewise, literature reviews might be repeated to document whether scholars have addressed research gaps that previous literature reviews pointed out.

Within IS research, various types of literature reviews exist, which can differ in their overarching research goal (e.g., summarization of knowledge, explanation building), their search strategy (e.g., selective, comprehensive), and their quality appraisal (Paré et al., 2015). In the context of this dissertation, only a single paper (Paper III) applies of a literature review to extract reasons from the related literature regarding why hand hygiene monitoring systems are often not adopted. Based on the review, the paper develops design recommendations for future hand hygiene systems to overcome the identified barriers to adoption.

4.4.3 Experiments

As most of the papers in this dissertation use experiments (Papers V to IX), the following subsection introduces the associated methods.

4.4.3.1 Foundations

Experiments have become the standard across numerous research disciplines to explore the relationships between different variables of interest. While everyday observations might provide initial hints of underlying relationships, there are several problems associated with such correlational analyses that might lead to spurious results. For example, in testing whether a specific medicine cures patients of a disease, there is potentially an infinite number of confounding variables (Cozby and Bates, 2020) that might serve as alternative explanations for a higher rate of cured patients (the dependent variable). In comparing patients that received the treatment to other groups, treated patients could, for example, have had a stronger immune system, a healthier weight, or less risky lifestyles before the onset of the treatment. As a consequence, a set of methodical approaches has been developed to minimize the risk that confounding variables, rather than independent variables (in this case, the medicine), are having an impact on the dependent variables and therefore lead to wrong conclusions.

A classical approach to reducing the influence of confounding variables is to conduct an experiment in which one group receives the treatment and the second group (control group), has all potentially confounding factors held constant. This allows experimenters to infer that the cause of the changes actually results from the treatment (Recker, 2021). Hence, both groups must be equal as possible before the treatment to reduce potential biases resulting from group differences. To achieve this, experimenters usually apply randomization procedures that randomly assign subjects to the study groups. Nevertheless, it is important to note that such randomization procedures do not guarantee that study groups are statistically equal. Particularly with a low number of subjects, it might be that the randomization is not successful due to, for instance, bad luck. To reduce the risk that this happens, randomization can also be stratified based on certain variables. Such a procedure attempts to distribute the stratified variables as evenly as possible among the study groups.

To test whether randomization procedures lead to balanced study groups, experiments also often comprise a period before the treatment group receives the treatment. This so-called baseline period allows consideration of whether the study groups are statistically equal before the onset of the treatment (Cozby and Bates, 2020). Sometimes, after the removal of the treatment, there is also a phase in which so-called post-treatment effects are measured. Here, experimenters might want to investigate whether the intervention has a lasting influence on the dependent variable(s) after the intervention was removed. Corresponding analyses have to be interpreted with caution as many study participants may never reach the later stages of an experiment, potentially leading to spurious results (see, e.g., Nunan et al., 2018). An example from the environmental domain is that a field experiment might report intriguingly high conservation effects on fuel usage in cars even after the removal of the treatment (e.g., a device that rings when one drives the car too fast). This may be because individuals with a low environmental attitude might have left the experiment, while those with a high environmental attitude—and who can relatively easily be convinced to drive more slowly—have stayed in the experiment.

Before outlining the typical experimental designs used in the papers of this dissertation, it is also important to consider that there are different types of experiments. Laboratory studies take place in an artificial environment for participants (Koch et al., 2019). One of the biggest advantages of laboratory studies is that experimenters have a high degree of control over confounding variables. More specifically, experimenters can ensure that all participants encounter the manipulation of independent variables under almost the same conditions. This is not the case in a field experiment. Field experiments take place in the natural environment of participants, which involves all sorts of confounding variables that are less controllable than those in the laboratory (Harrison and List, 2004). While this can dampen the internal validity of the associated findings, their external validity might be higher. A reason for this is that the findings do not depend strictly on the experimental setting, and hence they can be generalized to more settings. Beyond laboratory and field experiments, the literature often also distinguishes between online and natural field experiments (see, e.g., Koch et al., 2019). While the former take place via the Internet, meaning that participants can sit in front of their computers and fill out a web-based questionnaire, the latter takes place in the natural environment of the participants. Experimenters have no control over assigning specific participants to different study groups in natural environments. Instead, participants face potential manipulations in everyday scenarios and are thus often unaware of the experiment. A popular example of this is studies that have tested so-called towel and linen programs in hotels that encourage hotel guests to reuse their towels and linens (see, e.g., Dolnicar et al., 2017; Goldstein et al., 2008). Hotel guests face manipulation when they enter their room (the natural environment), but experimenters have little control over who is assigned to a room with the program and who is instead assigned to a room under a control condition (without the towel and linen program).

As this dissertation focuses on research questions with the aim of improving personal or societal outcomes under realistic conditions, all of its experiments are field experiments. The main rationale for this is that it would be difficult to observe the behaviors of individuals over a long period of time in the laboratory. While this clearly comes at the expense of internal validity, field experiments allow observation of how the interventions affect individuals under realistic scenarios. These include scenarios where not all participants pay full attention to an intervention, where there are distracting elements next to an intervention, or where an intervention sometimes fails.

4.4.3.2 Experimental designs

Apart from different types of experiments, the research papers in this dissertation also share common experimental designs. Such experimental designs represent patterns to analyze the interplay of multiple treatments by shedding light on interaction effects or to reduce the risk of bias from confounding variables. Interactions are effects that result when multiple independent variables are changed at the same time (Cozby and Bates, 2020). In an experiment that manipulates two independent variables and observes a change in a dependent variable, the observed change could be a result of a single independent variable or the combined result of both independent variables. Given that there are also cases where one independent variable has an impact on the dependent variable but no effect when both independent variables are combined, there are specific experimental designs to capture such relationships. For example, in the case of two treatments (i.e., independent variables), experimenters often apply a 2×2 factorial design. Consider the case of an experimenter (e.g., an insurance company) that wants to determine how financial rewards and consumption feedback have an impact on customers' sustainable driving style in the car. A corresponding 2×2 factorial design on car driving would involve four experimental conditions for study participants: (i) no feedback and no financial rewards, (ii) feedback and no financial rewards, (iii) no feedback and financial rewards, and (iv) feedback and financial rewards. Figure 10 visualizes these conditions.

Such an experiment illustrates not only how individual factors affect the dependent variable compared to the condition receiving no manipulation, it also demonstrates how the combination of both interventions affects the dependent variable. In the context of eco-driving, the experimental design would be valuable as it provides a more profound understanding of whether (i) financial incentives are effective in promoting sustainable driving (the dependent variable), (ii) whether consumption feedback is a cost-effective alternative to incentives, or (iii) whether financial incentives amplify the effects of consumption feedback. Due to these additional insights on the interplay of independent variables, this design was used in Papers IV and IX.

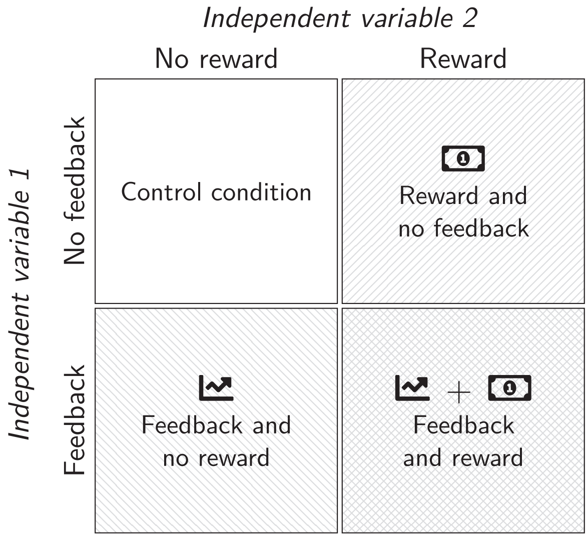


Figure 10: Exemplary 2 × 2 factorial design on eco-driving

A second experimental design used in Papers V, VI, VII, and VIII is the difference-in-differences design. This design is often chosen for longitudinal studies as it substantially improves the internal validity of the experiments (Meyer, 1995), because it contains a baseline phase in which the unmanipulated behavior of the study group is recorded, as well as at least one intervention phase in which the treatment group(s) receive the treatment. As a consequence, experimenters can (i) check whether the study groups are not statistically different before the onset of any treatment and (ii) control for confounding factors that may substantially influence the behavior under investigation. For example, when households in the treatment group receive letters with feedback information on their heating energy use in spring, the winter serves as a baseline phase. In spring, the weather is warmer, so participating households heat less. In the absence of a control group, it is difficult to extract the conservation effect that actually results from the letters due to the effects of the weather. A difference-in-difference design allows full elimination of the effects of the weather if both groups are influenced by the weather to the same extent.

Figure 11 presents a difference-in-difference design illustrating this example. A corresponding statistical method, which is not relevant in this context, estimates the average energy consumption of both groups at baseline and in the intervention phase. This is represented by the coefficients $\hat{y}_{c1}$ and $\hat{y}_{t1}$ in the baseline and by the coefficients $\hat{y}_{c2}$ and $\hat{y}_{t2}$ in the intervention phase. The treatment effect thus is expressed by the change in the dependent variable (i.e., heat energy use) that surpasses the change of in the control group, mathematically represented by $(\hat{y}_{t2} - \hat{y}_{t1}) - (\hat{y}_{c2} - \hat{y}_{c1})$.

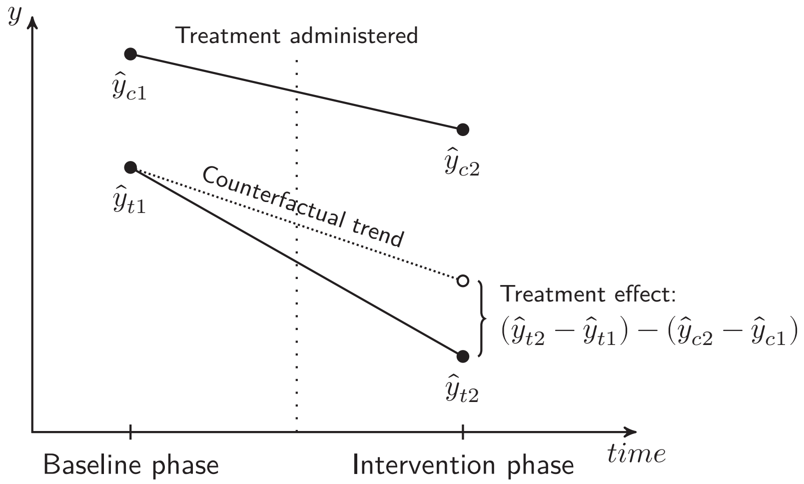


Figure 11: Difference-in-differences scheme (adapted from Dimick and Ryan, 2014)

4.4.3.3 Behavioral interventions of this dissertation

Regarding their experimental designs, most papers in this dissertation administer treatments in the form of behavioral interventions to individuals. While the dissertation's main focus is on feedback interventions, a few papers in the dissertation also test other types of interventions. For an overview, the behavioral interventions of the dissertation can be roughly categorized into three types of interventions, which are briefly summarized below.

- **Beliefs about consequences**: Knowledge about consequences and their influence on associated attitudes have been shown to explain individuals' actions across different domains (e.g., Ajzen, 1991; Schwartz, 1977). Two papers test how relatively simple messages influence individuals' subsequent behaviors (Papers IV and V). As an example, Paper V tests the extent to which an environmental appeal highlighting the energy intensity of an activity and its environmental consequences impacts subsequent conservation behavior.

- **Feedback**: In contrast to interventions that present information on general consequences, feedback interventions provide individuals with tangible information on their personal performance (e.g., on resource conservation). Yet both interventions are often combined as they target multiple barriers to behavior change. More specifically, information on consequences might increase the perceived relevance of feedback, while feedback might create transparency regarding the individual's performance. This dissertation uses feedback interventions to assess their impact on behavior as well as to test theoretical propositions from associated literature (Papers IV, V, VI, VII). For example, Paper VII explores to what extent a social normative feedback intervention providing detailed information on one's online learning time influences students' online learning behavior.

- **Financial incentives**: Financial incentives are a popular instrument to promote behavior change, though it is often claimed that they could crowd out intrinsic motivation for behavior change and induce adverse effects (see, e.g., Schwartz et al., 2019). In this context, Paper V tests the impact of performance-contingent financial incentives on resource conservation and compares their impact to other non-monetary interventions (i.e., interventions aiming at raising general knowledge or providing feedback).

4.5 Data collection techniques

To answer the underlying research questions, the papers of this dissertation use a variety of data collection techniques, as presented below.

4.5.1 Measurement data

Almost all papers in the dissertation rely on measurement data collected by hardware sensors or software components. These measurements were often processed in subsequent data analyses. For example, in the studies that use the presented amphiro shower meter, a crucial dependent variable is not provided by the meter itself: the required heat energy for a shower. However, this information can be extracted by inserting other variables (i.e., the measured water temperature, the measured volume in liters, and the specific heat capacity of water; see Kirtley, 2020) into corresponding physical equations. The following equation calculates the required heat energy for a shower (volume 44 liters, temperature 39 °C) while simultaneously assuming a cold water temperature of 12 °C:

$$\begin{aligned} Energy &= mass * capacity * \Delta\, temperature \\ &= 44\,kg * 4.184 \frac{kJ}{kg * k} * (39\,k - 12\,k) \\ &= 4970.592\,kJ \\ &= 1.38\,kWh \end{aligned} \tag{1}$$

Energy-related variables are often enhanced in this dissertation by considering real-world conditions that influence dependent variables, such as the losses associated with hot water generation (e.g., by considering the typical efficiency rating of water heating systems). Moreover, these numbers are sometimes translated into other equivalents of interest, like CO_2 emissions or monetary costs. This conversion helps to provide a broader perspective on the cost-effectiveness of investigated behavioral interventions.

4.5.2 Survey data

Aside from measurement data, the papers in this dissertation rely on survey data. In general, surveys are used to gather information via questionnaire-type instruments. These instruments can be used to learn more about a given situation from affected persons, to explore the reasons for a specific phenomenon, or to test theories or theoretical constructs (Recker, 2021). In the papers of this dissertation, surveys are primarily used to better understand the behavioral mechanisms induced by interventions as well as to gain information on participants' reception of the studied interventions. For instance, such information is used to conduct subgroup analyses with the aim of testing theoretical assumptions (e.g., do financial incentives only undermine the conservation behavior of intrinsically motivated individuals?). In other research contexts of this dissertation, in which certain consumption events should be identified based on measurement data, the survey data provides so-called ground truth data; this means that the corresponding log data provides important information required to objectively test the performance of an identification approach (more information on this topic is provided in Section 5.1).

4.5.3 Literature

The third data collection used in the papers of the dissertation is past literature. Past research was crucial for all papers to outline research gaps that they aim to overcome. In particular, Paper III performs a literature review to identify why the adoption of hand hygiene monitoring systems often fails. Based on the review's findings, Paper III proposes a novel system to overcome the identified barriers.

5 Data analysis

The papers of this dissertation use a plethora of data analysis techniques to answer the research questions. The reason for this is that the dissertation addresses all the links (i.e., the macro-micro, micro-micro, and micro-macro links) from the outlined research framework, with each of these links substantially differing in their research context. Initiating a behavioral intervention and testing its technical feasibility require different analytical techniques than evaluating the intervention's empirical impact on behavior. This section is therefore structured as follows: First, predictive methods are outlined to shed light on their use for building and technically evaluating approaches that enable novel feedback interventions (i.e., the macro-micro link). Subsequently, the main analytical techniques of this dissertation for testing causal relationships, which concern the micro-micro link and micro-macro link, are presented.

5.1 Supervised machine learning

5.1.1 Foundations

While supervised machine learning is a broad area with many promising real-world use cases (e.g., autonomous driving), the general aim of its underlying methods can be summarized briefly. Supervised machine learning tries to learn relationships between independent variables and an outcome variable (Hastie et al., 2009). To this end, a supervised machine learning algorithm is trained on data where the dependent variable is known. As a result of this training process, a model is created, which is then usually deployed in a real-world application in which the dependent variable is unknown. The model then tries to predict the dependent variable based on the previously learned relationships from the training data.

To accurately predict outcome variables, a variety of machine learning algorithms exist that learn such relationships differently (e.g., decision tree, random forest, support vector machine). The general opinion in the literature is that no best algorithm exists. In other words, the performance of these machine learning algorithms depends on the nature of the data, and it is therefore advised to test multiple algorithms side by side (e.g., Demšar, 2006; Mansilla and Ho, 2004). Irrespective of this, there are some applications in which specific algorithms appear to produce the highest performance levels (e.g., convolutional neural nets for image classification; see Zeiler and Fergus, 2014).

5.1.2 Feature generation

Before crucial aspects of the evaluation of such models are presented, it is important to understand how training data has to be processed to generate models. As supervised machine learning models try to learn relationships between independent variables and at least one dependent variable (continuous or discrete), such models benefit from powerful independent variables (also called features). Oftentimes, these features have to be thoughtfully generated in advance by an individual; a common approach to this is to closely inspect relationships in the data and then mathematically transform these into feature variables. For example, Figure 12 displays measurements of three showers from different individuals (each displayed in a different scatter plot). As the high-resolution measurement data indicate, these individuals appear to differ in their usage behavior by, for example, choosing a different water temperature or stopping the water flow in the middle of the shower.

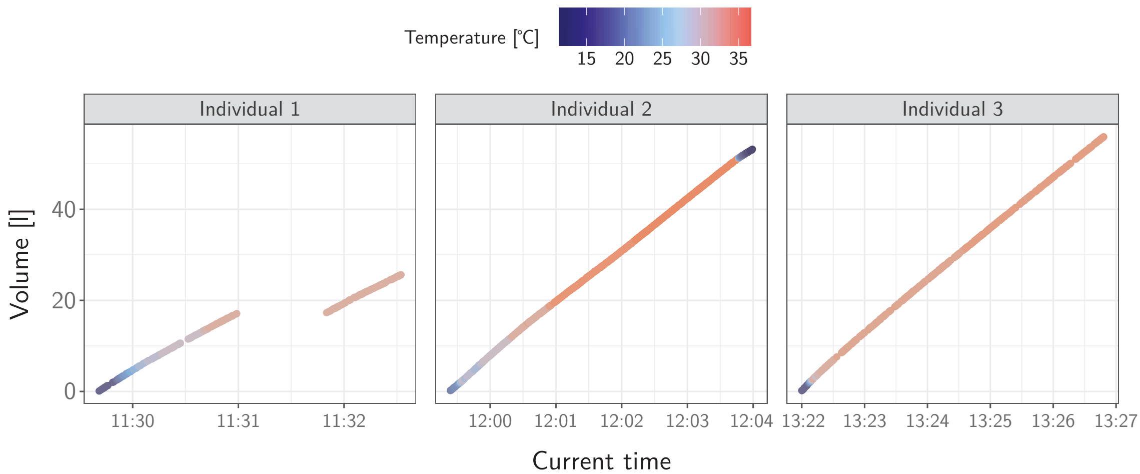


Figure 12: Example for deducing potentially powerful features

To enable the identification of different individuals, one could summarize all the measurement data of each shower into one feature vector (i.e., one row), subsequently annotate each vector with the respective person label, and input the resulting dataset of three vectors into a machine learning algorithm for the training of a model. In a subsequent real-world deployment, the model could potentially be used to provide personalized feedback to users after a shower. Both papers (Papers I and II) that evaluate the feasibility of supervised machine learning algorithms for feedback provision have described the generated features in more detail.

5.1.3 Evaluation

To check the feasibility of machine learning models, a set of different evaluation approaches exists. All these evaluation approaches provide a unique lens on the performance of the models, and they therefore often complement each other in understanding model performance. It is important to note that this section only sheds light on performance metrics for models that predict discrete outcomes.

Before outlining the performance metrics from the papers of this dissertation, it is crucial to describe the overall approach for the evaluation of supervised machine learning models. When testing the performance of a model, it is important not to provide data to the model that it has already seen because this could cause substantial overestimation of the performance of the model for new observations. For example, consider a model that has not learned any general patterns from the training data to discriminate between different outcomes of the dependent variable and has instead stored the training data with the associated outcome variable. In this case, the model would fail to predict novel data points, even though it would provide optimal performance for already-seen data. Generally, the performance of supervised machine learning models can be evaluated without this bias by splitting the data into a training and a testing set (Janiesch et al., 2021), with the training set used to instantiate a model and the testing set used to evaluate the prediction performance. As the evaluation approach should approximate the performance of a model with all the data, the training dataset is usually chosen to be larger than the test set. Moreover, with more training data, algorithms tend to detect more powerful patterns to correctly predict dependent variables (see, e.g., Amari et al., 1992).

In this dissertation, the papers apply k-fold cross-validation to evaluate supervised machine learning models (Hastie et al., 2009). In k-fold cross-validation, the dataset is initially split into k mutually exclusive folds. Subsequently, each of the k folds is used once to test the data, while the other folds are used to train the model. The performance values can then be averaged across the k-cross validation runs. The papers apply k-cross validation for two main reasons. First, the approach uses the data economically by considering each of the data points once for testing. Second, it demonstrates to what extent the performance levels of the machine learning models with slightly different training data might vary, which provides information on the robustness of the models.

To evaluate a model, many performance metrics exist. The accuracy metric is the most widespread and reflects the number of correctly predicted data instances divided by the total number of performed predictions (Hastie et al., 2009). Formally, this is expressed as follows:

$$Accuracy = \frac{Number\,of\,correct\,predictions}{Number\,of\,predictions} \tag{2}$$

The accuracy metric provides a tangible performance metric that is easily interpretable. However, it has several drawbacks for evaluating models that are trained on unbalanced data. In this context, unbalanced data refers to data with a skewed class distribution, meaning that, for instance, for a classification problem with two classes, one class is much more frequent than the second class. The issue with the accuracy metric is demonstrated via the following scenario: consider a classification problem in which one class has a 90% frequency and the second a 10% frequency. Even if a classifier has not learned meaningful patterns from the data, it could just always predict the majority class (i.e., the class with 90% frequency, as determined using the a priori distribution of the training set). The accuracy of the classifier would be rather high even though it provides no meaningful prediction power that exceeds simple statistical rules.

To counteract this, there are two main approaches. First, one could put the accuracy values of machine learning models in relation to statistical estimators, such as by comparing the accuracy values to an estimator that always predicts the majority class. This allows identification of the extent to which the models are better than a statistical estimator in providing prediction power. Second, there are performance metrics that are not prone to skewed class distributions. The “area under the curve” or the “f-measure” can be more adequate performance metrics in this context. The former metric considers the model’s predictions (in this case, probabilities) to differentiate between two classes (Davis and Goadrich, 2006). Based on the model-generated probability values and the true labels of the observations, the performance metric quantifies the prediction power by varying a decision threshold in deciding between both classes. Geometrically, the performance metric represents the area under a curve, which displays the tradeoff between correctly identifying an observation as a specific class (the so-called positive class) and classifying an observation from the other class as the positive class at varying decision thresholds (Bradley, 1997). This relationship is visualized in Figure 13, which has an area under the curve value of 0.821 (the shaded area in the chart). With an increase in the decision threshold (meaning a cutoff point differentiating between both classes), the classifier correctly identifies more observations as belonging to the positive class (the true positive rate, also called recall) but also makes more false positive predictions, causing an increase in the false positive rate. Overall, the performance value of the area under the curve can range between 0 (no differentiation) and 1 (perfect differentiation). It is important to note that a value of 0.5 is equivalent to a statistical estimator that randomly predicts the outcome classes; therefore, a model with this value has no meaningful prediction power. The

True class	Probability for class 1
0	0%
0	10%
0	20%
0	30%
1	40%
0	50%
1	60%
0	70%
0	80%
1	90%
1	100%

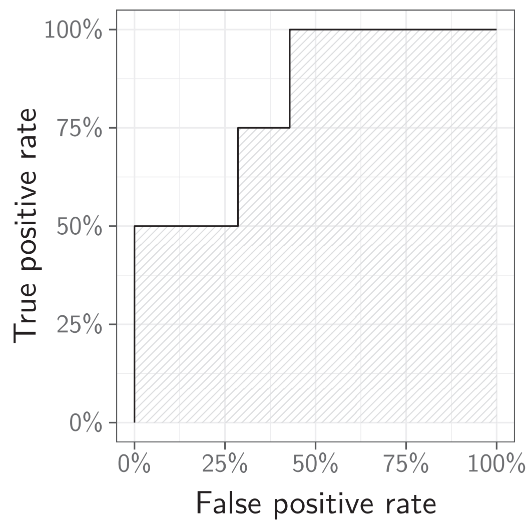


Figure 13: Example for area under the curve performance metric

latter metric (i.e., the f-measure) represents a harmonic mean between two measures: (i) the percentage of detecting all instances from a specific class (metric called true positive rate or recall) and (ii) the percentage of having correctly predicted the positive class out of all positive predictions (metric called precision) (Bekkar et al., 2013). The f-measure thus penalizes a model if the model detects all instances of the positive class but is ineffective in distinguishing the positive class from the other class (Bekkar et al., 2013). Formally, the metric, whose values range between 0 and 1, is expressed as follows:

$$f\text{-}measure = 2 * \frac{true\,positive\,rate\, *\, precision}{true\,positive\,rate\, +\, precision} \tag{3}$$

Regarding these two metrics, there is one further statement in the context of this dissertation to make: both cannot be applied to classification problems with more than two outcome types in the dependent variable (i.e., not a binary but a multi-class classification problem). For the area under the curve metric, however, multi-class variants exist (for an overview, see Ferri et al., 2009). One of these variants is calculated in Paper II. Likewise, the f-measure metric is used for a multi-class problem in Paper I. No specific adjustment for multi-class problems was made in Paper I; rather, the f-measure is calculated as follows. The multi-class problem is handled as a binary classification problem in which, for each outcome type, an f-measure is calculated by comparing the respective outcome type (i.e., the positive class) with the remaining outcome types. At the end of this procedure, an f-measure is obtained and reported for each of the outcome types.

5.2 Causal inference

5.2.1 Foundations

To determine whether a specific behavioral intervention has a desired outcome, inferential statistics is the most important analytical approach used in this dissertation (used in Papers V to IX). Inferential statistics allows the testing of hypothetical cause-and-effect relationships by relying on a few statistical assumptions. These methods enable estimation of whether a treatment is the causal reason for an observation or whether the observation has happened randomly. To make this estimation, a variety of inferential statistics techniques exist which are differently suitable depending on the context of the experimental design and the involved variables (i.e., dependent, independent, and control variables).

Before the specific inferential techniques of this dissertation are outlined, it is important to briefly note a similarity these techniques share: they balance two important error rates, which are visualized in Figure 14. When testing the influence of a behavioral intervention, there are only two true outcomes of the respective study, which the experimenter cannot know. Consider the case of an environmental campaign that aims to reduce electricity consumption in households. On the one hand, the intervention may have no impact on the dependent variable (here x: consumption per month in kWh) represented by the null hypothesis (H_0), the true population value of 260 kWh, and its density curve. On the other hand, the intervention may have an impact on the dependent variable, which is represented by the alternative hypothesis (H_1), the true consumption value of 160 kWh after being enrolled in the campaign, and its density curve. Both density curves follow a Gaussian distribution as it is known that average values of samples are distributed in that way around a true population value.

When the experimenter measures the dependent variable of the treatment group, it has to be recalled that the corresponding average value could be drawn from any of both distributions because one cannot know the impact of a new intervention in advance. Assuming that the treatment actually has an effect and the sample originates from H_1, it could be that the resulting mean value is so "significant" (e.g., in comparison to a sample from H_0; the control group) that one can assume a causal effect. Significance here means that the probability of mistakenly claiming an effect would be negligibly small (Field et al., 2012).

To determine statistical significance, a set of different statistical estimators exists to calculate certain aspects differently. What many of these estimators have in common is that they estimate significance level by testing how probable it is that both samples result from the same density curve (i.e., the same population). This probability can be rather high (see the blue area in Figure 14) so one cannot convincingly state that there is an effect. This would be a Type II error, as one cannot prove an effect that is actually there (Field et al., 2012). Analogously, in the case that an intervention has no true effect and the sample is consequently drawn from H_0, one can still obtain a sample with a statistically significant effect (see the red area). This, would be a Type I error, which occurs when an effect is reported that does not actually exist.

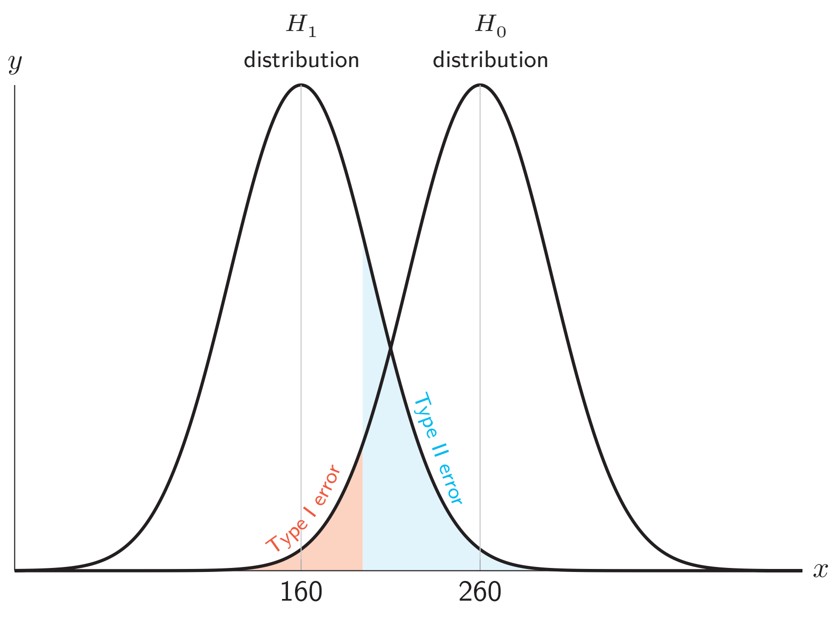


Figure 14: Visualization of Type I and Type II errors

It is important to note that such comparisons can be made not only between a control group and a treatment group but also between treatment groups. To make reasonable comparisons, one has to make certain assumptions about determining sufficiently large sample sizes to increase the power of finding statistically significant effects. The effect size is one of the most important parameters and determines, for example, how many participants should be recruited for a study. In Figure 14, the effect size determines how distant the centers of the density curves are. Importantly, the choice of a power test also depends on the chosen inferential statistics technique, though this is out of the scope of this dissertation (for an overview of power tests, see Cohen, 1988). The following subsection focuses on the inferential statistics techniques that have been used in the papers.

5.2.2 Analysis of variance (ANOVA)

A key statistical method that is used in all experimental papers of this dissertation is the Analysis of Variance (ANOVA). In its basic functionality, the ANOVA method compares a continuous dependent variable for at least two discrete factors. In this dissertation, these independent factors usually represent two or more study groups (e.g., control group, treatment group). Importantly, comparison by an ANOVA and the resulting significance test estimate whether at least one factor has a different effect on the dependent variable than the remaining factors. For example, in the 2 (no feedback, feedback) × 2 (no reward, reward) factorial design on sustainable car driving from page 51, the equation would be as follows:

$$y_i = \beta_0 + \beta_1 \, T_i^{RewardOnly} + \beta_2 \, T_i^{FeedbackOnly} + \beta_3 \, T_i^{RewardWithFeedback} + \epsilon_i \tag{4}$$

where the dependent variable y_i (consumption) is explained by a constant β_0 and the treatment indicators, which are modelled through the use of binary labels (i.e., $T_i^{RewardOnly}$, $T_i^{FeedbackOnly}$, and $T_i^{RewardWithFeedback}$) that take the value of 1 if an observation belongs to the respective group and otherwise are 0. Ultimately, an error term ϵ captures all unmodeled effects.

Within the papers of the dissertation, the ANOVA method is used for two cases. First, the conducted experiments often have a baseline phase in which there is no manipulation of the study groups conducted. Here, the ANOVA, like closely related non-parametric methods for ordinal variables (e.g., the Kruskal–Wallis test), rules out that the study groups are statistically different in key characteristics (see, e.g., Paper V or Paper VII). These so-called randomization checks are important for documenting that the study groups are not fundamentally different in potential confounds so that a causal link can be made to later differences in the dependent variables after the onset of interventions. The randomization checks in the papers indicate balanced study groups; for this reason, the randomization checks are not included in the introductory paper. Second, the ANOVA is also used for the main analysis of Paper IV by demonstrating that study groups are statistically different in the main dependent variable. Subsequently, pairwise tests (e.g., Student's t-tests) are applied between the study groups. As pairwise comparisons inflate the risk of making Type I errors, which is a common problem in the evaluation of differences among study conditions (e.g., four groups lead to six pairwise comparisons, with potentially spurious results), the corresponding paper has further corrected the resulting significance values by applying Tukey's honest significance test as a post-hoc test (see Field et al., 2012, for more information on this topic).

Apart from these applications, ANOVA is a promising statistical method for a variety of datasets as it can be used to flexibly model more complex analyses. For example, ANOVA allows the estimation of unbiased effects when multiple measurements per subject exist (meaning that the observations of a sample are not independent of each other) and can simultaneously assess the impact of an intervention on multiple dependent variables. In the papers of this dissertation, however, regression models are used for more complex analyses. Consequently, the following subsection provides more details on this type of model.

5.2.3 Regression

In contrast to ANOVA, a regression model allows for the inclusion of continuous independent variables to explain the dependent variable;* this is particularly interesting as it can provide more insights into the interplay of potentially moderating factors. For example, in environmental studies, it is often the case that baseline behavior (i.e., resource consumption) fundamentally moderates the impact of an intervention. As a result, the higher the baseline consumption, the stronger the conservation effect. Regression models can quantify such linear relationships, and these insights can then be leveraged by organizations and policymakers to potentially increase the impact of initiatives for behavior change by deliberately selecting subjects for whom an intervention might have the largest impact.

In the papers of this dissertation, regression models were used for experiments that follow the difference-in-differences experimental design. The corresponding studies involve multiple observations per individual, which are split across a baseline phase and at least one intervention phase. As a blueprint for the evaluation of the corresponding studies, the following equation applies:

$$y_{it} = \alpha_i + IN_{it} \times (\beta_1 + \beta_2 T_i) + d_t + \epsilon_{it} \tag{5}$$

where the dependent variable y_{it} is the measurement at time t from the individual i. The variable α_i is an individual-fixed effect for each individual i, which should eliminate time-invariant differences among the individuals. In the environmental context, such differences may result from infrastructural differences (e.g., shower heads with different flow rates). The variable IN_{it} is a binary variable that takes a value of 1 if the measurement is in the intervention phase and in the baseline phase it is 0. Likewise, T_i is a binary label that takes the value of 1

* While a regression model can model the same relationships as an ANOVA, there is often no uniform usage of these terms. Within the papers of this dissertation, there is a strict differentiation: the term ANOVA is only used for factorial designs (e.g., a 2 × 2 factorial design). Regression is used for experiments with more complex study designs (i.e., difference-in-differences).

if the individual i belongs to the treatment group. Otherwise, it takes the value of 0. The variable d_t is a time-fixed effect that aims to eliminate the influence of unobservable variables during each period t. Ultimately, the error term ϵ_{it} captures the unexplained variance.

Within the papers at hand, the equations built on Equation 5 differ slightly because some studies have multiple intervention phases (e.g., Paper V), test whether independent variables have a moderating impact (e.g., Paper VI), or do not include time-fixed effects to better understand the temporal dynamics of the investigated behavior (e.g., Paper VII). Whether the studies used ANOVA or regression, there are several final statistical aspects to consider, which are described in the following.

5.2.4 Statistical estimators for ANOVA and regression

When estimating the influence of the independent variables of a model, which means numerically quantifying model coefficients (e.g., α or β values), there is a choice of different estimators. Studies in this dissertation generally make use of the ordinary least squares method. This estimator determines point estimates (e.g., the average effects) by minimizing the sum of the squared errors between the fitted values and the dependent values of the dataset. Moreover, they often determine standard errors to test the statistical significance of independent variables. While there is a set of different assumptions that must be met to efficiently produce unbiased estimates (e.g., independence of collected observations, no autocorrelation; for an overview, see Greene, 2018), there is one assumption that could be easily violated in the analysis of the collected measurement data: within the difference-in-differences studies of this dissertation, it is usually the case that multiple measurements per observation unit (e.g., individual, room) are made. As a consequence, the resulting standard errors by which corresponding models calculate significance values might be biased (Croissant and Millo, 2019). A reason for this is that the observations are not independent of each other, and the resulting model errors could substantially vary across the observation units, which is a phenomenon called heteroskedasticity. To fix this, the papers that apply the ordinary least squares approach cluster the model errors across each observation unit. This approach acknowledges that there might be an issue in the model and therefore fixes the potentially affected estimates of the model. While the average effects of the independent variable on the dependent variable do not change as a result of this, the associated standard errors might change substantially. If this is the case, the results of the significance tests will also change. In the papers that apply ordinary least squares (Papers V, VI, and VII), only cluster-robust standard errors are presented.

In addition to adjusting standard errors through clustering, there is an alternative approach that was used in Paper IV that used a 2×2 factorial design. Specifically, Paper IV applies the generalized least squares method to infer the point estimate and the standard errors. Instead of retrospectively correcting the standard errors (as with the ordinary least squares method), this approach can take into account that data points stem from the same observation unit.

Interestingly, the approach can then consider this aspect when determining point estimates (e.g., the average effects) and the standard errors. The reasoning for the use of this approach is mainly that generalized least squares can be more efficient in finding unbiased effects when the source of heteroskedasticity is well known (e.g., Bai et al., 2021; Liu et al., 2016). As Paper IV had a relatively small sample of observation units (hostel rooms) with a high degree of heteroskedasticity in its unit-specific measurements, the paper applied generalized least squares to the associated consumption data. For Equation 4 on the 2 × 2 factorial design on sustainable driving, the following equation reflects that car driver j can have multiple measurements (which are clearly not independent of each other):

$$
\begin{aligned}
&y_{ij} = \beta_0 + \beta_1 T_i^{RewardOnly} + \beta_2 T_i^{FeedbackOnly} + \beta_3 T_i^{RewardWithFeedback} + \epsilon_{ij} \\
&\epsilon_{ij} \sim N(0, \sigma_j^2) \quad j = 1, ..., n
\end{aligned} \tag{6}
$$

The second part of the equation specifies the variance structure for each driver j, which allows all the n car drivers to have a different variance structure. The model estimates a factor for each driver based on how the variance differs and subsequently eliminates the resulting differences (and potentially the detected heteroskedasticity) when quantifying the influence of the independent variables. While there are more possibilities for modeling variance structures, Paper IV used the in Equation 6 presented variance structure to correct heteroskedasticity across its observation units (for an overview of different variance structures, see, e.g., Zuur et al., 2009).

5.2.5 Structural equation modeling

Many factors shape the outcomes of a behavior, such as beliefs about consequences, perceived behavioral control, and value orientations, as presented in Section 2 and Section 3. While the ANOVA and regression methods are primarily used to causally test the effects of a behavioral intervention, they are typically not used to test how a behavior forms due to such underlying factors, which may also be of interest when designing interventions. In contrast, structural equation modeling is a statistical method that can shed light on such relationships (Hair et al., 2021). By considering a structural model and a measurement model, this method can investigate complex, theoretical relationships to explain a specific dependent variable. As a consequence, structural equation modeling has been successfully used in many behavioral domains to test and extend theories regarding, among other topics, health behavior (e.g., Gaube et al., 2021; Rhodes et al., 2004), environmental behavior (e.g., Bamberg et al., 2007; Kaiser et al., 2005), and IS usage (e.g., Fang et al., 2014; Maier et al., 2015).

In this dissertation, only Paper IX considers this method. Paper IX specifies a structural model that aims to test the propositions of feedback intervention theory for behavioral interventions that target everyday behavior (e.g., handwashing, showering, etc.). Hence, the remainder of this subsection provides only a brief overview of the structural models of structural equation modeling. Figure 15 displays an alternative specification of Schwartz's norm activation model, presented in Section 3.1, to explain altruistic behavior.

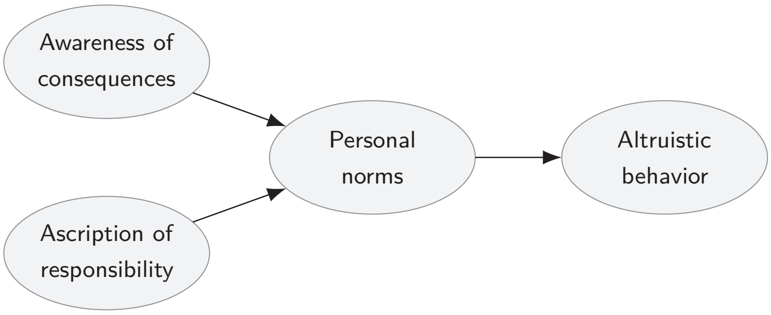


Figure 15: Alternative interpretation of the norm activation model (adapted from Steg and de Groot, 2010)

As outlined, Schwartz (1977) conjectured that individuals' altruistic behaviors depend on personal norms, which are feelings of personal obligation to engage in a specific behavior. According to Schwartz, personal norms may substantially depend on awareness of the consequences of performing or not performing a specific behavior and associated feelings of responsibility. The depicted structural model shows such a relationship. The oval boxes are latent variables, which are measured via surveys (e.g., validated scales that potentially reliably capture the respective variable). The associated arrows indicate a direct influence between one latent variable and another. The model specifies that the activation of personal norms depends on an individual's awareness of consequences as well as the individual's feeling of responsibility. Hence, awareness of consequences and ascription of responsibility are both independent variables; the variable personal norms is both a dependent and an independent variable. Ultimately, altruistic behavior is the main dependent variable of the model. Together with a measurement model, the hypothesized relationships between the variables of the structural model can be statistically estimated and tested. Similarly, the validity of different model specifications can be explored, as described for the norm activation model in Section 3.1.

It is important to note that such structural models can be highly complex. In the model of Figure 15, the variable awareness of consequences could also explain one's feelings of responsibility to perform a specific altruistic behavior. Likewise, concrete measurable outcomes of altruistic behavior that researchers study (e.g., "Which latent variables explain blood donation?") can be incorporated directly into a model as a dependent variable (see, e.g., Hair et al., 2021). However, as only Paper IX relies on structural equation modeling, it is out of the scope of this dissertation to provide further details on this topic.

6 Main results

This section presents the results of this dissertation. It first focuses on the results of the papers that instantiate and validate novel systems for feedback provision (Chapter 1). Subsequently, the section sheds light on the results of the papers that aim to evaluate the behavioral responses to feedback interventions (Chapter 2).

6.1 Chapter 1: Instantiation and validation

6.1.1 Paper I*: Enabling more specific feedback through an automated identification of different water consumption events

Highlights

- In the environmental domain, studies have suggested that feedback is most effective when it is highly specific. One reason for this is that individuals are often unaware of the environmental footprint of their activities. When they receive feedback showing them which activities are resource-intensive, they can direct their attention more efficiently to activities that better conserve resources. Yet the provision of feedback on different environmental activities usually requires multiple hardware sensors to capture the associated data on resource consumption.
- The present paper explores whether the measurement data of a single, central sensor attached to the main water pipe of a household is sufficient to identify specific consumption events from the data traces. Indeed, the paper provides evidence that such consumption events can be reliably identified through the use of supervised machine learning algorithms.
- Ultimately, the identification approach could be used by downstream IT systems to provide more specific and potentially more effective feedback to individuals.

Experiments in the environmental domain show that feedback is most effective when it is delivered clearly and on actions that recipients can influence (see, e.g., Gerster et al., 2020; Tiefenbeck, 2014). In this context, Paper I investigates a technical approach intended to identify water consumption activities that happen inside a household. The identification of such activities can be used by feedback systems to subsequently provide residents with activity-specific feedback on their water consumption. The main rationale for why such a system might be beneficial is that individuals often under- or overestimate the environmental footprint of their activities and therefore undertake ineffective curtailment efforts (Truelove et al., 2014). If individuals better comprehended the environmental outcomes of their activities, they could direct their attention more efficiently to activities where they could save more.

To investigate the feasibility of a technical approach in fostering activity-specific feedback, Paper I conducted a field study, collecting water consumption data from two households. In both households, an ultrasonic sensor was attached to the main water pipe that measured the water consumption. Beyond collecting measurement data, residents of the households took notes of water consumption events (i.e., type of water-consuming fixture/appliance, timestamp). This data was used to build feature vectors to train supervised machine learning models with the aim to identify and classify water consumption events.

∗ S. Schöb, S. A. Günther, K. Regensburger, and T. Staake (2018). “NIWM: Non-Intrusive Water Monitoring to Uncover Heat Energy Use in Households.” *Computer Science - Research and Development* 33 (1-2), pp. 127–133. DOI: 10.1007/s00450-017-0353-8.

Although the collected dataset was rather small with two households and 209 labeled water consumption events, the approach yielded good results. Overall, the approach identified 85.17% of the reported events, with most undetected events stemming only from small water extractions (e.g., cleaning the sink after brushing teeth). In distinguishing water extractions from five fixtures and appliances (namely "Dishwasher," "Shower," "Tap," "Toilet," and "Washing machine"), the random forest classifier achieved an average precision value of 81.62% and an average recall value of 80.42%. In addition, the random forest classifier achieved an area under the curve value above 0.90 for each appliance and fixture. Paper I therefore responds to the research framework of the dissertation (i.e., the macro-micro link) because these predictions can be used by downstream IT systems to provide activity-specific feedback. Such feedback could in turn empower significant resource conservation from individuals as it supports them in learning which water-consuming activities are particularly resource-intensive.

6.1.2 Paper II: Empowering personalized and potentially more effective feedback by detecting individuals in water consumption data*

A growing stream of research suggests that feedback could be more effective in promoting behavior change if it is tailored to the individual recipient (see, e.g., Costa and Kahn, 2013; Schultz et al., 2007; Sütterlin et al., 2011). The theoretical reasoning supporting this finding is that individuals have different value orientations and motives for engaging in curtailment behavior. For instance, while some individuals set a focus on associated financial benefits, others might be more strongly motivated by the associated environmental outcomes (e.g., CO_2 savings, kWh savings, etc.). A main challenge for tailoring feedback is that the sensors which measure environmental behavior are often used for multiple individuals at once. For example, the presented system of Paper I measures the aggregated consumption of all individuals of a household and cannot distinguish between specific individuals. Thus, to provide an individual with tailored feedback, the system must first determine from which individual it is collecting the corresponding consumption data. To facilitate this, a system might employ supervised machine learning that could recognize person-specific patterns in the measurement data (e.g., for water-consuming activities: time, water temperature, volume).

Highlights

- In numerous instances, hardware sensors measure the consumption data of several individuals, making it challenging for downstream feedback systems to provide person-specific feedback to its users (e.g., displaying consumption trends for a single user).
- The present paper explores whether consumption data exhibits unique patterns that can be leveraged to identify specific individuals. For the activity of showering, the paper provides evidence that there are person-specific patterns in the consumption data. These then allow for a reliable differentiation of individuals through supervised machine learning algorithms. Interestingly, the approach requires little data to achieve a satisfactory performance.
- Overall, the identification approach could be used to provide personalized and potentially more effective feedback. For example, the associated measurement device (amphiro b1) could display person-specific elements on its LED. Likewise, a companion app on a mobile phone could use this information to enable, e.g., energy efficiency competitions between individuals of a household.

* S. A. Günther, C. Stingl, V. C. Coroamă, S. Schöb, and T. Staake (2020). "Empowering Personalized Feedback on Hot Water Usage: A Field Study with Shower Meters." *Proceedings of the 35th Annual ACM Symposium on Applied Computing.* Brno, Czech Republic. DOI: 10.1145/3341105.3374053.

Paper II addresses the challenge of creating tailored feedback for the described shower metering device. Though the meter can promote high conservation effects, its impact is bound to the place of action and does not allow for subsequent person-specific motivating elements (i.e., a mobile app displaying one's personal conservation behavior). To overcome this limitation, Paper II presents the results of a field study that collected data from six locations: five residential apartments and one company in which employees regularly used the showers. For each of the showers, a shower metering device was installed with a nearby gateway that transmitted the measurement data to a server. While participants in the apartments took notes on their showers, the employees had a mobile app for this purpose. Both forms of logging required the participants to note down a personal identifier and a timestamp, which were used to build feature vectors for the training of supervised machine learning models. After several pre-processing steps, the filtered dataset ended with 691 shower events from 28 individuals.

For each of the locations, the classifiers had a high prediction power in identifying individuals. For instance, the random forest classifier achieved an average accuracy value of 98.8% when distinguishing between individuals from a two-person household. For the most challenging setting—differentiating between five users of a company shower—the accuracy dropped to an average value of 83.2%. Moreover, an additional analysis indicates that the approach yields satisfactory performance even with little training data (i.e., 3 to 5 showers), implying that users do not have to label very many showers to initialize the identification component (i.e., the training of the classifiers). Ultimately, Paper II addresses the macro-micro link of the research framework by demonstrating that consumption data can be used to identify individuals and provide them with personalized feedback. With regard to the shower meter, this could inspire new features for the mobile app of Amphiro AG to motivate users toward long-term resource conservation (e.g., efficiency competitions between individuals of a household).

6.1.3 Paper III: A novel approach to feedback provision that might help to overcome current adoption issues of hand hygiene monitoring systems*

Although proper hand hygiene is one of the most effective ways to reduce the spread of infectious diseases (Bloomfield et al., 2007), compliance with proper hygiene practices often appears to remain relatively low (see, e.g., Makhni et al., 2021). To support individuals in adhering to proper hand hygiene guidelines, a stream of monitoring systems has emerged that can provide feedback on hand hygiene, yet many of these systems still have low adoption rates (see, e.g., Braun et al., 2009; Durant et al., 2020) and therefore would lead to only modest behavioral improvements when extrapolating their effects to target populations such as hospitals.

* C. Stingl, S. A. Günther, and T. Staake (2021). "A Feedback Information System for Improving Hand Hygiene on a Personal and Organizational Level." *Proceedings of the 16th International Conference on Wirtschaftsinformatik*. Duisburg, Germany.

Paper III addresses this topic by conducting a literature review to shed light on the current challenges that hand hygiene monitoring systems face, with the aim of stimulating the development of new systems. Based on the results, the paper identifies four main barriers to the adoption of such systems, including costs, privacy issues, usability issues, and accessibility issues. Many systems bear high costs, record privacy-sensitive data, or require distinct hardware for each individual to receive feedback. To overcome these barriers, the paper provides six design recommendations for future hand hygiene systems, which are then considered for the conceptualization of a new feedback system.

Highlights

- Despite the potential of monitoring systems in improving individuals' adherence to proper hand hygiene compliance, their adoption rates are quite low. In this paper, a literature review is conducted to identify barriers for the adoption of such systems. Likewise, to overcome these barriers, the paper proposes several design recommendations for the development of novel monitoring systems.
- Based on the identified barriers and the outlined design recommendations, the paper presents the conceptualization of a new feedback system to improve hand hygiene. This feedback system has been implemented as a fully functional prototype which the paper describes.

The proposed feedback system consists of a touchless faucet, a soap sensor, a feedback display attached near the respective sink, and a gateway that transmits the data to a server (Figure 8, page 45). As soon as an individual washes their hands or uses soap, the gateway captures corresponding measurements from the faucet and the soap sensor. Consequently, the gateway can control the feedback display and provide feedback on an individual's handwashing behavior. Given that the system is based on relatively inexpensive hardware components (e.g., not cameras), privacy is maintained by not identifying its users, and any individual can receive feedback (i.e., even individuals who wash their hands only once at a particular faucet), the system should overcome the outlined barriers for adoption.

Thus, Paper III addresses the macro-micro link of the research framework by making recommendations on the design of hand hygiene monitoring systems to overcome existing adoption issues; this in turn can fuel the development of novel systems similar to the system outlined that could improve individuals' health at scale.

6.2 Chapter 2: Evaluation and recommendations

6.2.1 Paper IV: Adverse effects of corporate carbon offset programs and the potential of consumption feedback as a remedy*

More and more organizations are offsetting the emissions associated with consumers' demand through the use of carbon offset programs. Such programs reduce the carbon emissions elsewhere through the promotion of green initiatives like reforestation programs, energy efficiency campaigns, or clean energy projects (Gössling et al., 2007). By allowing consumers to maintain carbon-intensive lifestyles, carbon offset programs have been strongly questioned as a proper instrument to reduce climate change (Anderson, 2012). In particular, offset programs allow

∗ S. A. Günther, T. Staake, S. Schöb, and V. Tiefenbeck (2020). "The Behavioral Response to a Corporate Carbon Offset Program: A Field Experiment on Adverse Effects and Mitigation Strategies." *Global Environmental Change* 64, Article nr. 102123. DOI: 10.1016/j.gloenvcha.2020.102123.

consumers to feel entitled to engage in overconsumption without experiencing feelings of guilt about the consequences for the environment. As offset programs have a history of being weakly regulated in terms of documenting their real-world impact and have even been fraudulent (Walters and Martin, 2013), it is possible that their use by individuals or organizations could do more harm than good to the climate. Indeed, there is a body of research on the behavioral responses of consumers who actively opt into such offset programs. Based on these studies (i.e., Harding and Rapson, 2019; Jacobsen et al., 2012; Kotchen and Moore, 2008), the programs could encourage higher demand levels from customers depending on the additional price effects resulting from participation in offset programs. Companies' increasing offer of carbon-neutral products and services without explicitly passing the cost of offsetting on to customers raises the question of whether these initiatives dampen customers' environmental protection efforts. While the associated motives for using offset programs might genuinely aim to reduce climate change and the programs might be helpful for customers in choosing environmentally sustainable products and services, there is also the question of whether potentially resulting adverse effects can be mitigated.

To study these questions, Paper IV presents the results of a natural field experiment conducted in a youth hostel in Nuremberg, Germany. By following a 2 (no offset program, offset program) × 2 (no feedback, feedback) design, the experiment measured the conservation behavior of hostel guests when using the shower. The bathroom of each hostel room was equipped with a shower meter, and decals were placed on the shower walls. The decal and the content of the display varied across the conditions. The decal in the no offset condition informed guests that water consumption in the shower is resource-intensive. The decal in the offset condition additionally informed guests that their use of the shower is climate-neutral as the youth hostel contributes to a green project that mitigates associated CO_2 emissions. While the shower meter in the control condition only displayed the water temperature, the shower meter in the feedback condition displayed real-time feedback on resource consumption to the guests. Based on the 2 × 2 study design, the study comprised four groups: a no offset and no feedback condition (O^-F^-), an offset and no feedback condition (O^+F^-), a no offset and feedback condition (O^-F^+), and an offset and feedback condition (O^+F^+). For the experiment, these conditions were randomly assigned to the hostel rooms. The randomization was stratified by floor number and number of beds in the rooms to reduce potential biases resulting from infrastructure-related differences (e.g., different water pressure across the floors).

Highlights

- The paper presents the results of a natural field experiment (8,892 measurements) with a 2 (no offset program, offset program) × 2 (no feedback, feedback) experimental design. As previous studies have shown that feelings of guilt and responsibility for environmental consequences explain pro-environmental behaviors, the study investigates whether corporate carbon offset programs increase consumers' resource use. The conjecture is that consumers' actions—once they are made climate-neutral through organizations—have fewer environmental consequences, which might reduce feelings of guilt and responsibility. This might ultimately dampen consumers' conservation efforts.
- Indeed, individuals in the carbon offset condition had a higher consumption level than those in the control condition (i.e., no feedback and no offsetting condition).
- Real-time feedback fully counteracted the adverse effects of the carbon offset condition. The feedback might have reactivated consumers' perceived responsibility by highlighting that the consumers are still using precious environmental resources despite the climate neutrality of their associated actions.

Overall, this study design enabled evaluation of not only how individuals respond to the carbon offset program but also to what extent real-time feedback can cancel the impact of the offset program on conservation behavior. It is important to note that the experiment considers all the rooms with the no offset and no feedback condition (i.e., O^-F^-) as the control group.

After finding a statistically significant effect on energy consumption through the use of ANOVA ($F(3, 8888) = 31.82$, $p < .001$), the paper presents pairwise comparisons of the four study groups, which are displayed in Table 6.

Table 6: Main results of Paper IV

Contrast	Difference [kWh]	Standard error
O^+F^- – Control	$+0.175^{***}$	0.026
O^-F^+ – Control	-0.044	0.025
O^+F^+ – Control	-0.042	0.023
O^+F^- – O^-F^+	$+0.219^{***}$	0.026
O^+F^- – O^+F^+	$+0.217^{***}$	0.024
O^-F^+ – O^+F^+	-0.002	0.023

Notes: The table presents the treatment effects on energy use (in kWh) of a generalized least squares regression, which takes into account that the observations of each hostel room share a room-specific variance structure. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The p values have been adjusted according to Tukey's method for the comparison of a family of four estimates.

As Table 6 documents, consumption in the offset only condition is statistically higher than in the control condition. More specifically, consumers used 0.175 kWh (15.5%, p value $<$ .001) more per shower than the control group, with an average consumption of 1.13 kWh, demonstrating that offset programs can have adverse effects on consumers' environmental behaviors. However, this changed when consumers also received real-time feedback as the O^+F^+ group used slightly less heat energy on average than the control group (the difference is not statistically significant, p value $= 0.270$). Interestingly, the consumption of this group is statistically equal to the group that only received real-time feedback (O^-F^+) and statistically different from the group in the offsetting condition (O^+F^-), implying that real-time feedback eliminated the adverse effect of the offset program.

Overall, Paper IV demonstrates that organizational initiatives that make consumers' demands climate-neutral can have adverse effects on consumer behavior. As more and more organizations rely on offset programs to promote their products and services as climate-neutral, this paper questions the careless use of offset programs without considering unintended consequences. By additionally showing that real-time feedback can counteract adverse effects, Paper IV addresses the micro-micro link of the research framework by contributing to the understanding that seemingly small changes to individuals' decision-making context can have a large effect on subsequent conservation behavior. This could motivate organizations to rethink their use and communication of offset programs to consumers, which may improve the environmental outcomes of associated sustainability initiatives.

6.2.2 Paper V: The dominance of feedback over financial incentives and environmental appeals in promoting resource conservation*

Highlights

- The paper presents the results of a randomized controlled trial with 326 individuals (20,830 observations) using a 3 (appeals, financial incentives, no cues) × 2 (no feedback, feedback) design. To better understand how environmental campaigns can foster resource conservation, the study tests the predictions of the standard economic model and motivation crowding theory in the absence and presence of consumption feedback. Previous research has not convincingly analyzed the interplay of consumption feedback and financial incentives and did not compare their effect sizes with those of environmental appeals.
- Performance-contingent financial incentives alone stimulated resource conservation, which is consistent with the standard economic model and refutes predictions of motivation crowding theory.
- Consistent with previous studies, environmental appeals alone were ineffective in promoting resource conservation.
- Neither financial incentives nor environmental appeals amplified the conservation effects of consumption feedback. Feedback is dominant over financial incentives and environmental appeals.
- The observed behavioral effects are overall consistent with resource allocation theory (Kanfer and Ackerman, 1989), implying that the interventions have diminishing marginal returns on individuals' attention and therefore on conservation behavior.

In the environmental domain, a wide range of experiments have demonstrated that behavioral interventions can promote resource conservation (see, e.g., Abrahamse et al., 2005). In particular, non-monetary interventions, such as environmental appeals or feedback, have received substantial attention recently, while monetary interventions have often been said to crowd out intrinsic motivation for resource conservation (see, e.g., Schwartz et al., 2019). However, the literature is lacking a direct comparison of these types of interventions, which raises several practical and theoretical questions. To begin, it is totally unclear for practitioners how the effect sizes of different interventions vary. A comparison of different experiments that tested specific interventions in isolation might be biased as studies often target different activities with different difficulty levels of behavior change or have samples with different socio-demographic backgrounds that may influence the effects of interventions. A better comparison of interventions is therefore needed as climate change is an urgent issue, with limited time left to reach Paris Agreement goals (Höhne et al., 2020). Thus, interventions with modest effect sizes might not be enough to tackle climate change. In addition, the effect of financial incentives on resource conservation has been studied with too little focus on factors that potentially mask crowding-out effects on conservation behavior (e.g., intrinsic motivation at baseline). Moreover, insights from literature on whether financial incentives are helpful in promoting conservation behavior when the financial incentives are combined with consumption feedback are inconclusive. With reference to motivation crowding theory, financial incentives might cancel out the desirable effects of feedback, while the standard economic model would predict higher conservation effects in response to the additional provision of financial incentives.

To study these underlying research questions, Paper V uses a 3 (financial incentives, environmental appeals, no cues) × 2 (no feedback, feedback) experimental design to promote

∗ S. A. Günther, S. Schöb, V. C. Coroamă, V. Tiefenbeck, F. Mattern, and T. Staake (under review). "All Eyes on Consumption Feedback: A Randomized Controlled Trial on the Interplay of Financial Incentives, Environmental Appeals, and Consumption Feedback." Working Paper. Submitted to Experimental Economics.

resource conservation for a resource-intensive activity (i.e., showering). Showering is an everyday activity that not only uses a great deal of water but also high amounts of heat energy (Bertrand et al., 2017) that is often derived from carbon-intensive fossil fuels. For the randomized field experiment, 398 individuals were recruited from three student dorms in Bamberg, Germany. Each of the residents received a shower meter and was randomly assigned to one of the four study groups: an environmental treatment group, a financial treatment group, a treatment group without any cues, and a control group. The experiment comprised three main experimental phases: in the first phase, the treatment groups received a text message (SMS) with group-specific content (e.g., financial incentives for the financial treatment group, environmental appeals for the environmental group), in the second phase, the treatment groups additionally received different intensities of consumption feedback (first outcome feedback with group-specific content and then real-time feedback combined with outcome feedback), and in the third phase, the outcome feedback with the group-specific contents (e.g., appeals, financial incentives) was withdrawn. This study plan allows for both (i) comparison of appeals with financial incentives in terms of conservation effects and (ii) calculation of how these cues moderate the effect of consumption feedback and whether feedback has a lasting effect even when the cues are withdrawn. Table 7 presents the main results of this paper, namely the consumption change of the treatment groups in comparison to the control group (for the specification of the regression equation, see Paper V). These results are based on 20,830 observations from 326 individuals that remained after the data pre-processing steps.

As displayed in Table 7, financial incentives had a statistically significant effect on resource conservation (i.e., conservation effects of 8.5%), while environmental appeals had no measurable effect. However, these cues did not moderate the effect of consumption feedback. In other words, once consumption feedback was given to the individuals, the consumption behavior was not statistically different among the treatment groups (e.g., for the outcome feedback phase: $F(2, 19999) = 0.329$; p value = .720). Therefore, financial incentives are not required for feedback to foster resource conservation. In fact, outcome feedback alone induced even higher conservation effects of 15.4%. Additionally, it appears that the incremental provision of interventions has diminishing returns on resource conservation. For example, the higher intensity of consumption feedback led to marginally higher conservation effects that are only statistically significantly different when collapsing all the treatment groups in an additional analysis (+6.7%; p value = .092). Interestingly, regarding the post-treatment effects of cues, the results indicate that a motivational cue (i.e., incentives or appeal) is necessary to sustain long-term conservation behavior as the treatment group without any cues stopped saving resources.

Beyond the main treatment effects, the paper investigated whether specific factors shape behavioral response (i.e., conservation behavior) to the interventions. In contrast to predictions of motivation crowding theory, the results suggest that a high level of intrinsic motivation for resource conservation does not attenuate the behavioral response to feedback when financial

Table 7: Main results of Paper V

Independent variable	Energy use per shower [kWh] Financial group	Environmental group	General group	Temporal effect
Single text message	−0.236**	−0.072	−0.008	+0.020
	(0.116)	(0.134)	(0.132)	(0.104)
Outcome feedback	−0.376*	−0.556***	−0.448***	+0.163
	(0.212)	(0.187)	(0.158)	(0.147)
Outcome and real-time feedback (1)	−0.701***	−0.623***	−0.650***	+0.175
	(0.182)	(0.170)	(0.156)	(0.154)
Outcome and real-time feedback (2)	−0.513***	−0.553***	−0.465***	+0.193
	(0.162)	(0.146)	(0.166)	(0.161)
Real-time feedback only	−0.478***	−0.497**	+0.212	+0.054
	(0.162)	(0.214)	(0.367)	(0.186)
Observations	5,098	4,955	5,189	
Clusters	79	78	79	

Notes: The table presents the treatment effects on energy use (in kWh) using a fixed-effects ordinary least squares regression, controlling for individual- and time-fixed effects. The standard errors are shown in parentheses and adjusted for clustering at the individual level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Moreover, after applying the false discovery rate (see Benjamini and Hochberg, 1995) to the p values of the treatment effects, the statistically significant effects remain statistically significant. Control group observations: 5,588; control group clusters: 90; constant: 2.75; within R^2: 0.005.

incentives are given. Overall, the results indicate that financial incentives have no adverse effects on resource conservation and in fact produce desirable effects. Another question was whether the effects of financial incentives depend on participants' financial situation. A corresponding analysis did not find substantial support for this. Lastly, the analyses of the paper did not reveal heterogeneous effects of the tested environmental appeal.

In sum, Paper V extensively compares the effects of different important behavioral interventions on resource conservation. By demonstrating how these changes to individuals' decision-making context influence their subsequent conservation behavior, it responds to the micro-micro link of the research framework of the dissertation. It also addresses the associated micro-macro link by shedding light on the interplay between the effects of the interventions and individuals' socio-demographic backgrounds and personality variables. Consequently, Paper V lends insight into how these micro-level observations can be projected to the macro level, where the recipients of interventions may differ regarding investigated factors (e.g., intrinsic motivation for resource conservation, financial situation).

6.2.3 Paper VI* and Paper VII**: Social normative feedback to counteract procrastination in higher education students

Highlights

- Previous studies have shown that descriptive comparative feedback leads to mixed effects on learning behavior within digital learning platforms. The present papers argue that the mixed effects of comparative feedback can be explained through the lens of social norms theory: people who are already using a platform actively might be discouraged from online learning once they realize that others are much less active.
- Two randomized controlled trials (N=58 and 118) evaluate the effects of a comparative feedback intervention that is informed by social norms theory (featuring a descriptive and an injunctive norm to counteract adverse effects from active users). The feedback intervention is embedded into the learning platform of two university courses. The aim of the feedback intervention is to mitigate procrastination behavior (i.e., self-regulation issues) and test the implications of social norm theory.
- In both courses, there is substantial evidence that the intervention reduces procrastination. More specifically, in response to the feedback, individuals tend to increase their online learning behavior in the middle of the respective course (instead of cramming right before the exam). There is no evidence of heterogeneous treatment effects, which could explain the mixed effects observed in other studies. Overall, the results provide evidence for a decisive influence of injunctive norms, which related studies have neglected in their feedback design.
- Further analyses, outlined in this section of the dissertation, show that the social normative feedback intervention increased course performance (i.e., exam-taking rate or exam performance) in both courses.

Digital learning environments have been increasingly used in academic contexts to provide both instructors and students with the benefits of online learning (see, e.g., Adedoyin and Soykan, 2020; Norberg et al., 2011; Wanner and Palmer, 2015). For example, in contrast to face-to-face classroom sessions, asynchronous online learning permits students to better align their learning activities with their obligations (e.g., childcare, part-time work, etc.) while also enabling them to work through content (e.g., lecture videos) at their own pace. In a similar vein, instructors might have a smaller workload when running a course once its online content is set as the content might be viable for several years. Despite these advantages, there is one major pitfall of online learning for learners: Knowing that the content, such as lecture videos, is available at later stages of the course, students might postpone their learning (see, e.g., Tuckman, 2005) and instead engage in non-course activities with immediate benefits (i.e., leisure time). If students cannot catch up with the content before the graded submissions or exam of a course, this could harm their course performance and might even lead to severe side effects (e.g., dropping out of one's study program, having to pay an additional semester of tuition fees, etc.).

Paper VI and Paper VII explore the effects of a social normative feedback intervention that is embedded into a digital learning environment with the aim to reduce procrastination among participants of two university courses. Within both courses, students use the digital learning environment to work asynchronously through lecture videos, lecture slides, or quizzes. Apart from the online content, these courses offer synchronous tutorials in which the lecture materials are discussed and in which course participants can pose questions. Paper VI presents the experimental evaluation of the social normative feedback intervention in a master's level elective course, Paper VII repeats the study on a bachelor's level compulsory

* S. A. Günther, T. Veihelmann, and T. Staake (2020). "Leveraging Social Norms to Encourage Online Learning: Empirical Evidence from a Blended Learning Course." *Proceedings of the 41st International Conference on Information Systems*. Hyderabad, India.

** S. A. Günther (2021). "The Impact of Social Norms on Students' Online Learning Behavior: Insights from Two Randomized Controlled Trials." *Proceedings of the 11th International Conference on Learning Analytics & Knowledge*. Irvine, CA, USA. DOI: 10.1145/3448139.3448141.

course and compares the findings to the elective course. For this reason, both papers are described here together.

In both courses, a difference-in-differences experimental design is used, meaning that there is a baseline phase of six weeks in which the control and treatment groups can browse and work on the online content of the respective course. From the seventh week on, the participants in the treatment group can also see the social normative feedback intervention every time they access the course overview page. This feedback design consists of a descriptive element, which compares one's online learning time from the previous week with the average online learning time of other course participants in the previous week, and an injunctive element showing approval or disapproval of one's online learning time (see Figure 16). The overarching idea behind the intervention is that it provides feedback on an aspect that many students can easily change by simply spending more time on the online learning platform. This is in stark contrast to the provision of comparative performance feedback, through which students might develop feelings of being unable to achieve the corresponding performance levels of their peers, potentially increasing their risk of quitting the course (see Rogers and Feller, 2016).

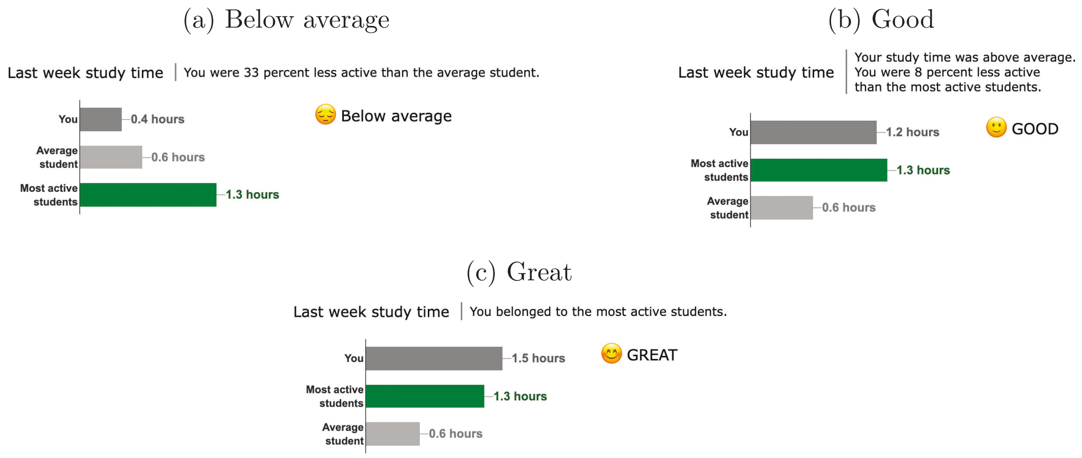


Figure 16: Comparative feedback embedded into the individual course overview page

It is important to note that the analyses of the papers explored the effect of the intervention on the dependent variable "online learning time" as the online learning time and corresponding online learning activities from the recorded log data are strongly correlated. Three analyses are presented in the papers to test (i) whether the overall online learning time increased in response to the intervention, (ii) whether students learned longer on the platform in the middle of the respective course and less before the exam, (iii) and whether there are heterogeneous effects of the feedback intervention. The main finding of the papers is from the second analysis, which is displayed in Table 8.

Table 8 demonstrates how the online learning time and the impact of the feedback differs among three time segments (I_{S_i}) in the intervention phase compared to the baseline phase. The main insights of the studies lie in the interaction between the time segments and the feedback

Table 8: Main results from the comparison of both experiments in Paper VII

	Online learning time per week [minutes]	
Independent variable	Elective	Compulsory
Intervention phase$_{S_1}$ (abbr. I$_{S_1}$)	$+3.92$	-15.27^{***}
	(5.81)	(3.26)
Intervention phase$_{S_2}$ (abbr. I$_{S_2}$)	$+5.81$	-15.31^{***}
	(5.82)	(5.08)
Intervention phase$_{S_3}$ (abbr. I$_{S_3}$)	$+58.13^{***}$	$+35.81^{**}$
	(22.35)	(17.47)
I$_{S_1}$ $\times$ Treatment	-8.88	$+15.90^{***}$
	(7.16)	(5.46)
I$_{S_2}$ $\times$ Treatment	$+23.00^{**}$	$+12.81$
	(11.69)	(8.13)
I$_{S_3}$ $\times$ Treatment	-33.71	-26.03
	(25.84)	(19.88)
Constant	28.66	38.78
Number of participants	58	118

Notes: The table presents the temporal treatment effects on online learning time using a fixed-effects ordinary least squares regression controlling for individual-fixed effects. The standard errors are presented in parentheses and adjusted for clustering at the individual level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

condition. For instance, in the elective course, the feedback increased participants' online learning time in the second time segment (i.e., the second half of the course in which still new content was published) by 80.25% on average. Moreover, participants who received the feedback spent less time cramming (i.e., intensively learning in short amounts of time) for the course right before the exam (I_{S_3} $\times$ Treatment) than the control group. However, this interaction effect is not statistically significant, potentially due to the relatively low number of participants ($N = 58$ participants). In the compulsory course ($N = 118$ participants), the treatment effects are more surprising. The three time segments suggest that control and treatment spent less time on the online course during the semester (S_1 and S_2) and more right before the exam (S_3). Conversely, the point estimates of the treatment effects indicate the opposite, cancelling the procrastination behavior observed from the control group in the form of the phase effects. A corresponding linear hypothesis test finds support for this (H_0: $\mathrm{I}_{S_1} + \mathrm{I}_{S_1} \times \text{Treatment} = 0 \wedge \mathrm{I}_{S_2} + \mathrm{I}_{S_2} \times \text{Treatment} = 0 \wedge \mathrm{I}_{S_3} + \mathrm{I}_{S_3} \times \text{Treatment} = 0$; $F(3, 2862) = 1.52$, p value $=$ .678). In contrast to the control group, it does not provide evidence that the treatment group changed their online learning time compared to the baseline phase.

However, this raises another question: why are the effects on online learning in the three time segments seemingly different among the control groups? In particular, the control group of the elective course did not learn shorter in the middle of the semester, in contrast to the control group of the compulsory course. Another outcome variable sheds light on the potential reason, which is masked in the associated regression analyses: in both courses, the number of dropouts is fundamentally different. Specifically, the elective course had an overall dropout rate of around 39.7%, while in the compulsory course the overall dropout rate was approximately 14.4%*. This leads to two potential considerations for underlying mechanisms: on the one hand, it appears that many participants in the control group dropped out of the elective course, potentially due to procrastination issues. Those who remained did not necessarily learn less weekly on the learning platform compared to the baseline phase. On the other hand, the feedback intervention appears to have increased the weekly online activity of the treatment group in the middle of the course. As the participants of the treatment group invested more time in learning early on, they might have perceived higher levels of "sunk costs" when considering quitting the course, thereby lowering their chances of dropping out. Indeed, an additional analysis indicates a substantial difference in dropout rates between the study groups of the elective course. Among the participants who registered for the exam (each student must register to be able to complete the exam), the treatment group had a significantly higher rate of taking the exam (91.7%) compared to the participants in the control group (52.0%; $X^2 = 9.44$; p value = .006). Interestingly, the groups did not perform statistically differently in the exam (these analyses are not reported in the papers**). While the treatment effect on the exam-taking rate does not exist in the compulsory course, there was a statistically significant effect on exam performance (+6.5 points out of 90 points in the final exam; t value = 2.036, p value = .044). This indicates that the feedback actually mitigated procrastination in students (i.e., self-regulation issues) because in both courses the individual treatment group benefited regarding academic outcomes (i.e., either through reduced dropout rate or improved exam performance).

Apart from these insights on temporal dynamics, additional analyses from Paper VII that consider the data of both courses indicate that the overall online learning time is not influenced by the intervention and that the feedback intervention does not produce heterogeneous treatment effects based on certain variables (i.e., baseline online learning time and participants' self-report on the frequency of procrastination). In other words, such findings would indicate that only high-performing students could benefit from the intervention, which could exacerbate educational inequality.

Overall, the papers provide evidence that a feedback intervention based on social norms theory can have a large impact on participants' online learning behavior. More precisely, the impact of social normative feedback is not represented in the overall engagement on the platform but in the temporal changes of learner engagement, which suggest that students' procrastination behavior is reduced. By shedding light on this, Paper VI and Paper VII address the micro-micro link of

* The relatively high dropout rates could be the result of university-specific guidelines whereby students can attempt to pass a course as many times as they like. They are, however, limited in their attempts to the maximum period of study set by the university and how often the exam takes place.

** The elective course was not graded yet when Paper VI was written. For the same reason, Paper VII had no access to the exam data for the compulsory course.

the overall research framework. Because the papers investigate—for two courses with different samples of students—whether the treatment effects vary across different baseline variables, they also address the micro-macro link of the research framework. Notably, they provide better insights into how the intervention might work if it were implemented on a large scale (i.e., as a feature that is enabled by default in digital learning environments). Ultimately, the findings might motivate practitioners to implement this intervention in additional learning environments, which could contribute to a more in-depth understanding of the induced behavioral mechanisms. This could shed light on, among other things, the robustness of the treatment effects in a wider variety of courses (e.g., with courses that require different self-regulated learning skills) as well as treatment effects on other outcome variables (e.g., assignments).

6.2.4 Paper VIII: Toward an evaluation of an intervention that seeks to provide personalized feedback by modeling the heterogeneity of learners*

Highlights

- According to self-regulated learning theory, learners substantially differ in their self-regulated learning characteristics (e.g., in metacognition, management of internal resources, etc.). Thus far, this aspect has been often neglected in the design of feedback interventions for digital learning environments, yet theory implies that feedback interventions need to consider such characteristics to provide meaningful feedback to all learners.
- The paper argues that machine learning models could be leveraged to provide personalized feedback at scale that considers the heterogeneity of learners. On the basis of such machine learning models, the paper presents a feedback element that has been instantiated to provide feedback to higher education students.
- The study describes the expected results of an ongoing randomized controlled trial in light of self-regulated learning theory. For example, it is expected that individuals with lower levels of metacognition could benefit more from the feedback since it could empower them to better schedule their learning plan.

As demonstrated in Papers VI and VII, feedback interventions can have large effects on students' online learning behavior. However, self-regulated learning theory suggests that students inherently differ in their learning strategies and their personality (see, e.g., Pintrich, 2004), which also has important implications for such interventions. While some students might make sense of the previously outlined social normative feedback to increase their learning time, others might need more specific instructions to be meaningfully supported in their learning process. Howevere, existing literature that tests the effects of feedback interventions on online learning, often neglects the heterogeneity of learners in the design of interventions.

Paper VIII argues that digital learning environments might allow for feedback interventions that consider the heterogeneity of learners on a large scale: The associated data from past runs of a course (e.g., log data, socio-demographic variables) could be used to instantiate supervised machine learning models that predict course success for new learners (e.g., in the form of exam scores). For the next run of the respective course, student data could be passed to the previously trained machine learning models to predict exam performance at a certain point in time. By subsequently using counter-

* S. A. Günther, F. Haag, K. Hopf, P. Handschuh, M. Klose, and T. Staake (forthcoming). "A Feedback Component That Leverages Counterfactual Explanations for Smart Learning Support." In: *Digitale Kulturen der Lehre entwickeln – Rahmenbedingungen, Konzepte und Werkzeuge.* Edited by L. Mrohs, J. Franz, D. Herrmann, K. Lindner, and T. Staake. Perspektiven der Hochschuldidaktik. Wiesbaden, Germany: Springer VS.

factual explanations, a feedback component can deduce what action(s) each individual student should take next to improve their exam performance. More precisely, counterfactual explanations is a method that tries to determine what change in input features (here representing the data of each student) is required to bring the dependent variable of the respective machine learning model (here the exam scores) to a specific value (for an overview of the method, see Fernández-Loría et al., 2022). Thus, for a student with a current prediction of 60 out of 90 points, a counterfactual method could yield several suggestions for improving their exam performance to 70 points. It is important to note that the generated suggestions to improve one's success do not have to form causal relationships. Rather, they present a mathematically derived way of changing the outcome variable to a specific value.

Paper VIII presents an instantiation of a feedback component, which is shown in Figure 17. The component provides personalized feedback in the form of up to three weekly learning actions to improve students' exam performance. The predicted exam scores and the associated potential for improvement are not displayed to avoid discouraging students as the prediction is to varying degrees erroneous. Additionally, Paper VIII outlines the (planned) empirical evaluation of the feedback component, which follows a difference-in-differences experimental design.

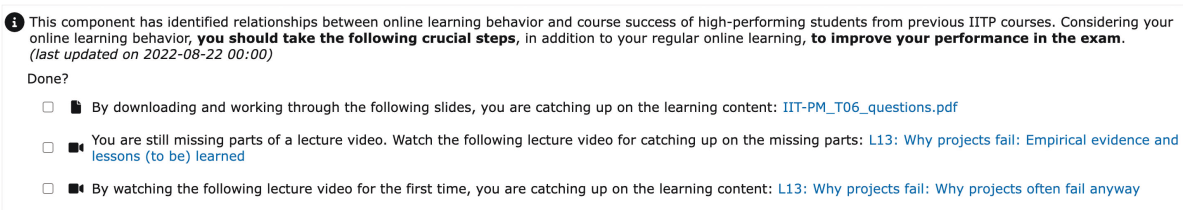


Figure 17: Personalized feedback to support online learning

To summarize, Paper VIII relates to the micro-micro link of the overall research framework by motivating the use of counterfactual explanations for the provision of personalized feedback in digital learning environments. The overarching approach has the potential to identify influential patterns in students' online learning data to master a specific course. By considering students' learning characteristics in this process (e.g., socio-demographic variables or antecedents of self-regulated learning), the component could recognize important differences between learners and adapt its feedback purposefully. In doing so, the feedback component might have the power to broadly support students' self-regulation in online learning. This feedback intervention is also interesting from a theoretical perspective. For example, it could be that metacognition moderates the effects of feedback on online learning. In other words, the feedback component might have more desirable effects on those students who have relatively weak metacognitive skills and hence have issues organizing their learning. Likewise, it could be that the feedback component especially helps students with procrastination problems that result from low self-efficacy beliefs. An underlying reason for this is that the learning actions of the feedback component are relatively straightforward for students to execute. The feedback component might therefore be an initial prompt for students to engage in learning. By following these personalized steps, the component might improve students' self-efficacy beliefs, which could motivate them to learn even more instead of procrastinating. Nonetheless, these relationships, which would be valuable for theory development, still need to be explored in future work.

6.2.5 Paper IX: Toward quantifying cause-effect relationships of feedback interventions on antecedents of low-involvement behavior*

Highlights

- Behavioral interventions such as feedback appear to affect multiple antecedents relevant for behavior change, yet, the related literature lacks mathematical formalizations that quantify such cause-effect relationships through field experiments.
- The present paper proposes an experimental evaluation to identify and quantify certain behavioral mechanisms associated with feedback intervention theory and related theoretical models.
- The experiment aims to investigate cause-effect mechanisms of feedback and social norms on hand-washing duration. To investigate the interplay of feedback and social norms, the experiment is based on a 2 (feedback, no feedback) $\times$ 2 (salient social norms, no social norms) design.
- Lastly, the paper outlines hypotheses on the anticipated effects of these interventions.

Across various domains, feedback interventions have shown considerable effects on everyday activities to improve associated personal or organizational outcomes (see, e.g., Bravata et al., 2007; Sanguinetti et al., 2020; Van der Kleij et al., 2015). More specifically, while small improvements to everyday activities such as reducing heat energy usage in the shower, driving more efficiently, or handling production machines with care may appear unimportant at the micro level, at the macro level such activities accumulate and shape relevant outcomes for organizations and policymakers. Despite the relevance of feedback interventions, there is a lack of understanding of how such interventions induce behavior change as associated behavioral campaigns often affect multiple determinants of behavior change (see, e.g., van Valkengoed et al., 2022). As an illustration, feedback interventions operate on several levels, such as by increasing attention on the target behavior, outlining discrepancies between individuals' aspirations and their actual behavior, and creating a sense of specific outcomes of the respective behavior (e.g., CO_2 intensity of showering). While there is a plethora of papers from psychology and behavioral economics providing suggestions on how interventions may influence the antecedents of behavior (e.g., Bandura, 1977; Kluger and DeNisi, 1996; Michie et al., 2011), there is a lack of mathematical models that formalize intervention-associated theories. A formalization in the form of a model allows hypotheses to be tested from established theories on behavioral interventions (e.g., feedback intervention theory) and to better quantify the effects of individual interventions on associated antecedents of behavior. This potentially allows deeper understanding of the conditions needed for interventions to create the highest impact on behavior change.

To integrate intervention-associated theories, Paper IX proposes a structural equation model that integrates many important behavioral theories associated with feedback interventions on everyday behavior, which is shown in Figure 18. On the one hand, the model formalizes propositions from feedback intervention theory: behavior is influenced by the antecedents

* C. Stingl, S. A. Günther, and T. Staake (2022). "The Behavioral Mechanisms Behind Feedback – A Preliminary Model for Quantifying Cause-Effect Relationships." *Proceedings of the 30th European Conference on Information Systems*. Timișoara, Romania.

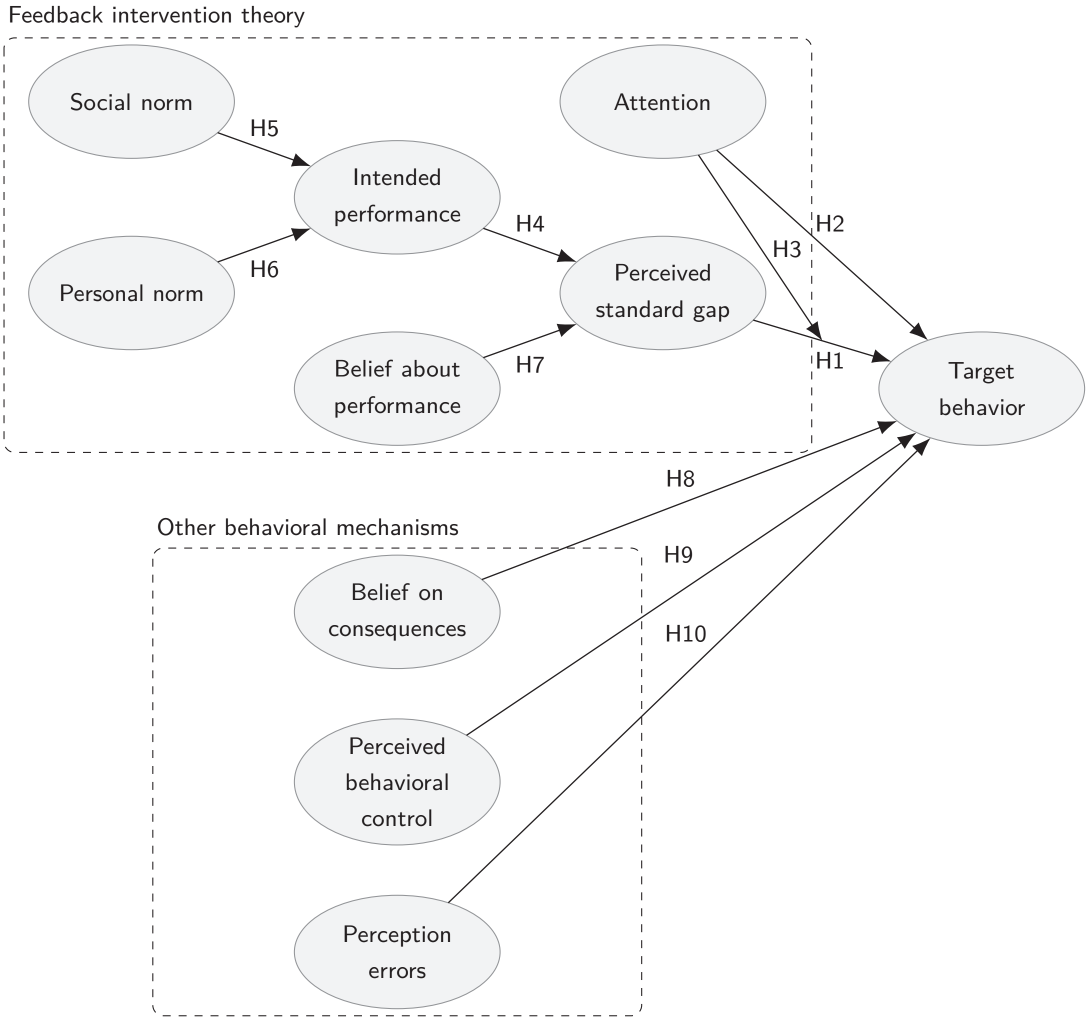


Figure 18: Path diagram of the developed structural equation model

of attention toward the behavior, a perceived standard gap that originates from one's belief about performance, and one's intended performance. The intended performance level results from the social norm, which consists of expectations of what behavior is socially approved or disapproved of, and the personal norm. Personal norms relate to a personal moral obligation toward behavior that results from personal value orientations (e.g., altruistic, ecological, or egoistic value orientations) and explain why individuals often resist social norms and act more according to their moral compass. On the other hand, the models build on several antecedents of behavior from other theories, such as the associated beliefs about consequences and the perceived behavioral control to regulate a behavior. Perception error is a new construct that is initially introduced in the paper. The construct of perception error accounts for potential differences between perceived task performance and objective task performance. Such a perception error might be that a person believes they are showing good hand hygiene in accordance with their personal goal while actually performing below their goal.

In addition to the model, the paper describes its planned evaluation in the form of a field experiment that uses a 2 (no feedback, feedback) $\times$ 2 (no norm, norm) experimental design to test the effects of different treatment conditions on the outlined antecedents of behavior. To do so, a system is set up that measures water and soap usage (the system is described in detail in Section 4.3). After washing their hands, individuals are asked to answer a few questions associated with the structural equation model when they leave the respective bathroom. In the event that individuals opt out, the handwashing data is discarded. Otherwise, the questionnaire data is linked to the behavioral measurements and allows testing of the model as well as of specific hypotheses regarding how the interventions affect the modeled antecedents of behavior.

To summarize, Paper IX responds to the micro-micro link of the research framework by proposing a model that potentially allows for better testing of behavioral propositions from different theories. In other words, the structural equation model might help to better indicate how changes to individuals' decision-making context could shape subsequent behaviors. Apart from this, the model could also help to shed light on which antecedents are most influential across various everyday behaviors, which could inform practitioners about creating more effective behavioral interventions.

7 Contributions and implications

By studying important questions around each link of the adapted belief–action–outcome framework from Section 1, the nine papers of this dissertation provide several insights into how IT-based opportunities for behavior change could have meaningful real-world impact on the micro and macro levels. This section first summarizes the associated contributions to literature before outlining practical implications.

7.1 Contributions to literature

The results of the nine papers are divided into two different chapters. Chapter 1 focuses on testing the technical or organizational feasibility of feedback systems (the macro-micro link of the overall research framework). Chapter 2 focuses on contributing to theory by increasing our understanding of the effects of feedback interventions and their interplay with other behavioral interventions (the micro-micro link). By taking differences in individuals into account, Chapter 2 also sheds light on the potential effects of feedback interventions at the macro level (the micro-macro link).

7.1.1 Chapter 1: Instantiation and validation

Key contributions of Chapter 1

– The chapter provides evidence that appliances, fixtures, and individuals leave unique patterns in water consumption data that can be used to provide more specific or personalized feedback to individuals.
– By presenting a feedback system to improve hand hygiene, the chapter also indicates that digital transformation enables novel opportunities for feedback provision that might help to overcome adoption issues reported on existing feedback systems.

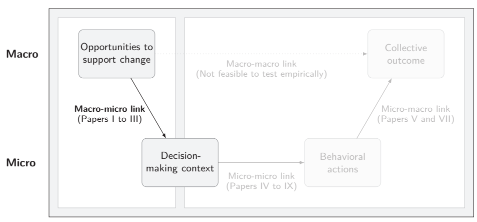
Chapter 1 of the overarching research framework is specifically concerned with developing and adapting IT tools to facilitate more effective feedback interventions. Prior research in the environmental domain suggests that feedback appears to be most effective when it is clearly formulated, delivered on actions that individuals can easily influence, and personalized (e.g., Costa and Kahn, 2013; Sütterlin et al., 2011; Tiefenbeck, 2014). Papers I and II contribute to the related literature by investigating whether more specific feedback interventions can be created by making more sense of measurement data. The joint problem both papers address is that, in many cases, there is only one sensor that records data on multiple variables of interest. The papers conjecture that machine learning algorithms might be able to extract these variables from the data.

By using supervised machine learning algorithms, Paper I shows that different water-consuming fixtures and appliances in a household can be identified by a single water sensor that is attached to the main water pipe of the household. This can then enable feedback systems that provide more tangible feedback. Such a system could reveal where the greatest potential for resource conservation and financial savings lies, which might increase conservation effects compared to feedback that only presents information on overall consumption to users (as demonstrated by Gerster et al., 2020 for feedback on electricity use). Similarly, Paper II demonstrates that individuals can be identified from their water traces in the shower, thereby empowering personalized feedback in multi-person households. For example, the amphiro shower meter could automatically adapt to individuals by presenting feedback differently depending on the identified user at the end of a shower. Likewise, downstream applications, such as the accompanying mobile app from Amphiro AG, could use this information to provide water-saving challenges to individuals in a household. It is important to note that Paper I advances existing research by demonstrating that more water-consuming fixtures and appliances can be reliably predicted with consumption data from a central sensor. Paper II extends the literature by demonstrating that supervised machine learning algorithms can recognize person-specific traces in water consumption data to identify individuals.

While Papers I and II aim to deliver more specific feedback to individuals, Paper III seeks to identify organizational barriers to feedback systems on hand hygiene in order to stimulate new technological developments. Specifically, the associated literature review of Paper III suggests that costs, privacy, usability, and accessibility issues are the main reasons why feedback systems for hand hygiene often fail. The paper also outlines recommendations that (future) feedback systems should comply with to empower more effective feedback. Beyond that, the paper specifically contributes to the literature by presenting a prototype that makes use of behavioral data to effectively overcome the identified barriers to feedback provision. To this end, the prototype samples the water and soap usage of handwashing events and presents real-time feedback to individuals.

Taken together, the three papers in Chapter 1 support that technical developments not only enable feedback interventions that are potentially more effective for behavior change but also help to overcome organizational barriers that may hinder the diffusion of feedback systems. Table 9 summarizes the contributions to the literature (macro-micro link).

Table 9: Contributions regarding the macro-micro link

<table>
<tr><th>Research context</th><th>Contribution</th></tr>
<tr><td>
</td><td>– Testing the feasibility of feedback systems on water usage:
- Demonstration that water consumption, measured by a single sensor that is attached to the main water pipe of a household, can be attributed to the corresponding water-consuming fixtures and appliances.
- Demonstration that some water consumption events have person-specific traces that could be used to provide personalized feedback on water consumption.
– Advancement of handwashing monitoring systems:
- Identification of barriers that current handwashing monitoring systems in the literature face.
- Provision of recommendations for overcoming the barriers.
- Exemplary instantiation of a prototype (i.e., feedback system) that conforms to the presented design recommendations.</td></tr>
</table>

7.1.2 Chapter 2: Evaluation and recommendations

Key contributions of Chapter 2

- The chapter provides evidence that corporate programs, which internalize the negative externalities associated with consumers' demand (i.e., greenhouse gas emissions), might actually encourage consumers to increase their resource use by alleviating their feelings of guilt. Feedback can counteract these adverse behavioral effects by potentially reactivating consumers' perceived responsibility for their environmental actions.
- The chapter indicates that performance-contingent financial incentives do not crowd out individuals' motivation for environmental protection but rather promote resource conservation, which refutes the predictions of motivation crowding theory. Moreover, feedback is dominant over financial incentives, meaning that the financial incentives do not amplify behavioral effects when feedback is already in place. Consistent with resource allocation theory, interventions might face diminishing returns in directing individuals' attention to resource conservation. Thus, more interventions are not necessarily better.
- In the learning domain, the chapter puts forth the argument that mixed effects of descriptive comparative feedback can be explained through the lens of social norms theory: some learners might be discouraged from feedback that resolves their erroneous beliefs that others are also active on the digital learning platform. In addition, the chapter provides suggestive evidence through two randomized controlled trials that injunctive norms counteract such adverse effects.

Instead of presenting novel technical possibilities for feedback, the papers of Chapter 2 focus on advancing understanding of how changes to individuals' decision-making context can influence their subsequent behaviors (addressing the micro-micro link of the research framework). In particular, Papers V to IX attempt to both reduce personal or societal issues through the use of behavioral interventions and contribute to theory development by investigating underlying behavioral mechanisms. It is important to note that this research is positioned in three different domains (environment, education, and health), as described below.

Prior research in the environmental domain has suggested that offsetting programs may lead to unintended consequences. After paying a fee to mitigate negative externalities associated with their demand, consumers may feel entitled to increase their demand, which might then increase overall carbon emissions. The contribution of Paper IV in this regard is twofold. First, the paper evaluates the behavioral response to such programs when consumers do not pay for the offsetting, which might increase the risk of adverse behavioral responses (i.e., higher consumption levels). Specifically, prior research has indicated that adverse effects are likely if the price of offsetting for consumers is low (Harding and Rapson, 2019; Jacobsen et al., 2012; Kotchen and Moore, 2008). Second, Paper IV presents the first study that evaluates a countermeasure to a potential adverse behavioral response of consumers by following a 2 (no offset program, offset program) × 2 (no feedback, feedback) design. Overall, the paper determines that climate-neutral products and services can indeed induce higher demand from consumers. However, feedback can completely mitigate this adverse behavioral effect, which raises new theoretical questions on the associated antecedents of behavior. For example, it is conceivable that the offsetting program may lower the ascribed responsibility of the consumers for environmental outcomes, reducing associated curtailment efforts. Likewise, consumers may perceive lower feelings of guilt as their consumption is now climate-neutral. Feedback, in contrast, might reactivate the consumers' perceived responsibility by highlighting that consumers—despite the climate neutrality of their consumption—are still using precious environmental resources (in this case, heat energy and water). While the study does not empirically test these possible explanations, it nonetheless provides a promising starting point for scholars to study the exact behavioral mechanisms behind the interplay of consumption feedback and offsetting programs.

In the environmental domain, there has been a longstanding debate about whether financial incentives might have hidden costs. Instead of motivating resource conservation, it is hypothesized that financial incentives might even dampen conservation efforts (see, e.g., Schwartz et al., 2019), which could also reduce the effects of feedback interventions. Despite this omnipresent fear, studies that test the real-world impact of financial incentives are scarce and ignore variables that might mask crowding-out effects (e.g., intrinsic motivation at baseline, financial situation of participants). By using a 3 (financial incentives, environmental appeals, no cues) $\times$ 2 (no feedback, feedback) experimental design and controlling for different potentially confounding factors, Paper V demonstrates that financial incentives for resource conservation do not harm curtailment efforts. Conversely, the results of the field experiment suggest that small, performance-contingent financial incentives can lead to modest resource conservation effects. Furthermore, Paper V contributes to the related literature by providing an in-depth comparison of the effects of multiple behavioral interventions. Related work offers scarce information on the role of motivational strategies in promoting pro-environmental behavior. Existing studies focus almost exclusively on one strand of intervention strategies: (i) financial incentives and environmental appeals or (ii) feedback interventions with either financial incentives or environmental appeals. As a result, the current body of research provides no clear picture of the cause-effect relationships between these strategies to elucidate whether financial incentives and environmental appeals induce behavior change to the same extent and what the marginal effect of adding feedback information would be. Additionally, it remains unclear whether financial or appeals are required to fully unlock the potential of feedback interventions or if they lead to lower levels of behavior change by shifting the focus away from the feedback, as feedback intervention theory may suggest. As outlined above, Paper V posits that feedback seems to be the dominant source of behavior change. Additional cues on top of the feedback information did not lead to substantially lower or higher conservation effects; however, the cues appear to decrease habituation to feedback, meaning that feedback has a more lasting impact on resource conservation. Finally, Paper V also contributes to the related literature by theorizing about the underlying behavioral mechanisms on a more general level. Specifically, the paper suggests that all its observations are consistent with resource allocation theory (Kanfer and Ackerman, 1989). This theory maintains that the incremental provision of interventions faces diminishing returns on the allocation of attention from individuals and therefore on their conservation behavior. For this reason, financial incentives led to conservation, while financial incentives in combination with feedback did not produce higher conservation levels. Ultimately, the paper indicates that such general attention-based theories deserve more attention in behavioral research on environmental sustainability.

Apart from these contributions to behavioral research on environmental sustainability, the dissertation also contributes to the area of digital learning. In this domain, more and more studies have evaluated the impact of comparative feedback, often motivated by the influential work of Festinger (1954). However, empirical evidence illustrates that such feedback can backfire, namely by discouraging online learning on the associated platforms (Li and Zhang, 2016). The present studies extend related work by investigating the impact of a comparative feedback intervention in two university courses based on social norms theory (Berkowitz, 2005). The papers hypothesize that comparative feedback may backfire through the lack of an injunctive component that provides approval or disapproval of one's behavior. Besides a descriptive component that provides feedback on online learning time, the feedback component also features an emoji showing approval or disapproval of one's behavior. In sum, Papers VI and VII extend previous research by demonstrating that a relatively simple intervention can considerably mitigate procrastination in digital learning environments. In both experiments, the online learning time was considerably higher in the middle of the respective course in response to the feedback (i.e., in the period in which new course content was still being published). Most importantly, an adverse behavioral response (e.g., lower login frequencies) to the feedback was not observed. While the papers do not provide sufficient evidence that the injunctive component is required for the feedback to not backfire (as the experiments lack a descriptive norm only condition), the experiments demonstrate that social influence can have a highly desirable impact on students' behavior, even in digital learning environments that feature little to no social interactions. Likewise, the studies provide a promising starting point for scholars to explore the role of the injunctive norm in encouraging online learning. Such a study could feature, for instance, a 2 (no descriptive norm, descriptive norm) $\times$ 2 (no injunctive norm, injunctive norm) design to identify the individual and the combined effects of descriptive and injunctive norm messages.

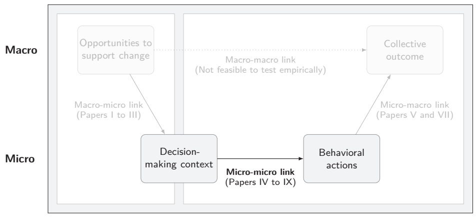
Given that self-regulation theory suggests that students differ in their learning strategies and their personalities (see, e.g., Pintrich, 2004), it is important to note that social normative feedback might not have the power to support all kinds of learners. Specifically, some learners might not be able to make sense of this feedback for their learning. As the heterogeneity of learners is largely neglected in studies testing the effects of feedback on online learning, the dissertation also tackled this topic. With Paper VIII, this dissertation presents a feedback component that has the power to automatically adapt to courses with online content and to their participants. The component uses data from previous runs of a course to predict participants' course success (i.e., points in the exam). By subsequently applying counterfactual explanations to the associated machine learning models, the feedback component can provide personalized learning suggestions to course participants to improve their course success. Beyond presenting this feedback design, Paper VIII outlines its planned evaluation, which follows a difference-in-differences experimental design. This is a promising approach to providing personalized feedback at scale to learners in online learning environments, though its potential behavioral effects have not yet been explored. An investigation of the behavioral responses to the feedback component can increase theoretical understanding of its effects. For example, the feedback component may especially help those with weaker metacognitive learning skills as they typically face more difficulties in organizing their learning. Such mechanisms need to be explored in the future.

As seen in the outlined papers in Chapter 2, feedback interventions often encompass different elements that target several behavioral barriers to behavior change. In addition to feedback on behavior, such interventions often try to increase knowledge on the consequences of non-compliance (e.g., environmental damage through climate change) or provide individuals with information on social norms (e.g., as in the outlined studies on online learning). Paper IX contributes to the related literature by proposing a model that aims to quantify the cause-effect relationship between behavioral interventions and key antecedents of low-involvement activities (e.g., showering, car driving, handwashing, brushing teeth). To this end, the paper provides an integrated structural equation model that incorporates many related theories, with the overarching goal of testing theoretical propositions from the literature (i.e., from feedback intervention theory). While the paper does not yet evaluate the outlined model, it provides a research plan in the form of an experiment with a 2 (no feedback, feedback) $\times$ 2 (no norm, norm) design on handwashing that is planned for the future. In this way, the paper contributes to the related literature not only by addressing an important topic (i.e., improving hand hygiene through behavioral interventions) but also by presenting a concrete concept for testing and quantifying the effects of specific behavioral mechanisms.

To summarize, four out of six papers in Chapter 2 demonstrate that relatively simple feedback can induce a substantial change in behavior. As demonstrated, feedback helps to address personal or societal issues (e.g., mitigating procrastination, fostering resource conservation) as well as to eliminate adverse consequences of other interventions (i.e., adverse effects of carbon offsetting). Interestingly, feedback information appears to be the dominant factor facilitating behavior change across all the papers that combine consumption feedback with information on environmental consequences or financial incentives. These studies thus provide empirical support for feedback intervention theory, which describes that the combined effect of feedback information and other motivational cues is not additive. While feedback intervention theory postulates that alternative cues moderate the effects of feedback information (e.g., by attenuating the effects of feedback), the studies do not provide evidence of this. A potential explanation is that feedback information is much more salient than the cues in the field and therefore received more attention, which might reduce the moderating effects of the cues in contrast to more controlled experimental settings (e.g., in a laboratory experiment). Table 10 summarizes the contributions of Chapter 2 to the literature (micro-micro link).

Additionally, Chapter 2 is concerned with increasing understanding of the extent to which the observed changes at the micro level may translate to the macro level (e.g., the population level), thus addressing the micro-macro link of the overall research framework. This is an important aspect to investigate as differences in individuals might fundamentally lead to different collective outcomes (see, e.g., Cintron et al., 2022; Murakami et al., 2022). For

Table 10: Contributions regarding the micro-micro link

Research context	Contribution
	– Evaluation of the consumer response to carbon offsetting in the presence and absence of feedback: - There is robust evidence that feedback can mitigate the adverse behavioral responses from consumers (i.e., higher consumption levels) that may ensue when organizations market their products and services as climate-neutral. - Raises new questions on how the related interventions (i.e., carbon offsetting and feedback) influence associated antecedents of environmental behavior. – Extensive field experiment on the interplay of different behavioral interventions (feedback, performance-contingent financial incentives, moral appeals): - There is strong evidence that financial incentives for resource conservation do not crowd out intrinsic motivation. - Inconsistent prediction of the standard economics model when financial incentives and feedback are combined: incentives appear not to unlock higher conservation levels. - Overall effects are consistent with the predictions of resource allocation theory: diminishing attentional effects of behavioral interventions on resource conservation. – Evaluation of two field experiments that test the impact of social normative feedback on online learning. – Presentation of ongoing evaluation of a novel intervention that could provide personalized feedback on online learning at scale. – Provision of a preliminary model and a corresponding experimental design to quantify cause-effect relationships of feedback interventions on everyday behaviors (e.g., handwashing).

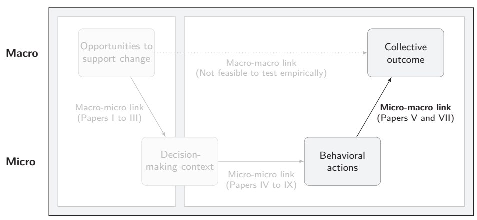
example, if an intervention only induces resource conservation for individuals with a strong pro-environmental value orientation, the behavioral outcome of an intervention may fundamentally be different when it is rolled out to a broader population of interest. To explore such mechanisms, Papers V and VII conduct additional analyses in the papers. These analyses do not rule out that heterogeneous effects are overlooked (e.g., as the sample size may be too low powered), but they provide important insights that papers from related work often do not investigate.

As outlined, Paper V tests multiple interventions side by side (i.e., environmental appeals, financial incentives, feedback). Prior research did not compare whether the behavioral responses to the given set of interventions depended on differences in the individuals. For instance, it could be that environmental appeals are only effective for individuals with a strong intrinsic motivation for resource conservation, while others may require financial incentives for behavior change. In this case, it would be cost-effective to approach individuals differently when rolling out a campaign for resource conservation. Across the different analyses, the analyses document that there is not a strong heterogeneity in the behavioral responses. Environmental appeals fostered conservation neither for individuals with a strong environmental attitude (the proxy variable of the paper for intrinsic motivation) nor for people with a relatively low environmental attitude. Likewise, the financial incentives stimulated relatively the same level of resource conservation, regardless of the financial situations of the individuals. Among these additional analyses, feedback remained the most effective intervention in inducing behavior change.

The contribution of Paper VII to the related literature is twofold: first, the paper compares both experiments on social normative feedback that were run in different courses. There is only scarce research in online learning that tests whether the results are replicable, especially in a university setting (see, e.g., Sønderlund et al., 2019). The paper indicates that, despite the fact that the course participants pursued different degrees (bachelor's level vs. master's level) and studied different course materials, the behavioral response is comparable: the feedback intervention led to substantially higher online learning time and online activity in the middle of the semester. Second, the paper identifies whether the behavioral response to the feedback is heterogeneous across the subgroups of the experiments. Prior research has often only reported average treatment effects, neglecting to consider whether behavioral interventions might lead to heterogeneous behavioral responses among students. By considering potentially important baseline characteristics, the paper does not find heterogeneity in the behavioral responses. Hence, there is evidence that all subgroups of participants benefited equally from the feedback intervention.

In sum, the two papers suggest that feedback is a powerful instrument for behavior change and does not induce substantially heterogeneous behavioral reactions. Still, this dissertation stresses that it is important to investigate such aspects to increase understanding of underlying behavioral mechanisms. With the scalability of IT-based interventions, it is feasible to personalize interventions depending on recipients' characteristics; this can not only enhance the resulting behavioral effects but also promote general acceptance of such interventions. Table 11 summarizes the contributions of Chapter 2 to the literature (micro-macro link).

Table 11: Contributions regarding the micro-macro link

<table>
<tr><th>Research context</th><th>Contribution</th></tr>
<tr><td>


</td><td>
– Extensive study on the interplay of different behavioral interventions (feedback, performance-contingent financial incentives, moral appeals):
- There is suggestive evidence that more affluent individuals respond more strongly to feedback when financial incentives are in place.
- Comparing the effects of the three interventions, feedback appears to be more effective than small financial incentives or environmental appeals for behavior change, irrespective of considered differences in the individuals (e.g., environmental attitude, financial situation).
– Evaluation of two field experiments that test the impact of social normative feedback on online learning:
- When comparing the two experiments with individuals pursuing different study degrees (bachelor's level vs. master's level), the feedback appears to consistently mitigate procrastination.
- When controlling for potentially important baseline characteristics of the individuals in each experiment (e.g., baseline online learning time, participants' self-report on the frequency of procrastination), the behavioral response to the feedback is not heterogeneous.
</td></tr>
</table>

7.2 Practical implications

Summary

- The dissertation underlines that IS and IT systems have the potential to improve numerous personal and societal issues through the provision of feedback.
- The effects of feedback interventions are large and preferable to those of other interventions (e.g., in cost efficiency or effect size). Moreover, there is no evidence that the behavioral responses to the feedback interventions would be different in slightly different settings than those evaluated.
- The dissertation thus encourages policymakers and organizations to implement these interventions in practice to induce behavior change at scale.

Beyond these contributions to literature, the two chapters of this dissertation also have several important implications for practice.

This dissertation shows that relatively simple feedback information is a powerful instrument for behavior change across different use cases. For instance, the papers in the environmental domain (i.e., Papers IV and V) demonstrate that feedback information can induce resource conservation in the absence of financial benefits for behavior change. Notably, other behavioral interventions for resource conservation, namely environmental appeals and financial incentives, were clearly less effective than consumption feedback, and they did not enhance the effects of feedback. As feedback information is potentially more scalable than providing financial incentives to induce behavior change, this dissertation highlights the potential of feedback interventions to reduce resource consumption in everyday behavior. These insights are very timely for organizations that are increasingly focusing on environmental sustainability (Amankwah-Amoah and Syllias, 2020). On a related note, the dissertation has another important implication for organizations that are marketing their products and services as climate-neutral. As shown, individuals may increase their demand in response to such programs, which could increase organizational costs. This might be especially true in scenarios where individuals do not bear the monetary consequences associated with their higher demand (e.g., hotel guests pay a fixed room rate for their consumption, employees do not bear the costs of their fuel use at work). Although the dissertation cautions organizations to transparently communicate their climate-neutrality measures for products and services, it also demonstrates that feedback can mitigate such adverse consequences in practice.

While one could argue that the previous examples are possible only through the intensive development of custom hardware (e.g., the shower meter), this dissertation provides evidence that this is not the case. Today, software and commercially available hardware (e.g., smartwatches, smartphones) are presenting an increasing number of interfaces that can empower feedback interventions. For instance, online platforms can be extended through software components that measure behavior and subsequently provide feedback to users. The studies on online learning (Papers VI and VII), support that even a small change to the main page of a digital course can considerably encourage online learning and mitigate procrastination among users. Once such developments have been sufficiently empirically evaluated and technically matured, they could be embedded in thousands of software solutions, potentially having a substantial societally desirable impact on behavior. To spark such processes, studies that aim to test theoretical propositions related to feedback, like Paper IX, can be especially helpful as they could provide a better understanding of the design of effective feedback interventions. While the dissertation has found evidence for some heterogeneous effects (e.g., high baseline consumption is associated with higher conservation effects through feedback), further studies could shed more light on the role of differences in individuals that influence the outcomes of behavioral interventions, which would further inform practice. Organizations and policymakers could then maximize the impact of behavioral campaigns by targeting those individuals who would respond most favorably to the respective treatment.

Furthermore, the dissertation highlights novel possibilities for feedback provision using IT to improve real-world outcomes. As outlined in the theoretical foundations section, feedback is most effective when individuals can easily interpret it and respond to it. In the environmental domain, individuals are usually not well versed in environmental outcomes. For most consumers, the units the outcomes are measured in are rather abstract (e.g., CO_2 emissions, energy in kWh), making it burdensome and time-intensive to make the most of the recommendations for behavior change. As demonstrated, digital technologies could make feedback information much more tangible to better support behavior change. Based on Paper I, an application could be built that displays the consumption of different water consumption events in a household; this could rule out perception errors on the resource use of different water-consuming events. In other words, individuals might notice, for example, that one's shower use is much more resource-intensive than the hot water consumption of a dishwasher. Therefore, individuals may direct their conservation efforts to fixtures or appliances where they could save the most. Additionally, novel features, such as the person identification in Paper II, might be used in practice to increase long-term adherence to associated feedback applications. For the shower meter, a mobile app already exists. This app could automatically identify individuals once the machine learning component of Paper II is set up. Hence, the app could provide savings challenges across the members of a household, potentially increasing attention to hot water use and enhancing conservation efforts. Utilities, which are increasingly required to increase energy efficiency (see, e.g., Alberini and Towe, 2015; European Union, 2021), could follow similar steps to induce higher conservation levels among consumers.

Lastly, this dissertation also underlines that technical developments from one domain (i.e., environmental sustainability) might be successfully used for feedback provision in a different domain (i.e., health). The core technical foundation of the shower meter (used in Papers II, IV, and V) and the prototype presented in Paper III are relatively similar. Both contain at least one turbine that harvests energy out of the water flow and thereby collects measurement data from the water flow through a microcontroller. While the shower meter uses the measurement data to provide consumption feedback, the presented prototype uses it for feedback on how well individuals wash their hands. This is just one example of such technical spillover effects. Widespread sensors that measure environmental resource use in households (e.g., smart meters) can be used for many other practical applications. For instance, sensors could identify specific events in households (e.g., unusually low or high consumption levels, water pipe bursts). In addition to damage prevention, this might be used by insurance companies to detect and automatically verify insurance claims. Likewise, similar information could be useful for elderly care, where changes in individuals' daily routines can be signs of diseases such as dementia (see, e.g., Enshaeifar et al., 2018). Hence, there may still be untapped potential in existing building blocks in organizations to create novel IT-based applications that can improve personal, organizational, or societal outcomes by delivering feedback. Table 12 summarizes the implications of this dissertation for practice.

Table 12: Detailed summary of practical implications

Scope	Implications
Immediate impact on behavior	– Non-monetary interventions, namely feedback, can induce substantial resource conservation in the absence of financial benefits for behavior change. – Feedback can induce larger conservation effects than reasonably sized performance-contingent financial incentives. The provision of additional financial incentives does not enhance the conservation effects of feedback. – Climate-neutral products and services can increase consumer demand. Feedback might mitigate this increase in consumption. – Social normative feedback, integrated into the main page of a course hosted on a digital learning platform, can mitigate the procrastination of course participants.
Expected impact at the macro level	– Evidence that the impact of the studied interventions is relatively insensitive to the tested characteristics of individuals. – Individuals with high baseline consumption benefit the most from the evaluated interventions.
Potential of technical developments	– Supervised machine learning algorithms can extract meaningful patterns from consumption traces. These patterns might be useful for real-world applications in creating new services for customers. – Technical developments in one domain (e.g., resource conservation) can facilitate novel behavioral interventions in other domains (e.g., handwashing feedback).

8 Limitations and future research

Summary

- The dissertation is primarily limited in the way that the feedback interventions have been tested within specific settings (samples, environments in which the feedback intervention is delivered, etc.).
- This might raise questions on the transferability of the associated findings (e.g., the behavioral effects).
- Given the high external validity of field experiments, the findings can be expected to be similar even in slightly different settings. This is especially true for the experiments on resource conservation, for which literature often provides supporting evidence.

Despite the best efforts in conducting the outlined research, this dissertation is limited in various ways. This section outlines the four main limitations of this dissertation before shedding light on future research avenues.

Although the dissertation provides important contributions to all links of the adapted belief–action–outcome framework, the underlying research is limited as it does not meaningfully address all three links simultaneously, which could provide a more coherent understanding of how novel opportunities for feedback interventions could improve micro- and macro-level outcomes. For example, the dissertation does not document whether a feedback system that is empowered by the papers of Chapter 1 (i.e., Papers I, II, or III) actually leads to the desired behavioral effects (micro-micro link), nor does it investigate how associated micro-level effects might project to the macro level (micro-macro link). The overarching reason for this is that this dissertation addressed more pressing research questions that otherwise would have been impossible to evaluate.

Additionally, the papers in Chapter 1 might have overestimated the potential of the research approaches for feedback provision. For example, the proposed hand hygiene system of Paper III was derived from literature, but there could be barriers for feedback systems that have not been mentioned there (e.g., legal aspects). In a similar vein, Papers I and II have focused on the technical feasibility of feedback provision by using preliminary hardware. Even if the feedback facilitated by these papers leads to the desired behavioral effects, there are many further steps required for the associated feedback systems to leave the prototype stage (e.g., cost-benefit analyses, privacy requirements, etc.) and ultimately have substantial real-world impact.

While the papers in Chapter 2 (i.e., Paper V to IX) have indicated that feedback interventions can mitigate important personal or societal issues by leveraging real-world behavioral data, there are concerns regarding their practical and theoretical implications. More precisely, field experiments are known for having relatively high external validity at the cost of internal validity, reducing the extent to which observed behavioral changes could be attributed to the tested behavioral interventions. For instance, there could be an indefinite number of unobserved confounding factors that may have influenced the behaviors of study participants independently of the interventions. At the same time, it is important to note that it might not be feasible to explore the behavioral responses to interventions over a longer time in a more controlled environment (e.g., observing individuals' pro-environmental behavior over multiple days in a laboratory setting). The dissertation addressed these concerns by using methodologically sound approaches to enhance the internal validity of its field experiments (e.g., the difference-in-differences design). Nonetheless, several concerns remain and could be eliminated with further experiments (such as with the experimental design presented in Paper IX).

The dissertation is further limited in its analyses of Chapter 2 exploring how the observed changes at the micro level might project to the macro level (concerning Papers V and VII). Specifically, both studies explore the behavioral responses to interventions with specific samples that may have a narrow range in their associated socio-demographic variables to detect heterogeneous effects. For example, the majority of individuals in Paper V have less than €100 in budget left for their living expenses per month, which indicates that it may be challenging to extrapolate specific findings on financial incentives (i.e., "small financial incentives induce resource conservation") to more affluent samples of individuals. It is possible that more affluent individuals do not react to such small incentives. In contrast, the related findings on the effects of feedback information are likely to be more robust: across studies on resource conservation in the shower (see, e.g., Fang et al., 2023; Tiefenbeck et al., 2018; Tiefenbeck et al., 2019), feedback has consistently led to impressive resource conservation effects. Another limitation is associated with the statistical power of the related experiments in this dissertation. The sample sizes in Paper V and Paper VII might be too small to detect more meaningful differences in behavioral responses. However, the associated analyses are still valuable as they could have revealed considerable differences in behavioral responses (e.g., adverse effects of financial incentives on individuals with high intrinsic motivation for resource conservation), which could have raised doubts about the "careless" use of related interventions in practice.

There are many promising avenues within the overall research framework for future research that go beyond addressing these limitations. One avenue relates to the currently underresearched topic of how individuals respond to feedback systems that provide feedback generated by machine learning approaches. Specifically, if the approaches of Paper I and Paper II were used for feedback provision, the feedback given to individuals would sometimes be erroneous. Here, it must be determined how the feedback could be presented so that recipients understand that the feedback does not have to represent the truth. IS research has recently explored consequences of making underlying decisions and outcomes of artificial intelligence understandable for users (e.g., Fernández-Loría et al., 2022; Storey et al., 2022), and this might therefore be an interesting research topic. Another promising research avenue would be to study machine learning approaches that could make the initialization phase of such feedback systems obsolete. As outlined, supervised machine learning algorithms require training data for feedback provision, which requires cooperation from potential feedback recipients. It is reasonable, however, that the efforts for such an initialization phase should be kept at a minimum as individuals could otherwise lose interest in the feedback systems. A potential way to overcome the initialization phase could be to use unsupervised machine learning approaches to detect patterns in new data without requiring training data.

Another promising research avenue is to increase understanding of how the behavioral effects of feedback depend on the respective characteristics of the behavior the feedback addresses.

For example, across studies on resource conservation, effect sizes vary substantially depending on the addressed behavior. Knowing which factors predict the effect sizes of interventions can inform future behavioral campaigns for resource conservation, which could then better select promising target behaviors. Similarly, further research is required to test the long-term effects of feedback, especially in the environmental domain. In numerous instances, initial investments of behavioral campaigns (e.g., feedback hardware) must amortize over time, bearing the risk that conservation effects wear off too much over time and render outcomes of behavioral campaigns negative (i.e., no net decrease in CO_2 emissions). Additionally, it may be interesting whether the treatment effects observed in the learning studies might translate to other courses. More precisely, individuals could have learned that their behavior change (i.e., spending more time studying in the middle of the semester) helped them pass the respective course. Consequently, they might adopt this time management strategy in future university courses.

To determine how the effects of interventions at the micro level extrapolate to the macro level (i.e., concerning the micro-macro transition of the research framework), future research could also use novel analytical methods to identify heterogeneous effects. In particular, causal machine learning algorithms might be valuable in detecting non-linear effects. With the relatively low number of individuals across the field studies, the analyses of this dissertation were deliberately conservative. This means that only linear relationships were considered to reduce the risk of producing spurious results. Especially with larger sample sizes, novel analytical methods for identifying heterogeneous effects could provide more insights into the interplay between interventions and individuals' characteristics (see, e.g., Künzel et al., 2019; McFowland III et al., 2021; Wager and Athey, 2018).

9 Conclusion

The aim of this dissertation was to explore how technical opportunities for behavior change could change important outcomes at the micro and macro level. By using an adaptation of the belief–action–outcome framework, the dissertation specifically focused on understanding the effects of feedback interventions to solve personal and societal issues. Studies in the educational domain indicate that relatively simple feedback information can considerably decrease students' procrastination within the online learning environments of university courses. Moreover, the presented studies on environmental sustainability support that feedback information can foster considerable conservation effects without requiring financial incentives for behavior change. Importantly, as demonstrated in the study on carbon offsetting, feedback information can even counteract adverse effects, namely higher consumption levels, from other behavioral programs. Besides studying such effects at the micro level, the dissertation provided insights into whether the observed effects might project to the macro level by shedding light on the heterogeneity in behavioral responses. In addition to the observed behavioral effects, the dissertation suggests that potentially more effective feedback systems can be implemented in practice; this is based on the evidence of the technical or organizational feasibility of novel feedback systems.

Due to the ubiquity of IT, many outcomes of our everyday actions can be increasingly captured and quantified. This not only helps to detect behavioral issues that individuals or organizations face but also informs interventions to overcome such issues on a large scale. Despite the associated potential of interventions, numerous questions regarding their effects have not yet been sufficiently studied. While this dissertation has substantially extended our knowledge of feedback interventions, there is still a strong need to better understand their effects across different behavioral domains, especially in view of factors that lead to heterogeneous effects. Ultimately, this could inform practitioners of how to harness the full potential of feedback interventions to improve personal, organizational, and societal outcomes.

References

Ableitner, L., Tiefenbeck, V., Hosseini, S., Schöb, S., Fridgen, G., and Staake, T. (2017). "Real-World Impact of Information Systems: The Effect of Seemingly Small Design Choices." *Proceedings of the Workshop on Information Technologies and Systems*. Seoul, South Korea.

Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. (2005). "A Review of Intervention Studies Aimed at Household Energy Conservation." *Journal of Environmental Psychology* 25 (3), pp. 273–291. DOI: 10.1016/j.jenvp.2005.08.002.

Adams, J. A. (1987). "Historical Review and Appraisal of Research on the Learning, Retention, and Transfer of Human Motor Skills." *Psychological Bulletin* 101 (1), pp. 41–74. DOI: 10.1037/0033-2909.101.1.41.

Adedoyin, O. B. and Soykan, E. (2020). "Covid-19 Pandemic and Online Learning: The Challenges and Opportunities." *Interactive Learning Environments* 31 (2), pp. 863–875. DOI: 10.1080/10494820.2020.1813180.

Agostinelli, G., Brown, J. M., and Miller, W. R. (1995). "Effects of Normative Feedback on Consumption among Heavy Drinking College Students." *Journal of Drug Education* 25 (1), pp. 31–40. DOI: 10.2190/XD56-D6WR-7195-EAL3.

Ajzen, I. (1991). "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes* 50 (2), pp. 179–211. DOI: 10.1016/0749-5978(91)90020-T.

Alberini, A. and Towe, C. (2015). "Information v. Energy Efficiency Incentives: Evidence from Residential Electricity Consumption in Maryland." *Energy Economics* 52 (Supplement 1 2015), S30–S40. DOI: 10.1016/j.eneco.2015.08.013.

Allcott, H. (2011). "Social Norms and Energy Conservation." *Journal of Public Economics* 95 (9-10), pp. 1082–1095. DOI: 10.1016/j.jpubeco.2011.03.003.

Amankwah-Amoah, J. and Syllias, J. (2020). "Can Adopting Ambitious Environmental Sustainability Initiatives Lead to Business Failures? An Analytical Framework." *Business Strategy and the Environment* 29 (1), pp. 240–249. DOI: 10.1002/bse.2361.

Amari, S.-i., Fujita, N., and Shinomoto, S. (1992). "Four Types of Learning Curves." *Neural Computation* 4 (4), pp. 605–618. DOI: 10.1162/neco.1992.4.4.605.

Anderson, J. R. (1982). "Acquisition of Cognitive Skill." *Psychological Review* 89 (4), pp. 369–406. DOI: 10.1037/0033-295X.89.4.369.

Anderson, K. (2012). "The Inconvenient Truth of Carbon Offsets." *Nature* 484 (7392), p. 7. DOI: 10.1038/484007a.

Ariely, D., Bracha, A., and Meier, S. (2009). "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review* 99 (1), pp. 544–555. DOI: 10.1257/aer.99.1.544.

Asensio, O. I. and Delmas, M. A. (2015). “Nonprice Incentives and Energy Conservation.” *Proceedings of the National Academy of Sciences* 112 (6), E510–E515. DOI: 10.1073/pnas.1401880112.

Bai, J., Choi, S. H., and Liao, Y. (2021). “Feasible Generalized Least Squares for Panel Data with Cross-Sectional and Serial Correlations.” *Empirical Economics* 60 (1), pp. 309–326. DOI: 10.1007/s00181-020-01977-2.

Bamberg, S., Hunecke, M., and Blöbaum, A. (2007). “Social Context, Personal Norms and the Use of Public Transportation: Two Field Studies.” *Journal of Environmental Psychology* 27 (3), pp. 190–203. DOI: 10.1016/j.jenvp.2007.04.001.

Bamberg, S. and Möser, G. (2007). “Twenty Years after Hines, Hungerford, and Tomera: A New Meta-Analysis of Psycho-Social Determinants of Pro-Environmental Behaviour.” *Journal of Environmental Psychology* 27 (1), pp. 14–25. DOI: 10.1016/j.jenvp.2006.12.002.

Bandiera, O., Barankay, I., and Rasul, I. (2007). “Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment.” *The Quarterly Journal of Economics* 122 (2), pp. 729–773. DOI: 10.1162/qjec.122.2.729.

Bandura, A. (1977). “Self-Efficacy: Toward a Unifying Theory of Behavioral Change.” *Psychological Review* 84 (2), pp. 191–215. DOI: 10.1037/0033-295x.84.2.191.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control.* New York, NY, USA: W. H. Freeman and Company. ISBN: 978-0-7167-2850-4.

Bargh, J. A. (2006). “What Have We Been Priming All These Years? On the Development, Mechanisms, and Ecology of Nonconscious Social Behavior.” *European Journal of Social Psychology* 36 (2), pp. 147–168. DOI: 10.1002/ejsp.336.

Bekkar, M., Djemaa, H. K., and Taklit, A. A. (2013). “Evaluation Measures for Models Assessment over Imbalanced Data Sets.” *Journal of Information Engineering and Applications* 3 (10), pp. 27–38.

Bem, D. J. (1972). “Self-Perception Theory.” In: *Advances in Experimental Social Psychology.* Edited by L. Berkowitz. Volume 6. New York, NY, USA: Academic Press, pp. 1–62. ISBN: 978-0-12-015206-3.

Bénabou, R. and Tirole, J. (2006). “Incentives and Prosocial Behavior.” *The American Economic Review* 96 (5), pp. 1652–1678. DOI: 10.1257/aer.96.5.1652.

Benjamini, Y. and Hochberg, Y. (1995). “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1), pp. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Berg, N. (2010). “Behavioral Economics.” In: *21st Century Economics: A Reference Handbook.* Edited by R. C. Free. Volume 2. Thousand Oaks, CA, USA: SAGE Publications, pp. 861–872. ISBN: 978-1-4129-6142-4.

Berkowitz, A. D. (2005). “An Overview of the Social Norms Approach.” In: *Changing The Culture Of College Drinking: A Socially Situated Health Communication Campaign.* Edited by L. C. Lederman and L. P. Stewart. Cresskill, NJ, USA: Hampton Press, pp. 193–214. ISBN: 978-1-57273-593-4.

Bertrand, A., Mastrucci, A., Schüler, N., Aggoune, R., and Maréchal, F. (2017). “Characterisation of Domestic Hot Water End-Uses for Integrated Urban Thermal Energy Assessment and Optimisation.” *Applied Energy* 186 (Part 2), pp. 152–166. DOI: 10.1016/j.apenergy.2016.02.107.

Bloomfield, S. F., Aiello, A. E., Cookson, B., O’Boyle, C., and Larson, E. L. (2007). “The Effectiveness of Hand Hygiene Procedures in Reducing the Risks of Infections in Home and Community Settings Including Handwashing and Alcohol-Based Hand Sanitizers.” *American Journal of Infection Control* 35 (10), pp. S27–S64. DOI: 10.1016/j.ajic.2007.07.001.

Boehler, M. L., Rogers, D. A., Schwind, C. J., Mayforth, R., Quin, J., Williams, R. G., and Dunnington, G. (2006). “An Investigation of Medical Student Reactions to Feedback: A Randomised Controlled Trial.” *Medical Education* 40 (8), pp. 746–749. DOI: 10.1111/j.1365-2929.2006.02503.x.

Boulstridge, E. and Carrigan, M. (2000). “Do Consumers Really Care about Corporate Responsibility? Highlighting the Attitude—Behaviour Gap.” *Journal of Communication Management* 4 (4), pp. 355–368. DOI: 10.1108/eb023532.

Bradley, A. P. (1997). “The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms.” *Pattern Recognition* 30 (7), pp. 1145–1159. DOI: 10.1016/S0031-3203(96)00142-2.

Braun, B. I., Kusek, L., and Larson, E. (2009). “Measuring Adherence to Hand Hygiene Guidelines: A Field Survey for Examples of Effective Practices.” *American Journal of Infection Control* 37 (4), pp. 282–288. DOI: 10.1016/j.ajic.2008.09.002.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., Stave, C. D., Olkin, I., and Sirard, J. R. (2007). “Using Pedometers to Increase Physical Activity and Improve Health: A Systematic Review.” *JAMA* 298 (19), pp. 2296–2304. DOI: 10.1001/jama.298.19.2296.

Broadbent, J. and Poon, W. (2015). “Self-Regulated Learning Strategies & Academic Achievement in Online Higher Education Learning Environments: A Systematic Review.” *The Internet and Higher Education* 27, pp. 1–13. DOI: 10.1016/j.iheduc.2015.04.007.

Broadbent, J. (2017). “Comparing Online and Blended Learner’s Self-Regulated Learning Strategies and Academic Performance.” *The Internet and Higher Education* 33, pp. 24–32. DOI: 10.1016/j.iheduc.2017.01.004.

Burton-Jones, A., Butler, B. S., Scott, S. V., and Xu, S. X. (2021). “Next-Generation Information Systems Theorizing: A Call to Action.” *MIS Quarterly* 45 (1), pp. 301–314. DOI: 10.25300/MISQ/2021/15434.

Burton-Jones, A. and Grange, C. (2013). “From Use to Effective Use: A Representation Theory Perspective.” *Information Systems Research* 24 (3), pp. 632–658. DOI: 10.1287/isre.1120.0444.

Campbell, D. T. (1963). "Social Attitudes and Other Acquired Behavioral Dispositions." In: *Psychology: A Study of a Science*. Edited by S. Koch. Volume 6. Investigations of man as socius: Their place in psychology and the social sciences. Study II. Empirical Substructure and Relations with Other Sciences. New York, NY, USA: McGraw-Hill, pp. 94–172. DOI: 10.1037/10590-000.

Carver, C. S. and Scheier, M. F. (1981). *Attention and Self-Regulation: A Control-Theory Approach to Human Behavior*. Springer Series in Social Psychology. New York, NY, USA: Springer. ISBN: 978-1-4612-5889-6. DOI: 10.1007/978-1-4612-5887-2.

Chen, W. and Hirschheim, R. (2004). "A Paradigmatic and Methodological Examination of Information Systems Research from 1991 to 2001." *Information Systems Journal* 14 (3), pp. 197–235. DOI: 10.1111/j.1365-2575.2004.00173.x.

Cho, M.-H. and Summers, J. (2012). "Factor Validity of the Motivated Strategies for Learning Questionnaire (MSLQ) in Asynchronous Online Learning Environments." *Journal of Interactive Learning Research* 23 (1), pp. 5–28.

Choudhary, V., Shunko, M., Netessine, S., and Koo, S. (2021). "Nudging Drivers to Safety: Evidence from a Field Experiment." *Management Science* 68 (6), pp. 4196–4214. DOI: 10.1287/mnsc.2021.4063.

Chowdhury, S. F. and Pychyl, T. A. (2018). "A Critique of the Construct Validity of Active Procrastination." *Personality and Individual Differences* 120, pp. 7–12. DOI: 10.1016/j.paid.2017.08.016.

Chun Chu, A. H. and Choi, J. N. (2005). "Rethinking Procrastination: Positive Effects of 'Active' Procrastination Behavior on Attitudes and Performance." *The Journal of Social Psychology* 145 (3), pp. 245–264. DOI: 10.3200/SOCP.145.3.245-264.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology* 58 (6), pp. 1015–1026. DOI: 10.1037/0022-3514.58.6.1015.

Cintron, D. W., Adler, N. E., Gottlieb, L. M., Hagan, E., Tan, M. L., Vlahov, D., Glymour, M. M., and Matthay, E. C. (2022). "Heterogeneous Treatment Effects in Social Policy Studies: An Assessment of Contemporary Articles in the Health and Social Sciences." *Annals of Epidemiology* 70, pp. 79–88. DOI: 10.1016/j.annepidem.2022.04.009.

Clark, E. V. (2016). *First Language Acquisition*. 3rd Edition. Cambridge, UK: Cambridge University Press. ISBN: 978-1-316-50760-5.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Hillsdale, NJ, USA: Lawrence Erlbaum Associates. ISBN: 978-0-203-77158-7. DOI: 10.4324/9780203771587.

Coleman, J. S. (1986). "Social Theory, Social Research, and a Theory of Action." *American Journal of Sociology* 91 (6), pp. 1309–1335. DOI: 10.1086/228423.

Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA, USA: Belknap Press of Harvard University Press. ISBN: 978-0-674-31225-8.

Conner, M., Norman, P., and Bell, R. (2002). "The Theory of Planned Behavior and Healthy Eating." *Health Psychology* 21, pp. 194–201. DOI: 10.1037/0278-6133.21.2.194.

Costa, D. L. and Kahn, M. E. (2013). "Energy Conservation 'Nudges' and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment." *Journal of the European Economic Association* 11 (3), pp. 680–702. DOI: 10.1111/jeea.12011.

Cozby, P. C. and Bates, S. C. (2020). *Methods in Behavioral Research.* 14th Edition. New York, NY, USA: McGraw-Hill Education. ISBN: 978-1-260-20558-9.

Credé, M. and Phillips, L. A. (2011). "A Meta-Analytic Review of the Motivated Strategies for Learning Questionnaire." *Learning and Individual Differences* 21 (4), pp. 337–346. DOI: 10.1016/j.lindif.2011.03.002.

Croissant, Y. and Millo, G. (2019). *Panel Data Econometrics with R.* 1st Edition. Hoboken, NJ, USA: John Wiley & Sons. ISBN: 978-1-118-94917-7. DOI: 10.1002/9781119504641.

Dahlinger, A., Tiefenbeck, V., Ryder, B., Gahr, B., Fleisch, E., and Wortmann, F. (2018). "The Impact of Numerical vs. Symbolic Eco-Driving Feedback on Fuel Consumption – A Randomized Control Field Trial." *Transportation Research Part D: Transport and Environment* 65, pp. 375–386. DOI: 10.1016/j.trd.2018.09.013.

Dangis, G., Terho, K., Graichen, J., Günther, S. A., Rosio, R., Salanterä, S., Staake, T., Stingl, C., and Pakarinen, A. (2023). "Hand Hygiene of Kindergarten Children—Understanding the Effect of Live Feedback on Handwashing Behaviour, Self-Efficacy, and Motivation of Young Children: Protocol for a Multi-Arm Cluster Randomized Controlled Trial." *PLOS ONE* 18 (1), Article nr. e0280686. DOI: 10.1371/journal.pone.0280686.

Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., and Houben, G.-J. (2017). "Follow the Successful Crowd: Raising MOOC Completion Rates through Social Comparison at Scale." *Proceedings of the 7th International Learning Analytics & Knowledge Conference.* Vancouver, BC, Canada. DOI: 10.1145/3027385.3027411.

Davis, J. and Goadrich, M. (2006). "The Relationship between Precision-Recall and ROC Curves." *Proceedings of the 23rd International Conference on Machine Learning.* Pittsburgh, PA, USA. DOI: 10.1145/1143844.1143874.

Day, V., Mensink, D., and O'Sullivan, M. (2000). "Patterns of Academic Procrastination." *Journal of College Reading and Learning* 30 (2), pp. 120–134. DOI: 10.1080/10790195.2000.10850090.

Deci, E. L. (1971). "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology* 18 (1), pp. 105–115. DOI: 10.1037/h0030644.

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin* 125 (6), pp. 627–668. DOI: 10.1037/0033-2909.125.6.627.

Deci, E. L. and Ryan, R. M. (2012). "Motivation, Personality, and Development within Embedded Social Contexts: An Overview of Self-Determination Theory." In: *The Oxford Handbook of Human Motivation.* Edited by R. M. Ryan. Oxford Library of Psychology. New York, NY, USA: Oxford University Press, pp. 85–107. ISBN: 978-0-19-539982-0. DOI: 10.1093/oxfordhb/9780195399820.001.0001.

De Groot, J. I. M. and Steg, L. (2009). "Morality and Prosocial Behavior: The Role of Awareness, Responsibility, and Norms in the Norm Activation Model." *The Journal of Social Psychology* 149 (4), pp. 425–449. DOI: 10.3200/SOCP.149.4.425-449.

Demšar, J. (2006). "Statistical Comparisons of Classifiers over Multiple Data Sets." *Journal of Machine Learning Research* 7, pp. 1–30.

DiClemente, C. C., Prochaska, J. O., Fairhurst, S. K., Velicer, W. F., Velasquez, M. M., and Rossi, J. S. (1991). "The Process of Smoking Cessation: An Analysis of Precontemplation, Contemplation, and Preparation Stages of Change." *Journal of Consulting and Clinical Psychology* 59 (2), pp. 295–304. DOI: 10.1037/0022-006x.59.2.295.

Dimick, J. B. and Ryan, A. M. (2014). "Methods for Evaluating Changes in Health Care Policy: The Difference-in-Differences Approach." *JAMA* 312 (22), pp. 2401–2402. DOI: 10.1001/jama.2014.16153.

Dobusch, L. and Kapeller, J. (2009). "'Why Is Economics Not an Evolutionary Science?' New Answers to Veblen's Old Question." *Journal of Economic Issues* 43 (4), pp. 867–898. DOI: 10.2753/JEI0021-3624430403.

Doepke, M. and Zilibotti, F. (2017). "Parenting With Style: Altruism and Paternalism in Intergenerational Preference Transmission." *Econometrica* 85 (5), pp. 1331–1371. DOI: 10.3982/ECTA14634.

Dolnicar, S., Knezevic Cvelbar, L., and Grün, B. (2017). "Do Pro-Environmental Appeals Trigger Pro-Environmental Behavior in Hotel Guests?" *Journal of Travel Research* 56 (8), pp. 988–997. DOI: 10.1177/0047287516678089.

Duncan, T. G. and McKeachie, W. J. (2005). "The Making of the Motivated Strategies for Learning Questionnaire." *Educational Psychologist* 40 (2), pp. 117–128. DOI: 10.1207/s15326985ep4002_6.

Durant, D. J., Willis, L., and Duvall, S. (2020). "Adoption of Electronic Hand Hygiene Monitoring Systems in New York State Hospitals and the Associated Impact on Hospital-Acquired C. Difficile Infection Rates." *American Journal of Infection Control* 48 (7), pp. 733–739. DOI: 10.1016/j.ajic.2020.04.005.

El Boghdady, M. and Alijani, A. (2017). "Feedback in Surgical Education." *The Surgeon* 15 (2), pp. 98–103. DOI: 10.1016/j.surge.2016.06.006.

Enshaeifar, S., Zoha, A., Markides, A., Skillman, S., Acton, S. T., Elsaleh, T., Hassanpour, M., Ahrabian, A., Kenny, M., Klein, S., Rostill, H., Nilforooshan, R., and Barnaghi, P. (2018). "Health Management and Pattern Analysis of Daily Living Activities of People with Dementia Using In-Home Sensors and Machine Learning Techniques." *PLOS ONE* 13 (5), Article nr. e0195605. DOI: 10.1371/journal.pone.0195605.

Esteves-Sorenson, C. and Broce, R. (2022). "Do Monetary Incentives Undermine Performance on Intrinsically Enjoyable Tasks? A Field Test." *The Review of Economics and Statistics* 104 (1), pp. 67–84. DOI: 10.1162/rest_a_00947.

European Union (2021). *Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on Energy Efficiency, Amending Directives 2009/125/EC and 2010/30/EU and Repealing Directives 2004/8/EC and 2006/32/EC*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02012L0027-20210101 (visited on 02/07/2023).

Evans, J. S. B. T. (2008). "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition." *Annual Review of Psychology* 59 (1), pp. 255–278. DOI: 10.1146/annurev.psych.59.103006.093629.

Falk, A. and Fischbacher, U. (2006). "A Theory of Reciprocity." *Games and Economic Behavior* 54 (2), pp. 293–315. DOI: 10.1016/j.geb.2005.03.001.

Fang, X., Goette, L., Rockenbach, B., Sutter, M., Tiefenbeck, V., Schoeb, S., and Staake, T. (2023). "Complementarities in Behavioral Interventions: Evidence from a Field Experiment on Energy Conservation." *Journal of Public Economics* 228, Article nr. 105028. DOI: 10.1016/j.jpubeco.2023.105028.

Fang, Y., Qureshi, I., Sun, H., McCole, P., Ramsey, E., and Lim, K. H. (2014). "Trust, Satisfaction, and Online Repurchase Intention: The Moderating Role of Perceived Effectiveness of E-Commerce Institutional Mechanisms." *MIS Quarterly* 38 (2), pp. 407–427. DOI: 10.25300/MISQ/2014/38.2.04.

Fernández-Loría, C., Provost, F., and Han, X. (2022). "Explaining Data-Driven Decisions Made by AI Systems: The Counterfactual Approach." *MIS Quarterly* 46 (3), pp. 1635–1660. DOI: 10.25300/MISQ/2022/16749.

Ferri, C., Hernández-Orallo, J., and Modroiu, R. (2009). "An Experimental Comparison of Performance Measures for Classification." *Pattern Recognition Letters* 30 (1), pp. 27–38. DOI: 10.1016/j.patrec.2008.08.010.

Festinger, L. (1954). "A Theory of Social Comparison Processes." *Human Relations* 7 (2), pp. 117–140. DOI: 10.1177/001872675400700202.

Field, A. P., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. London, UK: SAGE Publications. ISBN: 978-1-4462-0046-9.

Fishbein, M. and Yzer, M. C. (2003). "Using Theory to Design Effective Health Behavior Interventions." *Communication Theory* 13 (2), pp. 164–183. DOI: 10.1111/j.1468-2885.2003.tb00287.x.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* 40 (2), pp. 351–401. DOI: 10.1257/jel.40.2.351.

Gafni, R. and Filin, D. (2015). "Worldwide Video Use Patterns in E-Learning: Exploring Time, Completion Rate and Devices." *Online Journal of Applied Knowledge Management* 3 (1), pp. 74–92.

Gaube, S., Fischer, P., and Lermer, E. (2021). "Hand(y) Hygiene Insights: Applying Three Theoretical Models to Investigate Hospital Patients' and Visitors' Hand Hygiene Behavior." *PLOS ONE* 16 (1), Article nr. e0245543. DOI: 10.1371/journal.pone.0245543.

Geri, N., Gafni, R., and Winer, A. (2014). "The U-Curve of E-Learning: Course Website and Online Video Use in Blended and Distance Learning." *Interdisciplinary Journal of e-Skills and Lifelong Learning* 10, pp. 1–16. DOI: 10.28945/1959.

Gerster, A., Andor, M. A., and Goette, L. (2020). "Disaggregate Consumption Feedback and Energy Conservation." Working Paper. URL: https://ssrn.com/abstract=3638059.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25 (4), pp. 191–210. DOI: 10.1257/jep.25.4.191.

Gneezy, U. and Rustichini, A. (2000a). "A Fine Is a Price." *The Journal of Legal Studies* 29 (1), pp. 1–17. DOI: 10.1086/468061.

Gneezy, U. and Rustichini, A. (2000b). "Pay Enough or Don't Pay at All." *The Quarterly Journal of Economics* 115 (3), pp. 791–810. DOI: 10.1162/003355300554917.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels." *Journal of Consumer Research* 35 (3), pp. 472–482. DOI: 10.1086/586910.

Goodman, J. S. and Wood, R. E. (2004). "Feedback Specificity, Learning Opportunities, and Learning." *Journal of Applied Psychology* 89 (5), pp. 809–821. DOI: 10.1037/0021-9010.89.5.809.

Goodman, J. S., Wood, R. E., and Hendrickx, M. (2004). "Feedback Specificity, Exploration, and Learning." *Journal of Applied Psychology* 89 (2), pp. 248–262. DOI: 10.1037/0021-9010.89.2.248.

Gössling, S., Broderick, J., Upham, P., Ceron, J.-P., Dubois, G., Peeters, P., and Strasdas, W. (2007). "Voluntary Carbon Offsetting Schemes for Aviation: Efficiency, Credibility and Sustainable Tourism." *Journal of Sustainable Tourism* 15 (3), pp. 223–248. DOI: 10.2167/jost758.0.

Gradman, A. H., Basile, J. N., Carter, B. L., and Bakris, G. L. (2010). "Combination Therapy in Hypertension." *Journal of the American Society of Hypertension* 4 (2), pp. 90–98. DOI: 10.1016/j.jash.2010.03.001.

Greaves, M., Zibarras, L. D., and Stride, C. (2013). "Using the Theory of Planned Behavior to Explore Environmental Behavioral Intentions in the Workplace." *Journal of Environmental Psychology* 34, pp. 109–120. DOI: 10.1016/j.jenvp.2013.02.003.

Greene, W. H. (2018). *Econometric Analysis*. 8th Edition. New York, NY, USA: Pearson. ISBN: 978-0-13-446136-6.

Guerra, J., Hosseini, R., Somyurek, S., and Brusilovsky, P. (2016). "An Intelligent Interface for Learning Content: Combining an Open Learner Model and Social Comparison to Support

Self-Regulated Learning and Engagement." *Proceedings of the 21st International Conference on Intelligent User Interfaces*. Sonoma, CA, USA. DOI: 10.1145/2856767.2856784.

Günther, S. A., Schöb, S., Coroamă, V. C., Tiefenbeck, V., Mattern, F., and Staake, T. (under review). "All Eyes on Consumption Feedback: A Randomized Controlled Trial on the Interplay of Financial Incentives, Environmental Appeals, and Consumption Feedback." Working Paper. Submitted to Experimental Economics.

Günther, S. A. (2021). "The Impact of Social Norms on Students' Online Learning Behavior: Insights from Two Randomized Controlled Trials." *Proceedings of the 11th International Conference on Learning Analytics & Knowledge*. Irvine, CA, USA. DOI: 10.1145/3448139.3448141.

Günther, S. A., Haag, F., Hopf, K., Handschuh, P., Klose, M., and Staake, T. (forthcoming). "A Feedback Component That Leverages Counterfactual Explanations for Smart Learning Support." In: *Digitale Kulturen der Lehre entwickeln – Rahmenbedingungen, Konzepte und Werkzeuge*. Edited by L. Mrohs, J. Franz, D. Herrmann, K. Lindner, and T. Staake. Perspektiven der Hochschuldidaktik. Wiesbaden, Germany: Springer VS.

Günther, S. A., Staake, T., Schöb, S., and Tiefenbeck, V. (2020). "The Behavioral Response to a Corporate Carbon Offset Program: A Field Experiment on Adverse Effects and Mitigation Strategies." *Global Environmental Change* 64, Article nr. 102123. DOI: 10.1016/j.gloenvcha.2020.102123.

Günther, S. A., Stingl, C., Coroamă, V. C., Schöb, S., and Staake, T. (2020). "Empowering Personalized Feedback on Hot Water Usage: A Field Study with Shower Meters." *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. Brno, Czech Republic. DOI: 10.1145/3341105.3374053.

Günther, S. A., Veihelmann, T., and Staake, T. (2020). "Leveraging Social Norms to Encourage Online Learning: Empirical Evidence from a Blended Learning Course." *Proceedings of the 41st International Conference on Information Systems*. Hyderabad, India.

Hagel, S., Reischke, J., Kesselmeier, M., Winning, J., Gastmeier, P., Brunkhorst, F. M., Scherag, A., and Pletz, M. W. (2015). "Quantifying the Hawthorne Effect in Hand Hygiene Compliance Through Comparing Direct Observation With Automated Hand Hygiene Monitoring." *Infection Control & Hospital Epidemiology* 36 (8), pp. 957–962. DOI: 10.1017/ice.2015.93.

Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., and Sarstedt, M. (2021). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 3rd Edition. Thousand Oaks, CA, USA: Sage Publications. ISBN: 978-1-5443-9641-5.

Handgraaf, M. J., Van Lidth de Jeude, M. A., and Appelt, K. C. (2013). "Public Praise vs. Private Pay: Effects of Rewards on Energy Conservation in the Workplace." *Ecological Economics* 86, pp. 86–92. DOI: 10.1016/j.ecolecon.2012.11.008.

Harding, M. and Rapson, D. (2019). "Does Absolution Promote Sin? A Conservationist's Dilemma." *Environmental and Resource Economics* 73 (3), pp. 923–955. DOI: 10.1007/s10640-018-0301-5.

Harrison, G. W. and List, J. A. (2004). "Field Experiments." *Journal of Economic Literature* 42 (4), pp. 1009–1055. DOI: 10.1257/0022051043004577.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd Edition. Springer Series in Statistics. New York, NY, USA: Springer. ISBN: 978-0-387-84858-7.

Hattie, J. and Timperley, H. (2007). "The Power of Feedback." *Review of Educational Research* 77 (1), pp. 81–112. DOI: 10.3102/003465430298487.

Hensley, L. C. (2014). "Reconsidering Active Procrastination: Relations to Motivation and Achievement in College Anatomy." *Learning and Individual Differences* 36, pp. 157–164. DOI: 10.1016/j.lindif.2014.10.012.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). "Design Science in Information Systems Research." *MIS Quarterly* 28 (1), pp. 75–105. DOI: 10.2307/25148625.

Höhne, N., den Elzen, M., Rogelj, J., Metz, B., Fransen, T., Kuramochi, T., Olhoff, A., Alcamo, J., Winkler, H., Fu, S., Schaeffer, M., Schaeffer, R., Peters, G. P., Maxwell, S., and Dubash, N. K. (2020). "Emissions: World Has Four Times the Work or One-Third of the Time." *Nature* 579 (7797), pp. 25–28. DOI: 10.1038/d41586-020-00571-x.

Huis, A., van Achterberg, T., de Bruin, M., Grol, R., Schoonhoven, L., and Hulscher, M. (2012). "A Systematic Review of Hand Hygiene Improvement Strategies: A Behavioural Approach." *Implementation Science* 7 (1), Article nr. 92. DOI: 10.1186/1748-5908-7-92.

Jacobsen, G. D., Kotchen, M. J., and Vandenbergh, M. P. (2012). "The Behavioral Response to Voluntary Provision of an Environmental Public Good: Evidence from Residential Electricity Demand." *European Economic Review* 56 (5), pp. 946–960. DOI: 10.1016/j.euroecorev.2012.02.008.

Janiesch, C., Zschech, P., and Heinrich, K. (2021). "Machine Learning and Deep Learning." *Electronic Markets* 31 (3), pp. 685–695. DOI: 10.1007/s12525-021-00475-2.

Jung, J. H., Schneider, C., and Valacich, J. (2010). "Enhancing the Motivational Affordance of Information Systems: The Effects of Real-Time Performance Feedback and Goal Setting in Group Collaboration Environments." *Management Science* 56 (4), pp. 724–742. DOI: 10.1287/mnsc.1090.1129.

Juvan, E. and Dolnicar, S. (2014). "The Attitude–Behaviour Gap in Sustainable Tourism." *Annals of Tourism Research* 48, pp. 76–95. DOI: 10.1016/j.annals.2014.05.012.

Kahneman, D. (2012). *Thinking, Fast and Slow.* London, UK: Penguin Books. ISBN: 978-0-14-103357-0.

Kahneman, D. and Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2), pp. 263–292. DOI: 10.2307/1914185.

Kaiser, F. G. (2021). "Climate Change Mitigation within the Campbell Paradigm: Doing the Right Thing for a Reason and against All Odds." *Current Opinion in Behavioral Sciences* 42, pp. 70–75. DOI: 10.1016/j.cobeha.2021.03.024.

Kaiser, F. G., Byrka, K., and Hartig, T. (2010). "Reviving Campbell's Paradigm for Attitude Research." *Personality and Social Psychology Review* 14 (4), pp. 351–367. DOI: 10.1177/1088868310366452.

Kaiser, F. G. and Gutscher, H. (2003). "The Proposition of a General Version of the Theory of Planned Behavior: Predicting Ecological Behavior." *Journal of Applied Social Psychology* 33 (3), pp. 586–603. DOI: 10.1111/j.1559-1816.2003.tb01914.x.

Kaiser, F. G., Hübner, G., and Bogner, F. X. (2005). "Contrasting the Theory of Planned Behavior With the Value-Belief-Norm Model in Explaining Conservation Behavior." *Journal of Applied Social Psychology* 35 (10), pp. 2150–2170. DOI: 10.1111/j.1559-1816.2005.tb02213.x.

Kandul, S., Lang, G., and Lanz, B. (2020). "Social Comparison and Energy Conservation in a Collective Action Context: A Field Experiment." *Economics Letters* 188, Article nr. 108947. DOI: 10.1016/j.econlet.2020.108947.

Kanfer, R. and Ackerman, P. L. (1989). "Motivation and Cognitive Abilities: An Integrative/Aptitude-Treatment Interaction Approach to Skill Acquisition." *Journal of Applied Psychology* 74 (4), pp. 657–690. DOI: 10.1037/0021-9010.74.4.657.

Kelley, H. and Miltenberger, R. G. (2016). "Using Video Feedback to Improve Horseback-Riding Skills: Video Feedback and Horseback Riding." *Journal of Applied Behavior Analysis* 49 (1), pp. 138–147. DOI: 10.1002/jaba.272.

Kim, K. R. and Seo, E. H. (2015). "The Relationship between Procrastination and Academic Performance: A Meta-Analysis." *Personality and Individual Differences* 82, pp. 26–33. DOI: 10.1016/j.paid.2015.02.038.

King, P. E. (2016). "When Do Students Benefit From Performance Feedback? A Test of Feedback Intervention Theory in Speaking Improvement." *Communication Quarterly* 64 (1), pp. 1–15. DOI: 10.1080/01463373.2015.1078827.

Kirtley, J. L. (2020). *Electric Power Principles: Sources, Conversion, Distribution and Use.* 2nd Edition. Hoboken, NJ, USA: John Wiley & Sons. ISBN: 978-1-119-58517-6.

Klingsieck, K. B. (2018). "Kurz Und Knapp – Die Kurzskala Des Fragebogens 'Lernstrategien Im Studium' (LIST)." *Zeitschrift für Pädagogische Psychologie* 32 (4), pp. 249–259. DOI: 10.1024/1010-0652/a000230.

Kluger, A. N. and DeNisi, A. (1996). "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory." *Psychological Bulletin* 119 (2), pp. 254–284. DOI: 10.1037/0033-2909.119.2.254.

Koch, T., Peter, C., and Müller, P. (2019). *Das Experiment in der Kommunikations- und Medienwissenschaft: Grundlagen, Durchführung und Auswertung experimenteller Forschung.* Studienbücher zur Kommunikations- und Medienwissenschaft. Wiesbaden, Germany: Springer VS. ISBN: 978-3-658-19754-4. DOI: 10.1007/978-3-658-19754-4.

Kotchen, M. J. and Moore, M. R. (2008). "Conservation: From Voluntary Restraint to a Voluntary Price Premium." *Environmental and Resource Economics* 40 (2), pp. 195–215. DOI: 10.1007/s10640-007-9148-x.

Kulik, J. A. and Kulik, C.-L. C. (1988). "Timing of Feedback and Verbal Learning." *Review of Educational Research* 58 (1), pp. 79–97. DOI: 10.3102/00346543058001079.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2019). "Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning." *Proceedings of the National Academy of Sciences* 116 (10), pp. 4156–4165. DOI: 10.1073/pnas.1804597116.

Lam, C. F., DeRue, D. S., Karam, E. P., and Hollenbeck, J. R. (2011). "The Impact of Feedback Frequency on Learning and Task Performance: Challenging the 'More Is Better' Assumption." *Organizational Behavior and Human Decision Processes* 116 (2), pp. 217–228. DOI: 10.1016/j.obhdp.2011.05.002.

Levitt, S. D. and List, J. A. (2009). "Field Experiments in Economics: The Past, the Present, and the Future." *European Economic Review* 53 (1), pp. 1–18. DOI: 10.1016/j.euroecorev.2008.12.001.

Li, X. and Zhang, J. (2016). "The Effects of Monetary Incentives and Social Comparison on MOOC Participation: A Randomized Field Experiment." *Proceedings of the 37th International Conference on Information Systems*. Dublin, Irland.

Liebe, U., Gewinner, J., and Diekmann, A. (2018). "What Is Missing in Research on Non-Monetary Incentives in the Household Energy Sector?" *Energy Policy* 123, pp. 180–183. DOI: 10.1016/j.enpol.2018.08.036.

Liu, Q., Okui, R., and Yoshimura, A. (2016). "Generalized Least Squares Model Averaging." *Econometric Reviews* 35 (8-10), pp. 1692–1752. DOI: 10.1080/07474938.2015.1092817.

Locke, E. A. and Latham, G. P. (1990). *A Theory of Goal Setting & Task Performance*. Englewood Cliffs, NJ, USA: Prentice-Hall. ISBN: 0-13-913138-8.

Loock, C.-M., Staake, T., and Landwehr, J. (2011). "Green IS Design and Energy Conservation: An Empirical Investigation of Social Normative Feedback." *Proceedings of the 32nd International Conference on Information Systems*. Shanghai, China.

Lurie, N. H. and Swaminathan, J. M. (2009). "Is Timely Information Always Better? The Effect of Feedback Frequency on Decision Making." *Organizational Behavior and Human Decision Processes* 108 (2), pp. 315–329. DOI: 10.1016/j.obhdp.2008.05.005.

Maier, C., Laumer, S., Weinert, C., and Weitzel, T. (2015). "The Effects of Technostress and Switching Stress on Discontinued Use of Social Networking Services: A Study of Facebook Use: Effects of Technostress and Switching Stress." *Information Systems Journal* 25 (3), pp. 275–308. DOI: 10.1111/isj.12068.

Maier, M., Bartoš, F., and Wagenmakers, E.-J. (2023). "Robust Bayesian Meta-Analysis: Addressing Publication Bias with Model-Averaging." *Psychological Methods* 28 (1), pp. 107–122. DOI: 10.1037/met0000405.

Makhni, S., Umscheid, C. A., Soo, J., Chu, V., Bartlett, A., Landon, E., and Marrs, R. (2021). "Hand Hygiene Compliance Rate During the COVID-19 Pandemic." *JAMA Internal Medicine* 181 (7), pp. 1006–1008. DOI: 10.1001/jamainternmed.2021.1429.

Mansilla, E. B. and Ho, T. K. (2004). "On Classifier Domains of Competence." *Proceedings of the 17th International Conference on Pattern Recognition.* Cambridge, UK. DOI: 10.1109/ICPR.2004.1334026.

Manthei, K., Sliwka, D., and Vogelsang, T. (2023). "Talking About Performance or Paying for It? A Field Experiment on Performance Reviews and Incentives." *Management Science* 69 (4), pp. 2198–2216. DOI: 10.1287/mnsc.2022.4431.

McFowland III, E., Gangarapu, S., Bapna, R., and Sun, T. (2021). "A Prescriptive Analytics Framework for Optimal Policy Deployment Using Heterogeneous Treatment Effects." *MIS Quarterly* 45 (4), pp. 1807–1832. DOI: 10.25300/MISQ/2021/15684.

McSweeney, F. K. (2004). "Dynamic Changes in Reinforcer Effectiveness: Satiation and Habituation Have Different Implications for Theory and Practice." *The Behavior Analyst* 27 (2), pp. 171–188. DOI: 10.1007/BF03393178.

Mellström, C. and Johannesson, M. (2008). "Crowding Out in Blood Donation: Was Titmuss Right?" *Journal of the European Economic Association* 6 (4), pp. 845–863. DOI: 10.1162/JEEA.2008.6.4.845.

Melville (2010). "Information Systems Innovation for Environmental Sustainability." *MIS Quarterly* 34 (1), pp. 1–21. DOI: 10.2307/20721412.

Metcalfe, J., Kornell, N., and Finn, B. (2009). "Delayed versus Immediate Feedback in Children's and Adults' Vocabulary Learning." *Memory & Cognition* 37 (8), pp. 1077–1087. DOI: 10.3758/MC.37.8.1077.

Meyer, B. D. (1995). "Natural and Quasi-Experiments in Economics." *Journal of Business & Economic Statistics* 13 (2), pp. 151–161. DOI: 10.2307/1392369.

Michie, S., van Stralen, M. M., and West, R. (2011). "The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions." *Implementation Science* 6 (1), Article nr. 42. DOI: 10.1186/1748-5908-6-42.

Milkman, K. L., Gandhi, L., Patel, M. S., Graci, H. N., Gromet, D. M., Ho, H., Kay, J. S., Lee, T. W., Rothschild, J., Bogard, J. E., Brody, I., Chabris, C. F., Chang, E., Chapman, G. B., Dannals, J. E., Goldstein, N. J., Goren, A., Hershfield, H., Hirsch, A., Hmurovic, J., Horn, S., Karlan, D. S., Kristal, A. S., Lamberton, C., Meyer, M. N., Oakes, A. H., Schweitzer, M. E., Shermohammed, M., Talloen, J., Warren, C., Whillans, A., Yadav, K. N., Zlatev, J. J., Berman, R., Evans, C. N., Ladhania, R., Ludwig, J., Mazar, N., Mullainathan, S., Snider, C. K., Spiess, J., Tsukayama, E., Ungar, L., Van den Bulte, C., Volpp, K. G., and Angela L. Duckworth (2022). "A 680,000-Person Megastudy of Nudges to Encourage Vaccination in Pharmacies." *Proceedings of the National Academy of Sciences* 119 (6), Article nr. e2115126119. DOI: 10.1073/pnas.2115126119.

Mokhtari, R. B., Homayouni, T. S., Baluch, N., Morgatskaya, E., Kumar, S., Das, B., and Yeger, H. (2017). "Combination Therapy in Combating Cancer." *Oncotarget* 8 (23), pp. 38022–38043. DOI: 10.18632/oncotarget.16723.

Mullainathan, S. and Thaler, R. H. (2000). "Behavioral Economics." Working Paper. DOI: 10.3386/w7948.

Murakami, K., Shimada, H., Ushifusa, Y., and Ida, T. (2022). "Heterogeneous Treatment Effects of Nudge and Rebate: Causal Machine Learning in a Field Experiment on Electricity Conservation." *International Economic Review* 63 (4), pp. 1779–1803. DOI: 10.1111/iere.12589.

Myers, M. D. (2013). *Qualitative Research in Business & Management.* 2nd Edition. London, UK: SAGE Publications. ISBN: 978-0-85702-973-7.

Nisbett, R. E. and Wilson, T. D. (1977). "The Halo Effect: Evidence for Unconscious Alteration of Judgments." *Journal of Personality and Social Psychology* 35 (4), pp. 250–256. DOI: 10.1037/0022-3514.35.4.250.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2008). "Normative Social Influence Is Underdetected." *Personality and Social Psychology Bulletin* 34 (7), pp. 913–923. DOI: 10.1177/0146167208316691.

Norberg, A., Dziuban, C. D., and Moskal, P. D. (2011). "A Time-based Blended Learning Model." *On the Horizon* 19 (3), pp. 207–216. DOI: 10.1108/10748121111163913.

Nunan, D., Aronson, J., and Bankhead, C. (2018). "Catalogue of Bias: Attrition Bias." *BMJ Evidence-Based Medicine* 23 (1), pp. 21–22. DOI: 10.1136/ebmed-2017-110883.

Onwezen, M. C., Antonides, G., and Bartels, J. (2013). "The Norm Activation Model: An Exploration of the Functions of Anticipated Pride and Guilt in Pro-Environmental Behaviour." *Journal of Economic Psychology* 39, pp. 141–153. DOI: 10.1016/j.joep.2013.07.005.

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., and Sinz, E. J. (2011). "Memorandum on Design-Oriented Information Systems Research." *European Journal of Information Systems* 20 (1), pp. 7–10. DOI: 10.1057/ejis.2010.55.

Özer, B. U., Demir, A., and Ferrari, J. R. (2009). "Exploring Academic Procrastination Among Turkish Students: Possible Gender Differences in Prevalence and Reasons." *The Journal of Social Psychology* 149 (2), pp. 241–257. DOI: 10.3200/SOCP.149.2.241-257.

Panadero, E. (2017). "A Review of Self-Regulated Learning: Six Models and Four Directions for Research." *Frontiers in Psychology* 8, Article nr. 422. DOI: 10.3389/fpsyg.2017.00422.

Paré, G., Trudel, M.-C., Jaana, M., and Kitsiou, S. (2015). "Synthesizing Information Systems Knowledge: A Typology of Literature Reviews." *Information & Management* 52 (2), pp. 183–199. DOI: 10.1016/j.im.2014.08.008.

Patrzek, J., Grunschel, C., and Fries, S. (2012). "Academic Procrastination: The Perspective of University Counsellors." *International Journal for the Advancement of Counselling* 34 (3), pp. 185–201. DOI: 10.1007/s10447-012-9150-z.

Peffers, K., Tuunanen, T., and Niehaves, B. (2018). "Design Science Research Genres: Introduction to the Special Issue on Exemplars and Criteria for Applicable Design Science Research."

European Journal of Information Systems 27 (2), pp. 129–139. DOI: 10.1080/0960085X.2018.1458066.

Pellerano, J. A., Price, M. K., Puller, S. L., and Sánchez, G. E. (2017). "Do Extrinsic Incentives Undermine Social Norms? Evidence from a Field Experiment in Energy Conservation." *Environmental and Resource Economics* 67 (3), pp. 413–428. DOI: 10.1007/s10640-016-0094-3.

Pintrich, P. R., Smith, D. A. F., Garcia, T., and McKeachie, W. J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Report. Ann Arbor, MI, USA: National Center for Research to Improve Postsecondary Teaching and Learning. URL: https://eric.ed.gov/?id=ED338122.

Pintrich, P. R. (2000). "The Role of Goal Orientation in Self-Regulated Learning." In: *Handbook of Self-Regulation*. Edited by M. Boekaerts, P. R. Pintrich, and M. Zeidner. San Diego, CA, USA: Academic Press, pp. 451–502. ISBN: 978-0-12-109890-2. DOI: 10.1016/B978-0-12-109890-2.X5027-6.

Pintrich, P. R. (2004). "A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students." *Educational Psychology Review* 16 (4), pp. 385–407. DOI: 10.1007/s10648-004-0006-x.

Prochaska, J. O. and Velicer, W. F. (1997). "Behavior Change: The Transtheoretical Model of Health Behavior Change." *American Journal of Health Promotion* 12 (1), pp. 38–48. DOI: 10.4278/0890-1171-12.1.38.

Recker, J. (2021). *Scientific Research in Information Systems: A Beginner's Guide*. 2nd Edition. Progress in IS. Cham, Switzerland: Springer. ISBN: 978-3-030-85436-2. DOI: 10.1007/978-3-030-85436-2.

Rhodes, R. E., Courneya, K. S., and Jones, L. W. (2004). "Personality and Social Cognitive Influences on Exercise Behavior: Adding the Activity Trait to the Theory of Planned Behavior." *Psychology of Sport and Exercise* 5 (3), pp. 243–254. DOI: 10.1016/S1469-0292(03)00004-9.

Richardson, M., Abraham, C., and Bond, R. (2012). "Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis." *Psychological Bulletin* 138 (2), pp. 353–387. DOI: 10.1037/a0026838.

Rivera, M., Qiu, L., Kumar, S., and Petrucci, T. (2021). "Are Traditional Performance Reviews Outdated? An Empirical Analysis on Continuous, Real-Time Feedback in the Workplace." *Information Systems Research* 32 (2), pp. 517–540. DOI: 10.1287/isre.2020.0979.

Roell, K. R., Reif, D. M., and Motsinger-Reif, A. A. (2017). "An Introduction to Terminology and Methodology of Chemical Synergy—Perspectives from Across Disciplines." *Frontiers in Pharmacology* 8, Article nr. 158. DOI: 10.3389/fphar.2017.00158.

Rogers, R. W. (1975). "A Protection Motivation Theory of Fear Appeals and Attitude Change." *The Journal of Psychology* 91 (1), pp. 93–114. DOI: 10.1080/00223980.1975.9915803.

Rogers, T. and Feller, A. (2016). “Discouraged by Peer Excellence: Exposure to Exemplary Peer Performance Causes Quitting.” *Psychological Science* 27 (3), pp. 365–374. DOI: 10.1177/0956797615623770.

Roos, D. and Hahn, R. (2019). “Understanding Collaborative Consumption: An Extension of the Theory of Planned Behavior with Value-Based Personal Norms.” *Journal of Business Ethics* 158 (3), pp. 679–697. DOI: 10.1007/s10551-017-3675-3.

Rosenstock, I. M. (1974). “Historical Origins of the Health Belief Model.” *Health Education Monographs* 2 (4), pp. 328–335. DOI: 10.1177/109019817400200403.

Sanguinetti, A., Queen, E., Yee, C., and Akanesuvan, K. (2020). “Average Impact and Important Features of Onboard Eco-Driving Feedback: A Meta-Analysis.” *Transportation Research Part F: Traffic Psychology and Behaviour* 70, pp. 1–14. DOI: 10.1016/j.trf.2020.02.010.

Scalco, A., Noventa, S., Sartori, R., and Ceschi, A. (2017). “Predicting Organic Food Consumption: A Meta-Analytic Structural Equation Model Based on the Theory of Planned Behavior.” *Appetite* 112, pp. 235–248. DOI: 10.1016/j.appet.2017.02.007.

Schöb, S., Günther, S. A., Regensburger, K., and Staake, T. (2018). “NIWM: Non-Intrusive Water Monitoring to Uncover Heat Energy Use in Households.” *Computer Science - Research and Development* 33 (1-2), pp. 127–133. DOI: 10.1007/s00450-017-0353-8.

Schultz, P. W., Estrada, M., Schmitt, J., Sokoloski, R., and Silva-Send, N. (2015). “Using In-Home Displays to Provide Smart Meter Feedback about Household Electricity Consumption: A Randomized Control Trial Comparing Kilowatts, Cost, and Social Norms.” *Energy* 90 (Part 1), pp. 351–358. DOI: 10.1016/j.energy.2015.06.130.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). “The Constructive, Destructive, and Reconstructive Power of Social Norms.” *Psychological Science* 18 (5), pp. 429–434. DOI: 10.1111/j.1467-9280.2007.01917.x.

Schwartz, D., Fischhoff, B., Krishnamurti, T., and Sowell, F. (2013). “The Hawthorne Effect and Energy Awareness.” *Proceedings of the National Academy of Sciences* 110 (38), pp. 15242–15246. DOI: 10.1073/pnas.1301687110.

Schwartz, D., Milfont, T. L., and Hilton, D. (2019). “The Interplay between Intrinsic Motivation, Financial Incentives and Nudges in Sustainable Consumption.” In: *A Research Agenda for Economic Psychology*. Edited by K. Gangl and E. Kirchler. Elgar Research Agendas. Cheltenham, UK: Edward Elgar Publishing, pp. 87–103. ISBN: 978-1-78811-606-0. DOI: 10.4337/9781788116060.

Schwartz, S. H. (1977). “Normative Influences on Altruism.” In: *Advances in Experimental Social Psychology*. Edited by L. Berkowitz. Volume 10. New York, NY, USA: Academic Press, pp. 221–279. ISBN: 978-0-12-015210-0.

Screpanti, E. and Zamagni, S. (2005). *An Outline of the History of Economic Thought*. 2nd Revised and Expanded Edition. Oxford, UK: Oxford University Press. ISBN: 978-0-19-927913-5.

Seo, E. H. (2012). "Cramming, Active Procrastination, and Academic Achievement." *Social Behavior and Personality: An International Journal* 40 (8), pp. 1333–1340. DOI: 10.2224/sbp.2012.40.8.1333.

Shmueli, G. (2010). "To Explain or to Predict?" *Statistical Science* 25 (3), pp. 289–310. DOI: 10.1214/10-STS330.

Shmueli, G. and Koppius, O. R. (2011). "Predictive Analytics in Information Systems Research." *MIS Quarterly* 35 (3), pp. 553–572. DOI: 10.2307/23042796.

Sierzchula, W., Bakker, S., Maat, K., and van Wee, B. (2014). "The Influence of Financial Incentives and Other Socio-Economic Factors on Electric Vehicle Adoption." *Energy Policy* 68, pp. 183–194. DOI: 10.1016/j.enpol.2014.01.043.

Sievertsen, H. H., Gino, F., and Piovesan, M. (2016). "Cognitive Fatigue Influences Students' Performance on Standardized Tests." *Proceedings of the National Academy of Sciences* 113 (10), pp. 2621–2624. DOI: 10.1073/pnas.1516947113.

Simon, H. A. (1955). "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69 (1), pp. 99–118. DOI: 10.2307/1884852.

Sirois, F. M., Melia-Gordon, M. L., and Pychyl, T. A. (2003). "'I'll Look after My Health, Later': An Investigation of Procrastination and Health." *Personality and Individual Differences* 35 (5), pp. 1167–1184. DOI: 10.1016/S0191-8869(02)00326-4.

Smith, E. R. and DeCoster, J. (2000). "Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems." *Personality and Social Psychology Review* 4 (2), pp. 108–131. DOI: 10.1207/S15327957PSPR0402_01.

Solomon, L. J. and Rothblum, E. D. (1984). "Academic Procrastination: Frequency and Cognitive-Behavioral Correlates." *Journal of Counseling Psychology* 31 (4), pp. 503–509. DOI: 10.1037/0022-0167.31.4.503.

Sønderlund, A. L., Hughes, E., and Smith, J. (2019). "The Efficacy of Learning Analytics Interventions in Higher Education: A Systematic Review." *British Journal of Educational Technology* 50 (5), pp. 2594–2618. DOI: 10.1111/bjet.12720.

Steel, P. (2007). "The Nature of Procrastination: A Meta-Analytic and Theoretical Review of Quintessential Self-Regulatory Failure." *Psychological Bulletin* 133 (1), pp. 65–94. DOI: 10.1037/0033-2909.133.1.65.

Steel, P. and König, C. J. (2006). "Integrating Theories of Motivation." *Academy of Management Review* 31 (4), pp. 889–913. DOI: 10.5465/amr.2006.22527462.

Steg, L. and de Groot, J. (2010). "Explaining Prosocial Intentions: Testing Causal Relationships in the Norm Activation Model." *British Journal of Social Psychology* 49 (4), pp. 725–743. DOI: 10.1348/014466609X477745.

Steg, L. and Nordlund, A. (2019). "Theories to Explain Environmental Behaviour." In: *Environmental Psychology: An Introduction*. Edited by L. Steg and J. I. M. de Groot. 2nd Edition. BPS Textbooks in Psychology. Hoboken, NJ, USA: John Wiley & Sons, pp. 217–227. ISBN: 978-1-119-24111-9. DOI: 10.1002/9781119241072.

Stingl, C. (2023). “Using Digital Performance Feedback for Behavior Change: A Literature Review.” *Proceedings of the 29th Americas Conference on Information Systems*. Panama City, Panama.

Stingl, C., Günther, S. A., and Staake, T. (2021). “A Feedback Information System for Improving Hand Hygiene on a Personal and Organizational Level.” *Proceedings of the 16th International Conference on Wirtschaftsinformatik*. Duisburg, Germany.

Stingl, C., Günther, S. A., and Staake, T. (2022). “The Behavioral Mechanisms Behind Feedback – A Preliminary Model for Quantifying Cause-Effect Relationships.” *Proceedings of the 30th European Conference on Information Systems*. Timișoara, Romania.

Storey, V. C., Lukyanenko, R., Maass, W., and Parsons, J. (2022). “Explainable AI: Opening the Black Box or Pandora’s Box?” *Communications of the ACM* 65 (4), pp. 27–29. DOI: 10.1145/3490699.

Sudarshan, A. (2017). “Nudges in the Marketplace: The Response of Household Electricity Consumption to Information and Monetary Incentives.” *Journal of Economic Behavior & Organization* 134, pp. 320–335. DOI: 10.1016/j.jebo.2016.12.015.

Sütterlin, B., Brunner, T. A., and Siegrist, M. (2011). “Who Puts the Most Energy into Energy Conservation? A Segmentation of Energy Consumers Based on Energy-Related Behavioral Characteristics.” *Energy Policy* 39 (12), pp. 8137–8152. DOI: 10.1016/j.enpol.2011.10.008.

Sweller, J. (2011). “Cognitive Load Theory.” In: *Psychology of Learning and Motivation*. Edited by J. P. Mestre and B. H. Ross. Volume 55. Cognition in Education. San Diego, CA, USA: Academic Press, pp. 37–76. ISBN: 978-0-12-387691-1.

Tang, S.-H. and Hall, V. C. (1995). “The Overjustification Effect: A Meta-Analysis.” *Applied Cognitive Psychology* 9 (5), pp. 365–404. DOI: 10.1002/acp.2350090502.

Taube, O. and Vetter, M. (2019). “How Green Defaults Promote Environmentally Friendly Decisions: Attitude-conditional Default Acceptance but Attitude-unconditional Effects on Actual Choices.” *Journal of Applied Social Psychology* 49 (11), pp. 721–732. DOI: 10.1111/jasp.12629.

Tiefenbeck, V. (2014). “Behavioral Interventions to Reduce Residential Energy and Water Consumption: Impact, Mechanisms, and Side Effects.” PhD thesis. Zurich, Switzerland: ETH Zurich. DOI: 10.3929/ethz-a-010263933.

Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., and Staake, T. (2018). “Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation.” *Management Science* 64 (3), pp. 1458–1476. DOI: 10.1287/mnsc.2016.2646.

Tiefenbeck, V., Wörner, A., Schöb, S., Fleisch, E., and Staake, T. (2019). “Real-Time Feedback Promotes Energy Conservation in the Absence of Volunteer Selection Bias and Monetary Incentives.” *Nature Energy* 4 (1), pp. 35–41. DOI: 10.1038/s41560-018-0282-1.

Tisocco, F. and Liporace, M. F. (2023). “Structural Relationships Between Procrastination, Academic Motivation, and Academic Achievement Within University Students: A Self-

Determination Theory Approach." *Innovative Higher Education* 48 (2), pp. 351–369. DOI: 10.1007/s10755-022-09622-9.

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., and Vandenbergh, M. P. (2014). "Positive and Negative Spillover of Pro-Environmental Behavior: An Integrative Review and Theoretical Framework." *Global Environmental Change* 29, pp. 127–138. DOI: 10.1016/j.gloenvcha.2014.09.004.

Tuckman, B. W. (2005). "Relations of Academic Procrastination, Rationalizations, and Performance in a Web Course With Deadlines." *Psychological Reports* 96 (3), pp. 1015–1021. DOI: 10.2466/pr0.96.3c.1015-1021.

Tversky, A. and Kahneman, D. (1981). "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481), pp. 453–458. DOI: 10.1126/science.7455683.

Van der Kleij, F. M., Feskens, R. C. W., and Eggen, T. J. H. M. (2015). "Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis." *Review of Educational Research* 85 (4), pp. 475–511. DOI: 10.3102/0034654314564881.

Van Valkengoed, A. M., Abrahamse, W., and Steg, L. (2022). "To Select Effective Interventions for Pro-Environmental Behaviour Change, We Need to Consider Determinants of Behaviour." *Nature Human Behaviour* 6 (11), pp. 1482–1492. DOI: 10.1038/s41562-022-01473-w.

Venkatesh, V., Brown, S. A., and Bala, H. (2013). "Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems." *MIS Quarterly* 37 (1), pp. 21–54. DOI: 10.25300/MISQ/2013/37.1.02.

Vesely, S. and Klöckner, C. A. (2020). "Social Desirability in Environmental Psychology Research: Three Meta-Analyses." *Frontiers in Psychology* 11, Article nr. 1395. DOI: 10.3389/fpsyg.2020.01395.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. (2009). "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process." *Proceedings of the 17th European Conference on Information Systems*. Verona, Italy.

Von Neumann, J. and Morgenstern, O. (2007). *Theory of Games and Economic Behavior*. 60th Anniversary Commemorative Edition. Princeton Classic Editions. Princeton, NJ, USA: Princeton University Press. ISBN: 978-1-4008-2946-0. DOI: 10.1515/9781400829460.

Vosniadou, S. (2020). "Bridging Secondary and Higher Education – The Importance of Self-regulated Learning." *European Review* 28 (S1), pp. S94–S103. DOI: 10.1017/S1062798720000939.

Wager, S. and Athey, S. (2018). "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523), pp. 1228–1242. DOI: 10.1080/01621459.2017.1319839.

Walters, R. and Martin, P. (2013). "Crime and the Commodification of Carbon." In: *Emerging Issues in Green Criminology: Exploring Power, Justice and Harm*. Edited by R. Walters, D. S.

Westerhuis, and T. Wyatt. Critical Criminological Perspectives. Basingstoke, UK: Palgrave Macmillan, pp. 93–107. ISBN: 978-1-137-27399-4. DOI: 10.1057/9781137273994.

Wanner, T. and Palmer, E. (2015). "Personalising Learning: Exploring Student and Teacher Perceptions about Flexible Learning and Assessment in a Flipped University Course." *Computers & Education* 88, pp. 354–369. DOI: 10.1016/j.compedu.2015.07.008.

Weber, R. (2004). "Editor's Comments: The Rhetoric of Positivism versus Interpretivism: A Personal View." *MIS Quarterly* 28 (1), pp. iii–xii. DOI: 10.2307/25148621.

Webster, J. and Watson, R. T. (2002). "Analyzing the Past to Prepare for the Future: Writing a Literature Review." *MIS Quarterly* 26 (2), pp. xiii–xxiii.

Weinstein, N. D. (1988). "The Precaution Adoption Process." *Health Psychology* 7 (4), pp. 355–386. DOI: 10.1037/0278-6133.7.4.355.

Welton, N. J., Caldwell, D. M., Adamopoulos, E., and Vedhara, K. (2009). "Mixed Treatment Comparison Meta-Analysis of Complex Interventions: Psychological Interventions in Coronary Heart Disease." *American Journal of Epidemiology* 169 (9), pp. 1158–1165. DOI: 10.1093/aje/kwp014.

Werch, C. E., Pappas, D. M., Carlson, J. M., DiClemente, C. C., Chally, P. S., and Sinder, J. A. (2000). "Results of a Social Norm Intervention to Prevent Binge Drinking Among First-Year Residential College Students." *Journal of American College Health* 49 (2), pp. 85–92. DOI: 10.1080/07448480009596288.

Wild, K.-P. and Schiefele, U. (1994). "Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens." *Zeitschrift für Differentielle und Diagnostische Psychologie* 15 (4), pp. 185–200.

Winter, R. (2008). "Design Science Research in Europe." *European Journal of Information Systems* 17 (5), pp. 470–475. DOI: 10.1057/ejis.2008.44.

Wulf, G., Chiviacowsky, S., Schiller, E., and Ávila, L. T. G. (2010). "Frequent External-Focus Feedback Enhances Motor Learning." *Frontiers in Psychology* 1, Article nr. 190. DOI: 10.3389/fpsyg.2010.00190.

Wulf, G., Shea, C. H., and Matschiner, S. (1998). "Frequent Feedback Enhances Complex Motor Skill Learning." *Journal of Motor Behavior* 30 (2), pp. 180–192. DOI: 10.1080/00222899809601335.

Yamada, M., Goda, Y., Matsuda, T., Saito, Y., Kato, H., and Miyagawa, H. (2016). "How Does Self-Regulated Learning Relate to Active Procrastination and Other Learning Behaviors?" *Journal of Computing in Higher Education* 28 (3), pp. 326–343. DOI: 10.1007/s12528-016-9118-9.

Zeiler, M. D. and Fergus, R. (2014). "Visualizing and Understanding Convolutional Networks." *Proceedings of the 13th European Conference on Computer Vision*. Zurich, Switzerland. DOI: 10.1007/978-3-319-10590-1_53.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Red. by M. Gail, K. Krickeberg, J. M. Samet,

A. Tsiatis, and W. Wong. Statistics for Biology and Health. New York, NY, USA: Springer. ISBN: 978-0-387-87458-6. DOI: 10.1007/978-0-387-87458-6.

Chapter 1: Instantiation and validation

Paper I

NIWM: Non-intrusive water monitoring to uncover heat energy use in households

Samuel Schöb
University of Bamberg

Sebastian A. Günther
University of Bamberg

Karl Regensburger
LAGRAR

Thorsten Staake
University of Bamberg

Computer Science – Research and Development 33 (1-2), pp. 127–133, 2018, https://doi.org/10.1007/s00450-017-0353-8

Paper II

Empowering personalized feedback on hot water usage: A field study with shower meters

Sebastian A. Günther
University of Bamberg

Carlo Stingl
University of Bamberg

Vlad C. Coroamă
ETH Zurich

Samuel Schöb
Massachusetts Institute of Technology

Thorsten Staake
University of Bamberg

Proceedings of the 35th Annual ACM Symposium on Applied Computing, pp. 763–766, Brno, Czech Republic, 2020, https://doi.org/10.1145/3341105.3374053

Paper III

A feedback information system for improving hand hygiene on a personal and organizational level

Carlo Stingl
University of Bamberg

Sebastian A. Günther
University of Bamberg

Thorsten Staake
University of Bamberg

Proceedings of the 16th International Conference on Wirtschaftsinformatik, Duisburg, Germany, 2021, https://aisel.aisnet.org/wi2021/WCreating/Track04/7/

Chapter 2: Evaluation and recommendations

Paper IV

The behavioral response to a corporate carbon offset program: A field experiment on adverse effects and mitigation strategies

Sebastian A. Günther
University of Bamberg

Thorsten Staake
University of Bamberg

Samuel Schöb
University of Bamberg

Verena Tiefenbeck
University of Erlangen-Nuremberg

Global Environmental Change 64, Article nr. 102123, 2020, https://doi.org/10.1016/j.gloenvcha.2020.102123

Paper V

All eyes on consumption feedback: A randomized controlled trial on the interplay of financial incentives, environmental appeals, and consumption feedback

Sebastian A. Günther
University of Bamberg

Samuel Schöb
University of Bamberg

Vlad C. Coroamă
ETH Zurich

Verena Tiefenbeck
University of Erlangen-Nuremberg

Friedemann Mattern
ETH Zurich

Thorsten Staake
University of Bamberg

Submitted to Experimental Economics (under review)

Introduction

Everyday activities like choosing from a menu, taking a shower, or operating a production machine are typically perceived as insignificant and inconsequential actions. Yet, given their frequency and pervasiveness, when aggregated, related decisions shape important outcomes such as national healthcare costs, global carbon emissions, and firms' profitability. Consequently, a comprehensive strand of research has emerged that aims at encouraging "desirable" behaviors for the enormous number of everyday activities in both personal and corporate contexts (e.g., Allcott, 2011; Berger et al., 2018; Chadi et al., 2022; Hoffmann and Thommes, 2020; Kramer et al., 2019; Zizzo et al., 2021). To this end, many studies explore the impact of classical instruments for behavior change, including financial incentives to increase the overall utility of acting responsibly (e.g., providing incentives to exercise, carbon emission taxes; see Charness and Gneezy, 2009; Grieder et al., 2021) or campaigns to raise awareness of behavioral consequences (e.g., environmental consequences, consequences of preventable diseases; see Dolnicar et al., 2017; O'Keefe and Jensen, 2007).

Recently and in addition to the established instruments, great hopes are placed in digital technologies to achieve the triple goal of individual-level effectiveness, scalability, and social acceptability of behavior change campaigns (Burton-Jones et al., 2021). With the ubiquity of digital technologies, numerous data streams are now available that—given the technology's low transaction costs—make highly specific interventions practically viable even for activities whose single execution seems to be of small consequence, yet which are very widespread and occur frequently. Thus, when taken together, these activities are indeed impactful at the macro level. In the environmental context where an individual's resource usage is in particular often quantifiable, automated sensing and processing capabilities enable the provision of timely feedback to individuals on their everyday activities, making the environmental consequences of one's activities salient (e.g., carbon emissions associated with car driving or the use of air conditioning). Several studies have shown that feedback interventions on activities that recipients can easily influence can have large, cost-effective conservation effects (see, e.g., Brülisauer et al., 2020; Gerster et al., 2020; Tiefenbeck et al., 2018). At the same time, the declining transaction costs of digital interventions allow for small but frequent financial incentives as performance-contingent rewards and situation-specific moral appeals to foster behavior change.

While financial incentives, moral appeals, and recently also timely (even real-time or "live") feedback understandably received considerable attention in academia, studies that investigated feedback interventions, financial incentives, and moral appeals side by side are still sparse. To the best of our knowledge, no empirical study has been published yet in the environmental domain that explores whether financial incentives and moral appeals substantially moderate the impact of feedback. As a result, it is still unclear whether and how these instruments should be combined.

From a theoretical viewpoint, analyzing these three interventions in combination is interesting, as predictions from three strands of intervention-associated theories are in conflict and diverge. First, standard economic theory predicts a complementary effect of feedback and financial incentives. The conjecture is that the transparency created by feedback can help individuals to reap the economic benefits of financial incentives. In classical economics, this could even be true if the financial incentives per activity are very small, which is a typical practical requirement given the large number of everyday actions. In contrast to that first perspective, a second viewpoint is that small financial incentives might not be able to trigger the desirable behavior as predicted by standard economic theory, but could on the contrary even induce adverse effects. This widespread concern primarily emerges from a large strand of literature in psychology that goes back more than fifty years and is now grouped under the notion of motivation crowding theory. Motivation crowding theory postulates that financial incentives undermine a person's intrinsic motivation (a motor that feedback could leverage), namely the process of seeking an associated task or activity for one's internal motives (e.g., interest, enjoyment) and not for external motives such as monetary rewards (e.g., Deci, 1971). Putting this into the context of environmental campaigns, the theory suggests that financial incentives could actually undermine existing conservation efforts*. A third viewpoint that builds upon feedback intervention theory (Kluger and DeNisi, 1996) leads to yet another prediction. Feedback intervention theory emphasizes that individuals have limited cognitive resources. Therefore, additional incentives do not necessarily moderate the impact of feedback. If this assumption of feedback information theory holds, environmental campaigns do not have to provide financial benefits for resource conservation as feedback could be the dominant factor facilitating behavior change.

A deeper understanding of the interplay between feedback and financial incentives is also crucial for policymakers and managers that can (and need to) leverage more and more fine-grained behavioral data in a plethora of domains including energy (smart electricity meters), health (devices measuring physical activity), and safety (products making data on driving behavior) to improve related outcomes.

Against this backdrop, we conducted a four-month randomized controlled field trial with 326 individuals (20,830 observations) using a 3 (financial incentives, appeals, no cues) $\times$ 2 (feedback, no feedback) design to assess the effect sizes of different interventions as well as to test the explanatory power of standard economic theory, motivation crowding theory, and feedback intervention theory. We chose a target activity that has a relatively high impact on the environment. Showering is an everyday activity that consumes considerable amounts of energy, as hot water represents the second largest energy use in households after space heating (see,

* Numerous studies have investigated motivation crowding theory in the field, mostly focusing on the extensive margin for relatively infrequently performed activities such as for blood donation, gym attendance, or online reviewing. In contrast, our paper focuses on an everyday activity in the environmental domain (the intensive margin). In this setting, the literature is still lacking a rigorous analysis on the impact of financial incentives and their interplay with feedback.

e.g., European Statistical Office, 2023). More precisely, showering alone accounts for around 80% of domestic hot water use (Bertrand et al., 2017) and typically relies on carbon-intensive fossil fuels (European Statistical Office, 2023). An average shower from our sample consumes 2.75 kWh in less than 6 minutes, which is more than the average daily consumption by lighting in a typical EU household (Odyssee-Mure, 2021). Yet, most individuals are not aware of the resource-intensity of showering (Tiefenbeck et al., 2018), making it both an interesting and relevant target activity to explore the behavioral responses to different interventions.

Our study makes four important contributions. First, as predicted by the standard economic model, we find that financial incentives alone lead to resource conservation. Even when controlling for important confounds not considered in previous studies (i.e., ex-ante intrinsic motivation for resource conservation, financial situation of participants, fatigue effects), we do not find any evidence of motivation crowding theory. While financial incentives have a modest impact, we find that appeals pronouncing environmental consequences have no meaningful impact on resource conservation, corroborating the results of macro- and microeconomic studies (see, e.g., Dolnicar et al., 2017; Luyben, 1982). Second, feedback information far outperforms the financial incentive in promoting resource conservation. This is especially interesting as participants in our experiment pay a flat fee for their resource consumption and therefore have no financial benefits in responding to feedback per se. Third, we find that the provision of financial incentives does not amplify the impact of feedback on resource conservation (which appears to be inconsistent with the standard economic model given that the tested financial incentives alone induce resource conservation). A plausible explanation for the observation that financial incentives promoted resource conservation but did not amplify the impact of feedback is provided by resource allocation theory (Kanfer and Ackerman, 1989), which extends the outlined theories by theorizing at a more general level (i.e., not dependent on specific interventions such as feedback). According to resource allocation theory, individuals have a limited amount of attention that they can allocate to external cues: Therefore, feedback information could be so dominant that financial incentives, even though they are effective in isolation, can only generate little additional attention from recipients, making them ineffective for inducing substantially higher resource conservation. Fourth, our experiment demonstrates that digital technologies provide an intriguing platform to study and compare interventions over time. This may inspire future researchers to not only deepen our understanding of resulting behavioral mechanisms, but also to continue developing digital tools that could substantially improve personal, organizational, and societal outcomes at scale.

Literature overview & research gaps

While numerous field experiments tested the effects of behavioral interventions on resource conservation, the literature falls short on a more in-depth understanding (i) of the interplay between feedback and financial incentives and (ii) of appeals as a non-monetary alternative to financial incentives. To begin, the environmental studies that reward pro-environmental behavior through financial incentives are usually motivated by testing motivation crowding theory. Instead of inducing resource conservation through a potential higher level of utility of behavior change, the theory implies that financial incentives could backfire by harming intrinsic motivation (in this case the motivation for resource conservation). There are indeed many mechanisms that could explain such adverse behavioral responses. For instance, an individual might feel controlled by the provision of a financial incentive (Deci, 1972), an individual might attribute the causes of their (past) behavior to the presence of an external reward (Tang and Hall, 1995), or the financial incentive could distract individuals from the moral dimension of behavior and push them toward an evaluation of whether the financial incentive is worth the efforts (Heyman and Ariely, 2004). Another reason for adverse effects could be related to the concept of image motivation. Individuals often engage in prosocial behavior because of a desire to be liked and respected by others based on their altruistic actions (Ariely et al., 2009). The presence of financial incentives for prosocial behavior, however, may cast doubt on whether one performs the behavior primarily for societal outcomes or for purely egoistic reasons (i.e., financial benefits). Following this reasoning, financial incentives could crowd out desirable behaviors, as individuals do not want to appear “greedy” or selfish (Bénabou and Tirole, 2006).

Studies in the environmental domain testing the effects of mere financial incentives usually indicate that even relatively small performance-contingent financial incentives increase the utility of eliciting pro-environmental behavior. For instance, studies on residential water or electricity use (e.g., Azarova et al., 2020; de Sousa and Dias Fouto, 2019; Royal and Rustamov, 2018; Wang et al., 2022; Weber et al., 2017) or on resource use in corporate contexts (e.g., Günther et al., 2020; Mizobuchi and Takeuchi, 2013; Schall and Mohnen, 2017) indicate that financial incentives rather promote than crowd out pro-environmental behaviors. It is important to note that such studies are quite heterogeneous in incentivizing individuals. Specifically, the amount and the mode of payment (e.g., direct payments, vouchers) make it difficult to anticipate under which conditions a specific financial incentive actually increases individuals’ utility of acting pro-environmentally. Yet, literature suggests that financial incentives should not be too low (Gneezy and Rustichini, 2000).

In the studies that combine financial incentives with feedback and compare it to an alternative set of interventions, financial incentives tend to attenuate the impact of consumption feedback on resource conservation. For instance, Sudarshan (2017) reports that the additional provision of financial incentives cancels out the conservation effects of comparative feedback that has been observed in a different experimental group. Handgraaf et al. (2013) show in an experiment that consumption feedback with monetary rewards is ineffective in inducing conservation (i.e., reducing computer energy usage at the workplace) compared to a feedback intervention comprising social rewards. A similar observation is reported in a study that tests framing effects instead of providing additional financial incentives for behavior change. Specifically, Asensio and Delmas (2015) report that feedback highlighting financial benefits of electricity conservation is substantially less effective in inducing electricity conservation than feedback highlighting negative health and environmental externalities associated with electricity production.

To summarize, vast amounts of research have been conducted to understand the role of financial incentives or feedback, and many studies highlight their potential to induce pro-environmental behavior. However, there is still a lack of understanding from both theoretical and practical side. So far, most studies report an overall behavioral response to financial incentives but they do not consider that a crowding-out effect may only occur for those individuals who were ex-ante intrinsically motivated (as discussed by Schwartz et al., 2019). At the same time, no study has yet compared the impact of financial incentives to environmental appeals or feedback information, which potentially could be cost-efficient alternatives in promoting behavior change. Such an analysis may also be valuable for testing heterogeneous effects, since the effects of the interventions could be quite different depending on participants' attitudes to the target behavior. For instance, while environmental appeals might work only for individuals with high intrinsic motivation for resource conservation, financial incentives could only be effective for individuals with low intrinsic motivation and otherwise harm curtailment efforts. Beyond that, no study has yet holistically investigated whether financial incentives or environmental appeals enhance the behavioral effects induced by feedback. An analysis of the additivity of such factors requires not only to assess the individual effects of financial incentives or environmental appeals before combining them with feedback, but also to assess the effect of feedback in isolation. This, in turn, could shed light on (i) whether financial incentives increase the conservation effects of feedback, (ii) whether environmental appeals are an equally effective non-monetary intervention, or (iii) whether feedback is dominant over such motivational cues. From a theoretical perspective, such an analysis also tests associated theoretical assumptions, namely that individuals have unlimited information-processing capabilities and will exert more efforts when feedback is combined with financial incentives (as predicted by the standard economic model under the assumption that financial incentives alone increase the utility of conservation). In contrast, it could be that the additional provision of financial incentives does not necessarily unlock higher conservation effects in response to feedback, as individuals are bounded in cognitive resources (as described by feedback intervention theory; Kluger and DeNisi, 1996).

Empirical evaluation

Recruitment

To address the outlined research gaps, we recruited 398 participants from three student dorms near Nuremberg, Germany for the experiment*. Each participant lived in a single apartment

* While we aimed to recruit as many participants as possible, we estimated with the help of former longitudinal studies on showering (e.g., Fang et al., 2023; Tiefenbeck et al., 2018) that we require at least 60 participants per group (the control group and the three treatment groups) to detect small effects (i.e., 5% conservation

with a private shower. Recruiting started on April 2, 2018, one week before the start of the summer semester, by hanging posters in prominent spots in the student dorms and sending letters to the inhabitants. These promoted the study as a research program to better understand the use of hot water in showers, emphasizing that showering is cost- and resource-intensive. Moreover, the posters and letters emphasized that participants had an opportunity to win one of 10 available energy trackers for the shower, and informed individuals that their data would be treated confidentially and would not be shared with the respective dorm management.

A week after this announcement, we began approaching the residents individually at their apartments to ask whether they would like to participate in the study. To this end, residents had to fill out a short questionnaire and give their consent for data on their showering behavior to be collected and analyzed. During the three-week recruitment process, we encountered 543 out of the 608 students living in the dorms, and 398 of them agreed to participate (73.3% of those encountered). We implemented three experimental conditions and one control condition. Participants were assigned to these groups based on a custom stratified randomization approach that aimed to distribute the groups randomly and as evenly as possible across the floors of the student dorms (to take infrastructure-related confounding variables into account). As a consequence, participants could be assigned to the control group or to the financial, environmental, or to the general treatment group. As we depended on participants' mobile phone numbers for the interventions, there was one exception made to the randomization: Those who did not want to share their phone numbers were assigned to a fifth group, the privacy-sensitive group, whose participants also received no treatment throughout the study (like in the control condition). A total of 11 participants (2.8%) did not share their phone numbers and were assigned to this group.

Implementation of interventions

During the recruitment process, all participants received a shower meter that the research assistants mounted between the shower hose and the hand-held shower head. The shower meter measured and recorded the water consumption and water temperature of every shower and, additionally, transmitted live data (i.e., the amount of water and energy consumed up to that point during an ongoing shower) to gateways that were previously placed at strategic locations across the student dorms (outside the apartments). These gateways transmitted the data for storage to a server, which potentially then triggered subsequent interventions

with an alpha value of 10%). A reason for the seemingly small required sample size is that the variance in shower consumption is much smaller than, for example, in related conservation studies on household electricity use. A potential reason for that is that the dependent variable "household electricity use" is subject to standby loads and various other sources of noise, which reduce the observed behavioral effect. In contrast, we measure the behavioral responses directly at the place of action that we target (i.e., showering). Nevertheless, we further conducted a formal power analysis for multiple regression (see, e.g., Cohen, 1988): At a power level of 80%, an alpha value of 10%, a relatively small f^2 effect size of 6.5%, and 15 independent variables (that result from the five comparisons between the control group and each of the three treatment groups; see Section 5), the analysis suggested an overall sample size of 251 participants.

(e.g., environmental appeals or outcome feedback) depending on the group assignment of a participant.

The functionality of the shower meter goes far beyond the mere measurement of consumption: It has an adjustable screen that turns on as soon as the water begins flowing and that can display various content, such as feedback on resource consumption. The energy for its operation is harvested from the water that flows through the meter itself, so the device does not require a battery. And because it does not have a battery, the shower meter only operates for up to three minutes after an interruption to the water flow; thereafter the meter and the screen turn off. Irrespective of shorter interruptions to the water flow (less than 3 minutes), the meter considers water extractions until it turns off as one coherent shower and stores the respective measurement data in a chronological order. While the meter cannot record time stamps from the shower (due to the absence of a battery), the gateways and the server enrich the received measurement data with time stamps.

Implementation of financial incentives and environmental appeals

The first interventions were the financial incentives and environmental appeals. These messages were sent via text messages (SMS) to the participants depending on their treatment condition.

The financial message announced the possibility that participants could now earn money by taking shorter showers. As an anchor for the financial incentives, the text message informed them that they would receive €0.11 per shower (approximately $0.13 at that time) if they cut their average shower time by one minute. The financial incentives of this anchor were calculated on the basis of the average costs of showering from a previous study (Tiefenbeck et al., 2018) and, thus, served in our study as an estimate of the costs of showering. In addition, the financial message contained one hyperlink that pointed to a specific web page outlining the cost intensity of showering by projecting its costs over one year. To rule out potential adverse effects (i.e., participants exploiting the financial incentives by purposefully creating many small water extractions that might be considered to be real showers), the web page informed them that the financial savings are calculated for a maximum of one shower per day, taking the longest shower of each day into consideration. Lastly, participants were informed that they would have to register to receive their remuneration (by entering their bank account details on the web page) and that the remuneration would be paid out at the end of the study.

The environmental appeals followed the same structure. First, a text message prompted individuals to help the environment by taking shorter showers. To illustrate the energy-intensity of showering, the message informed participants that they could save about 140 g of CO_2 emissions if they cut their shower by one minute. This statement was followed by an analogy stating that this is equivalent to the emissions from watching television for four hours. Moreover, the appeals contained a hyperlink directing participants to a web page that described the environmental impact of showering. On this web page, two further analogies highlighted the energy intensity of showering, and its environmental impact in terms of energy consumption was projected over a period of one year. At the bottom of the web page, participants had an opportunity to register for a personal energy and CO_2 report on their shower consumption, to be distributed after the completion of the study. The welcoming message reminded participants in this group of being part of an experiment and informed them that they would soon receive information about their water and energy use in the shower. The accompanying web page displayed general information about the study, such as the timeline and the data handling policy in terms of privacy and confidentiality.

Implementation of consumption feedback

The second, third, and fourth interventions were outcome feedback, outcome feedback combined with real-time feedback, and real-time feedback alone, respectively, which participants received depending on their experimental condition and their study progress (see next subsection). The different intensities of feedback allowed us to explore whether the potential controlling nature of the financial reward and its adverse effects could become stronger depending on the immediacy of the accompanying interventions.

Outcome feedback was sent via text messages to the mobile phones of participants in the treatment groups and informed them of the total water (in liters*) and energy usage (in kWh) of their most recent shower. To this end, the infrastructure monitored the live consumption of the shower meters and triggered a text message as soon as water extractions were interrupted for more than three minutes. This was based on the principle that a meter considers a shower as completed if the water flow is interrupted for more than three minutes.

Shower meters in feedback mode displayed certain pieces of information to participants while they were taking a shower (see Figure 1). The device displayed (i) the amount of water used (in 0.1 liter increments) since the start of the shower, (ii) the current temperature of the water flow in degrees Celsius, (iii) an energy efficiency class rating the current shower with the letters A (most efficient) to G (least efficient), (iv) and a polar bear on an ice floe, which shrinks depending on the amount of energy used. As soon as the water flow is interrupted, the display toggles between water usage (in liters) and energy usage (in [k]Wh).

Experimental design

Participants in the control group received neither appeals nor consumption feedback throughout the course of the study. The only manipulation of the control group was the presence of a meter in the shower that displayed the current water temperature. We deliberately decided to display the current water temperature rather than a blank screen, so that participants in the control group were aware that the device was working correctly and could notify us in the

* Please note that in Europe, the basic unit of volume is the liter. One liter is equal to approximately 0.2642 US gallons.

Figure 1: Provision of real-time feedback on the shower meter screen. Credit: Amphiro AG

event of a malfunction. Figure 2 illustrates the timing of the experimental conditions on the financial, environmental, and general treatment group in comparison with the control group.

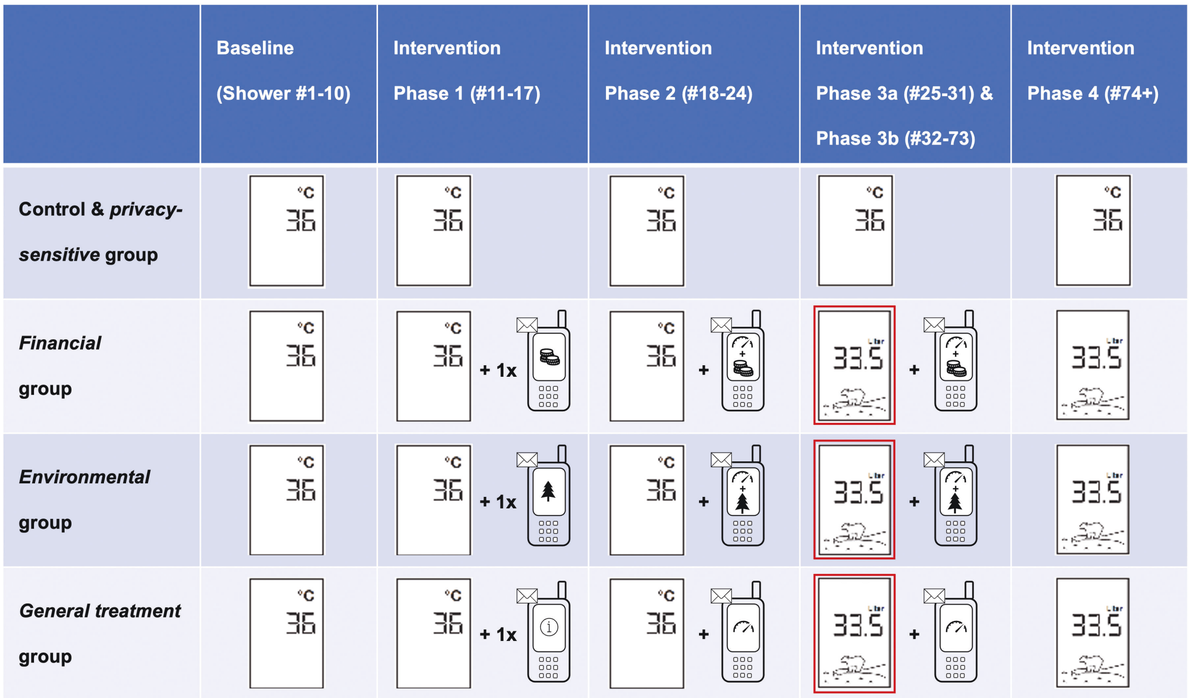


Figure 2: Experimental design and timing of the treatments

To establish a baseline, during the first ten showers, prior to any behavioral intervention, we measured the consumption of the treatment groups under the same study conditions as the control group (i.e., only current water temperature displayed, no text messages). Once a shower meter had recorded ten showers, a text message was sent to the respective participant with differing content based on their group: The financial group received one message promising financial incentives for behavior change, the environmental group received one environmental appeal, and the general treatment group received one welcoming message reminding them of their study participation (see Section 5). After a total of 17 showers, outcome feedback started to be sent via text messages to the participants after each shower. This feedback was supplemented with a financial incentive for the financial group and with an environmental appeal for the environmental group (see Table 1). After the next seven showers, the shower meter switched to feedback mode. In this phase, participants received both real-time feedback while showering and outcome feedback after the completion of a shower. After 48 showers in this phase, no further outcome feedback reports were sent to the participants, but participants continued receiving real-time feedback until the end of the study.

Table 1: Interventions sent via text messages

Group	#11-17	#18-72
Financial group	Earn money by taking shorter showers. You get 11 cents per shower if you shorten your average shower time by 1 minute. Find more information at [link].	You used X liters and Y kWh of energy in your most recent shower. Earn money by taking shorter showers. You get 11 cents per shower if you shorten your average shower time by 1 minute. Find more information at [link].
Environmental group	Help the environment by taking shorter showers. By reducing your shower time by one minute, you save about 140 g CO2 per shower (= emissions from watching TV for 4 hours). Find more information at [link].	You used X liters and Y kWh of energy in your most recent shower. Help the environment by taking shorter showers. By reducing your shower time by one minute, you save about 140 g CO2 per shower (= emissions from watching TV for 4 hours). Find more information at [link].
General treatment group	Welcome to a study on energy and water use in the shower by the University of Bamberg. After some days, we will start sending you information about your water and energy use. Find more information at [link].	You used X liters and Y kWh of energy in your most recent shower. Find more information at [link]

Measurement data and questionnaire data

For each shower taken, the shower meter stores an identification number, the water consumption (in liters), the average temperature in degrees Celsius, and the shower time. Moreover, the meters calculated the required minimum energy use to heat the water based on the standard engineering formula for heat energy ($Q = m * c_p * \Delta T$, with heat energy Q, mass of water m, specific heat capacity of water c_p, and ΔT the difference between the measured water temperature and the cold water temperature). For an ongoing shower, this information was communicated by the respective shower meter to nearby gateways. For the subsequent evaluation, the average heat losses are taken into account as in (Tiefenbeck et al., 2018).

In a short questionnaire during the recruitment process, we asked participants about their socio-demographics (i.e., age, gender), for their phone number, and posed questions related to their personality and environmental attitude. All participants filled out this questionnaire. In a subsequent longer questionnaire, which was sent at the end of the study per text message, we collected supplemental data about participants' socio-demographics (i.e., money available at the end of the month, monthly budget, sources of income), and asked questions about how participants perceived the shower meter, the text messages (only the treatment groups), and the financial incentives (only the financial group).

Sample and data pre-processing

Out of the initial 398 participants, 48 participants dropped out of the study for technical reasons: 17 shower meters apparently became defective during the course of the study (i.e., broken temperature sensor, data inconsistencies, shower meter was never observed by the communication infrastructure). We excluded 30 participants from the data analysis since they either received no text messages or only a small number of text messages (less than 50%) due to technical problems with the gateways or the delivery of the text messages. Due to the technical implementation of the shower meter, we had to exclude one more participant with extremely short water extractions (between 2 and 5 liters), since the meter considered only a small subset of these water extractions (above 3.5 liters) as real showers: However, a thorough analysis of the recorded live data indicates that most of these small extractions were apparently real showers by the respective participant.

Of the remaining 350 individuals, we further excluded 24 participants in subsequent data cleaning steps: One participant opted out of the study, 20 participants had taken fewer than 25 showers and hence did not reach the third intervention phase, and three participants who took, on average, more than three showers a day. We additionally tested whether the dropout of the participants correlates with the study groups. A corresponding chi-squared test does not indicate statistically significant group differences regarding the dropout of the participants ($X^2 = 6.27$; p value $= .182$).

For the following analyses, we perform some additional data cleaning steps on the 26,993 observations, largely following the filter criteria of (Tiefenbeck et al., 2018). More specifically, we remove data points with a consumption of less than 4.5 liters, as they most likely result from cleaning activities. In the same vein, we remove data points that do not relate to typical showering behavior (e.g., when the average temperature was higher than 47 °C). Lastly, we limit the experimental design to the first 94 observations for each participant, since only one third of the participants showered more often (98 of 326 participants), potentially causing attrition bias in the last intervention phase. A total of 326 participants with 20,830 data points remain for the subsequent analyses, of which 70.6% (230 participants) answered the second questionnaire.

Manipulation checks and knowledge about the interventions

In the second questionnaire, two questions served as manipulation checks for the participants in the treatment groups. A total of 97.1% of the participants who answered the second questionnaire reported that they had received study-related text messages during the experiment. With regard to their content, 90.7% participants of the financial group remembered the financial incentive and 84.2% participants of the environmental group remembered the environmental appeal. The manipulation checks thus indicate that most participants successfully received the behavioral interventions, and that many of them read the text messages at least occasionally.

Descriptive statistics, randomization checks, and collapse check

To explore whether the randomization process successfully created balanced groups across the study groups, in Table 2 we compare the socio-demographic backgrounds, the average resource consumption during the baseline period, and the environmental attitude between the groups. The third and fourth columns contain the mean values for the control group and the privacy-sensitive group, and the fifth to seventh columns contain the mean values for the treatment groups. The eighth column contains the test statistics from a randomization check (F-test with p values in parentheses), which checks the two-sided hypothesis that the correlation of the mean value of the control group and the mean values of the other study groups (i.e., the privacy-sensitive group and the treatment groups) with the individual variable is zero. As the test statistics indicate, the randomization process has indeed produced balanced variables across the different groups.

Next, we explore whether the control group differs from the privacy-sensitive group. Both groups received no treatment during the experiment, but participants in the latter group did not share their phone number. To check whether privacy-sensitive participants differ from participants in the control group, we repeated the randomization check by only taking these two groups into account. As the ninth column of Table 2 shows, we find no evidence that these groups differ in the key characteristics. Consequently, we collapsed the control group and the privacy-sensitive group into a single control group for the subsequent analyses.

Hypotheses

Based on our experimental design, we derive 15 testable hypotheses that specifically target the outlined research gaps. More specifically, we structure our hypotheses along the four intervention phases before turning to potential subgroup effects that we will explore across all intervention phases. In this context, it is important to note that we did not create a pre-analysis plan for the experiment, as this was less common in field research at that time.

In general, our literature review indicates that financial incentives have a desirable effect on resource conservation, even for relatively small payments (see, e.g., Royal and Rustamov,

Table 2: Randomization checks and collapse check

Variable	Full sample	Control group	Privacy-sensitive group	Financial group	Environ-mental group	General treatment group	F-Statistics (p value) All groups	Collapse check
Mean baseline water use per shower (l)	50.96 (37.76)	47.07 (29.74)	52.16 (31.87)	51.80 (43.27)	49.30 (39.47)	55.46 (38.41)	0.537 (.709)	0.278 (.599)
Mean baseline energy use per shower (kWh)	2.82 (2.10)	2.60 (1.71)	2.96 (2.00)	2.87 (2.37)	2.76 (2.25)	3.02 (2.06)	0.434 (.784)	0.421 (.518)
Mean baseline water temperature (°C)	31.50 (2.20)	34.82 (2.19)	35.55 (2.63)	35.20 (2.26)	35.37 (2.14)	34.96 (2.16)	0.845 (.498)	1.023 (.314)
Age	22.63 (3.08)	23.04 (3.01)	22.82 (3.68)	22.58 (2.56)	22.04 (2.92)	22.84 (3.63)	1.172 (.323)	0.049 (.826)
Proportion of women	0.59 (0.49)	0.58 (0.50)	0.55 (0.52)	0.51 (0.50)	0.64 (0.48)	0.62 (0.49)	0.871 (.481)	0.053 (.819)
Environmental attitude[a]	3.22 (0.93)	3.24 (0.90)	3.55 (0.82)	3.19 (0.87)	3.21 (1.04)	3.18 (0.94)	0.407 (.804)	1.106 (.296)
N	326	79	11	79	78	79		

Notes: [a] Scored using five-point Likert scale

2018; Wang et al., 2022). Consequently, we hypothesize that our financial incentives might induce resource conservation through an increase in utility of behavior change. On the other hand, adverse effects might also be possible. Specifically, motivation crowding theory is backed by over fifty years of research, which has systematically shown that extrinsic rewards could crowd out intrinsic motivation for incentivized tasks/activities (see, e.g., Deci et al., 1999). It is important to note that the related literature discussed that context-specific aspects have a strong influence on the behavioral response to financial incentives. For instance, a crowding-out effect may only occur for those individuals who were ex-ante intrinsically motivated (Schwartz et al., 2019). Following this reasoning, incentives could have an antithetical effect in our study, as we target a relatively young sample (university students) who might have stronger motives for environmental protection than the samples of related studies (by being potentially more affected by climate change in the future). Moreover, in comparing children to undergraduates, a highly cited meta-review has provided evidence that crowding-out effects might be stronger for younger individuals, as they might be less able to ignore the controlling aspect of extrinsic incentives (Deci et al., 1999). It hence could be that the crowding-out effect can be seen in students, but not in older individuals, on which related environmental studies tend to focused on (i.e., the studies predominantly focused on the behavioral response of households/employees). Therefore, in the context of our experiment, desirable and adverse overall behavioral responses are both conceivable. We hypothesize:

Hypothesis 1a: The provision of small financial incentives to reduce consumption leads to an increase in consumption.

Hypothesis 1a': The provision of small financial incentives to reduce consumption leads to resource conservation.

In terms of the interventions that should serve as a reference for financial incentives, previous studies in the environmental domain suggest that interventions aimed at filling knowledge gaps have no substantial effects on conservation behavior (Dolnicar et al., 2017; Geller, 1981; Luyben, 1982), in contrast to similar interventions in other behavioral domains (e.g., Apesteguia et al., 2013; Bott et al., 2020). Underlying reasons might be that the previously tested interventions might be too inconspicuous (e.g., in the form of decals in hotel rooms), not delivered in a timely way at the place of action, or given at the macro level, which makes it difficult to capture associated behavioral responses. As the environmental appeal of our study is delivered as a potentially highly salient SMS right after the target behavior (showering), we hypothesize:

Hypothesis 1b: An environmental appeal calling for a reduction in consumption leads to resource conservation.

To control for potential attentional effects on the behavior as well as for recipients' desire to act in socially acceptable ways (i.e., social desirability bias) (Schwartz et al., 2013), we further assess the impact of a placebo message that reminds individuals of their participation in a study.

Hypothesis 1c: A message reminding individuals that they are participating in a study leads to resource conservation.

In the second stage, we provide activity-specific feedback information upon completion of the associated activity. We expect that feedback information that provides tangible numbers enables individuals to put their behavior into context, making these interventions more persuasive overall for inducing change. While we assume that financial incentives and moral appeals have either a subtractive or additive effect on the overall response, we expect that such cues are not necessary for feedback to induce behavior change: Even in the absence of motivational cues, individuals may find inherent reasons for why a behavior change may be appropriate, possibly resulting in a desire to resolve the discrepancy between their conservation behavior highlighted by the intervention and their standards. Indeed, feedback intervention theory predicts that cues may even attenuate the impact of feedback information impact by shifting attention away from the task (Kluger and DeNisi, 1996). It is, however, also conceivable that feedback combined with an environmental appeal leads to a stronger desired change by purposefully strengthening recipients' intrinsic motivation and shifting their attention toward the task (Kluger and DeNisi, 1996). In contrast, when feedback is combined with financial incentives, one could once again argue that the impact could be observed in both directions, resulting in the financial incentives having either an adverse or a desirable impact when compared to the impact of feedback alone. We therefore hypothesize:

Hypothesis 2a: Combining small financial incentives with outcome consumption feedback leads to a lower conservation level than without the incentives.

Hypothesis 2a': Combining small financial incentives with outcome consumption feedback leads to a higher conservation level than without the incentives.

Hypothesis 2b: Combining an environmental appeal with outcome consumption feedback leads to a higher conservation level than without the appeal.

Returning again to the role of feedback information, empirical studies suggest that the effect size depends greatly on the immediacy of the feedback (Tiefenbeck, 2014). The conservation behavior, irrespective of the presence of a motivational cue, should therefore be proportionally stronger when feedback is also provided during the activity. Likewise, it is conceivable that feedback is dominant over financial incentives and environmental appeals. In this case, the financial incentives and the environmental appeals should not induce higher (or lower) effects when being combined with feedback compared to a treatment condition that only receives feedback.

Hypothesis 3: The additional provision of feedback during the activity results in greater effects on resource conservation.

Hypothesis 4: Consumption feedback is dominant over alternative motivational cues (i.e., small financial incentives or environmental appeals) in inducing resource conservation.

The crowding-out literature discusses how the trade-off between the economic effect and the potential controlling nature of the financial incentives might initially mask the damage to intrinsic motivation but after the financial incentives have been removed, then the adverse effects should become observable (see, e.g., Esteves-Sorenson and Broce, 2022). Our research design allows us to capture these effects, as there is a phase in which all treatment groups receive the same treatment (i.e., feedback during the activity). In the case of damage to intrinsic motivation, the group that received the initial financial incentives may then either (i) stop conserving resources or (ii) conserve fewer resources than its reference groups. In contrast, it is also conceivable that there is no undermining effect of financial incentives, meaning that the treatment effect is not statistically different from the conservation behavior of the other reference groups. We hypothesize:

Hypothesis 5: Once financial incentives are discontinued and only the provision of feedback is maintained, individuals stop conserving resources.

Hypothesis 5': Once financial incentives are discontinued and only the provision of feedback is maintained, individuals continue to conserve resources. The conservation effect is as high as if the incentives had never been offered.

Hypothesis 5'': Once financial incentives are discontinued and only the provision of feedback is maintained, individuals continue to conserve resources. The conservation effect is as high as if individuals had received environmental appeals instead of the incentives.

In addition to determining main effects, we explore whether the impact of the interventions differs depending on the recipients' attitude or their financial situation during all of these intervention stages. As the literature suggests, we expect that stronger environmental attitudes or a higher level of wealth are negatively related to the impact of incentives. Moreover, we expect that stronger environmental attitudes are positively related to the impact of the environmental appeals. We hypothesize:

Hypothesis 6a: The higher the intrinsic motivation level at baseline, the stronger the adverse effects of financial incentives on resource conservation.

Hypothesis 6a': The higher the intrinsic motivation level at baseline, the stronger the conservation effects of environmental appeals on resource conservation.

Hypothesis 6b: The better the financial situation of an individual (i.e., money left at the end of a month), the stronger the adverse effects of financial incentives on resource conservation.

Results

Descriptive evidence

Before presenting the estimates of the employed inferential methods, we shed light on the effects of the interventions through descriptive statistics. Figure 3 presents the average change in heat energy use for each study group between the baseline phase and the individual intervention phase. The underlying descriptive statistics have been generated as follows: First, we are calculating the average consumption per shower for each individual in each phase. To subsequently determine the change in consumption for each individual, the average consumption from the baseline phase is then subtracted from the average consumption in each intervention phase. From the yielded change in consumption values for each individual, we then take the overall phase-specific mean value (representing the change from the baseline) and calculate the associated standard error.

As Figure 3 shows, it appears that only financial incentives led to conservation behavior in the absence of consumption feedback; the environmental appeal and the welcoming message had no measurable effect. However, once individuals of the treatment groups receive consumption feedback, the financial incentives appear to have no additional impact on conservation behavior. Interestingly, it seems that environmental appeals or financial incentives are still important over time, as the group receiving solely consumption feedback (i.e., the general treatment group) stopped saving heat energy in the last intervention phase.

Main results

To formally test the effects of the behavioral interventions on the energy use per shower, we first conduct several specification tests for the inferential model. We first compare a pooled regression model to a fixed-effects regression model controlling for both individual and time effects (the fixed-effects model is based on Equation 1). An F-test rejects the null hypothesis that all fixed effects are zero ($F(807, 19999) = 25.9$; p value $< .001$). Apart from that, we checked the residuals for certain sources of biases. A Breusch-Pagan test (Breusch and Pagan,

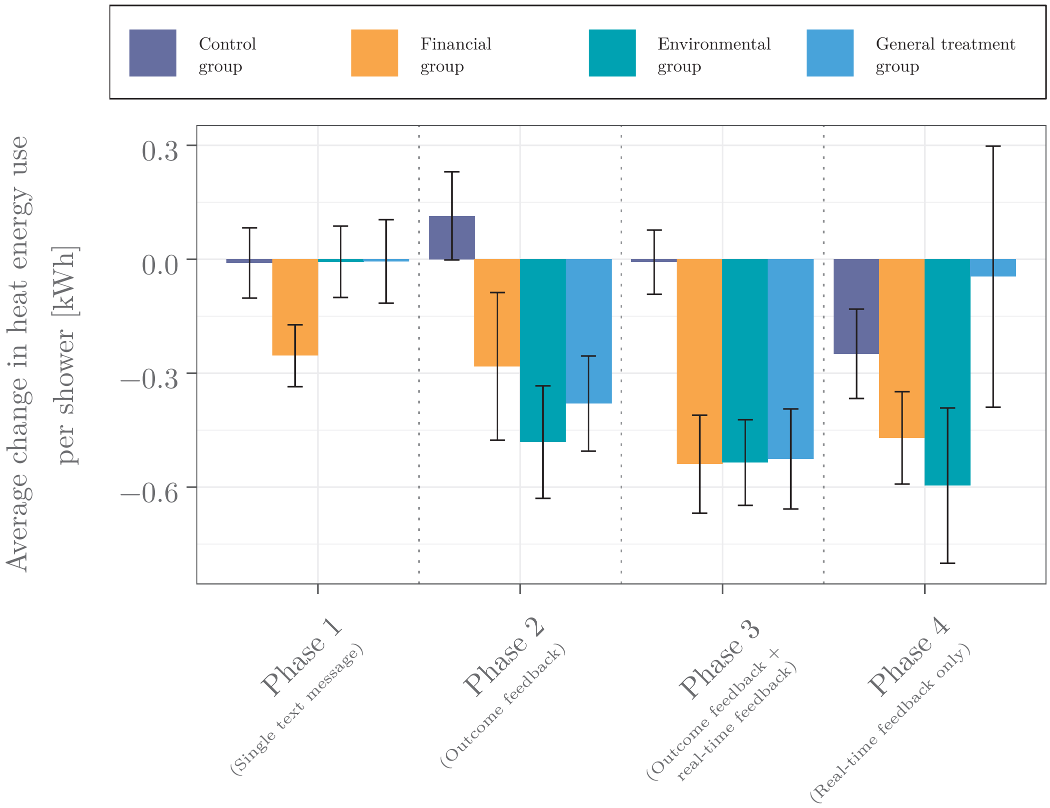


Figure 3: Descriptive evidence

1979) indicates heteroskedasticity across the participants (p value $< .001$) and a Breusch-Godfrey test (Breusch, 1978) indicates serial correlation (p value $< .001$). To account for these potential sources of bias, all the standard errors in the following are clustered based on the White-Arellano method (Arellano, 1987). Lastly, a heteroskedasticity robust regression-based Hausman test (see, e.g., Wooldridge, 2010) indicates that a fixed-effects model should be used over a random-effects model ($X^2(20) = 46.5$; p value $< .001$). Therefore, we choose the fixed-effect model based on the following equation, using ordinary least squares:

$$
\begin{aligned}
y_{it} = \alpha_i &+ IN_{it}^{phase\,1} \times (\beta_0 + \beta_1 T_i^{Financial} + \beta_2 T_i^{Environmental} + \beta_3 T_i^{General}) \\
&+ IN_{it}^{phase\,2} \times (\gamma_0 + \gamma_1 T_i^{Financial} + \gamma_2 T_i^{Environmental} + \gamma_3 T_i^{General}) \\
&+ IN_{it}^{phase\,3a} \times (\delta_0 + \delta_1 T_i^{Financial} + \delta_2 T_i^{Environmental} + \delta_3 T_i^{General}) \\
&+ IN_{it}^{phase\,3b} \times (\varphi_0 + \varphi_1 T_i^{Financial} + \varphi_2 T_i^{Environmental} + \varphi_3 T_i^{General}) \\
&+ IN_{it}^{phase\,4} \times (\omega_0 + \omega_1 T_i^{Financial} + \omega_2 T_i^{Environmental} + \omega_3 T_i^{General}) \\
&+ d_{it} + \varepsilon_{it}
\end{aligned}
\tag{1}
$$

where y_{it} is the energy use of participant i in shower t. We include an individual-fixed effects parameter α_i for each household to eliminate the variance resulting from fixed differences across the participants (e.g., different showerheads leading to different flow rates). The binary variables

$IN_{it}^{phase\,1}$, $IN_{it}^{phase\,2}$, $IN_{it}^{phase\,3a}$, $IN_{it}^{phase\,3b}$, $IN_{it}^{phase\,4}$ are all 0 during the baseline phase and take the value of 1 during the respective intervention phase. $T_i^{Financial}$, $T_i^{Environmental}$, and $T_i^{General}$ are treatment group indicators that take the value of 1 if the participant belongs to the respective group. It is important to note that the phase variables return to the value of 0 after the respective intervention phase. Consequently, all the effects must be interpreted as consumption changes compared to the baseline phase. Since the time stamps are known for 97.9% of the showers (and unknown time stamps have been linearly interpolated), we can also control for time effects (e.g., the outside temperature on the date of shower t). To this end, we include the variable d_{it} in the model, which denotes the chronological order of showers taken by participant i on the date of shower t. Finally, the variable ε_{it} is an error term that captures any unmodelled effects.

Table 3 presents the interventions in the first column and the response of the treatment groups compared to the control group in the second to fourth columns. The standard errors are given in parentheses and, as previously described, they are adjusted for clustering at the participant level. As we test multiple interventions side by side, it might inflate the risk of falsely rejecting the null hypothesis (see, e.g., List et al., 2019). It is important to note that all the statistically significant effects in Table 3 remain statistically significant after adjusting the p values of the associated treatments through the false discovery rate (see Benjamini and Hochberg, 1995). Moreover, we report the corresponding adjusted p values in the following.

Table 3: Main treatment effects on energy use per shower

Independent variable	Financial group	Environmental group	General group	Temporal effect
Single text message	−0.236**	−0.072	−0.008	+0.020
	(0.116)	(0.134)	(0.132)	(0.104)
Outcome feedback	−0.376*	−0.556***	−0.448***	+0.163
	(0.212)	(0.187)	(0.158)	(0.147)
Outcome and real-time feedback (1)	−0.701***	−0.623***	−0.650***	+0.175
	(0.182)	(0.170)	(0.156)	(0.154)
Outcome and real-time feedback (2)	−0.513***	−0.553***	−0.465***	+0.193
	(0.162)	(0.146)	(0.166)	(0.161)
Real-time feedback only	−0.478***	−0.497**	+0.212	+0.054
	(0.162)	(0.214)	(0.367)	(0.186)
Observations	5,098	4,955	5,189	
Clusters	79	78	79	

Notes: The table displays the treatment effects on energy use (in kWh), controlling for individual- and time-fixed effects. Standard errors are presented in parentheses, adjusted for clustering at the participant level. Coefficients were estimated using the within estimator. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Control group observations: 5,588; control group clusters: 90; constant: 2.75; within R^2: 0.005.

With regard to the first intervention, the results indicate that the financial incentives induce energy conservation with an average saving effect of 0.236 kWh per shower (-8.5%, p value $=$ .042, p value$_{adjusted} = .057$). Hence, we find support for hypothesis 1a' and reject hypothesis 1a, which had predicted the opposite based on motivation crowding theory. Interestingly, neither the environmental appeal nor the welcome message had any measurable impact on shower behavior (both have p values greater than .591). We thus reject hypotheses 1b and 1c. In the second intervention phase, outcome consumption feedback that was combined with an additional message for the financial and environmental group, induced statistically significant savings across all the treatment groups. More specifically, the financial group responds with average savings of 0.376 kWh (12.9%, p value $= .076$, p value$_{adjusted} = .095$), the environmental group with average savings of 0.556 kWh (19.1%, p value $= .003$, p value$_{adjusted} = .007$), and the general treatment group with savings of 0.448 kWh (15.4%, p value $= .005$, p value$_{adjusted} =$ .008). An F-test reveals no statistically significant differences in behavioral responses between the different treatment groups ($F(2, 19999) = 0.329$; p value $= .720$), indicating that neither financial incentives nor environmental appeals have any immediate, subtractive, or additive effect on resource conservation when combined with outcome feedback. Hence, we reject hypotheses 2a, 2a', and 2b.

Moving on, the third intervention phase estimates the joint effect of outcome feedback and real-time feedback (with motivational cues for the financial group and the environmental group). In contrast to the first and second intervention phases, the third intervention phase is seven times longer and consequently may be more susceptible to time effects (i.e., behavior changes may diminish over time as the novelty effect of the feedback wears off). To better compare the behavioral response to the different interventions, we have therefore split the third intervention phase into two time segments (see Figure 2). The first segment comprises the same number of observations (7 showers) as the previous intervention phases, and the second segment comprises the remaining observations (42 showers). In the first segment of this phase (i.e., Phase 3a), the treatment groups respond more strongly to the bundled intervention of outcome feedback and real-time feedback than to outcome feedback alone. More specifically, the financial group responds with average savings of 0.701 kWh (24.0%, p value $< .001$, p value$_{adjusted} < .001$), the environmental group with average savings of 0.623 kWh (21.3%, p value $< .001$, p value$_{adjusted} < .001$), and the general treatment group with average savings of 0.650 kWh (22.2%, p value $< .001$, p value$_{adjusted} < .001$). We find no evidence that participants respond differently among the groups ($F(2, 19999) = 0.081$; p value $= .922$). However, a separate analysis in which all the treatment conditions of this phase are collapsed (see Table A1) suggests that real-time feedback enhances the conservation effects by additional 6.7% (p value $= .092$). We therefore find support for hypothesis 3. In the second time segment of this phase (i.e., Phase 3b), the average savings decrease to a level of 0.513 kWh for the financial group (17.4%, p value $= .002$, p value$_{adjusted} = .005$), to 0.553 kWh for the environmental group (18.8%, p value $< .001$, p value$_{adjusted} < .001$), and to 0.465 kWh for the general treatment group (15.8%, p value = .005, p value$_{adjusted}$ = .008). Again, the behavioral responses between the treatment groups appear not to differ ($F(2, 19999) = 0.129$; p value = .879). As a next step, we test whether consumption feedback is dominant over motivational cues. To this end, we construct a linear hypothesis test on the corresponding model coefficients (H_0: $\gamma_1 = \gamma_3 \wedge \gamma_2 = \gamma_3 \wedge \delta_1 = \delta_3 \wedge \delta_2 = \delta_3 \wedge \varphi_1 = \varphi_3 \wedge \varphi_2 = \varphi_3$). The test fails to reject the null hypothesis that the conservation effects are different between the general treatment group and the other treatment groups in the phases 2 to 3b ($F(6, 19999) = 0.422$; p value = .865). We hence find support for hypothesis 4 which predicted that feedback is the dominant factor inducing behavior change and that motivational cues do not increase conservation effects.

When text messages with outcome feedback were no longer sent to the participants (last intervention phase), the financial group and the environmental group still respond to real-time feedback with average savings of 0.478 kWh (17.0%, p value = .003, p value$_{adjusted}$ = .007) and 0.497 kWh (17.7%, p value = .020, p value$_{adjusted}$ = .030) per shower, respectively. In contrast, the behavioral response of the general treatment group is no longer statistically significantly different from zero (p value = .564, p value$_{adjusted}$ = .634). We thus reject hypothesis 5, which had predicted a strong undermining effect (i.e., no savings) after the withdrawal of the financial incentives. When we focus on the treatment differences in the last phase, the financial group saves as much as the environmental group (p value = .936). Interestingly, it appears that the monetary context is better for sustaining pro-environmental behavior than providing no motivational cue, as the difference to the general treatment group, which no longer conserves heat energy, is marginally statistically significant (p value = .066). We thus find support for hypotheses 5' and 5'', which had predicted that the treatment effects are not (negatively) affected by a preceding period of incentives, providing evidence against motivation crowding theory.

We further explore whether participants respond to these interventions by, e.g., replacing long showers with several short showers, thereby increasing their shower frequency. To this end, we estimated the treatment effects on the shower frequencies (i.e., showers per day). The results clearly indicate that the interventions had no effect on the average shower frequencies of the treatment groups (H_0: No impact of the treatments: $F(15, 325) = 0.685$; p value = .800; treatment coefficients reported in Table 4); instead, an additional analysis indicates that the savings were achieved primarily through shorter showers (treatment effects on shower time reported in Table A2).

To analyze whether participants' intrinsic task motivation has changed in response to the interventions, we use the five-point Likert scale question "I act environmentally responsible, even if this is associated with much higher costs and efforts." A corresponding Kruskal–Wallis test fails to reject the null hypothesis that there was no change in attitude between the baseline and the intervention phases for the treatment groups (N = 228 participants; $H_{Env.attitude}(3)$ = 2.44; p value = .486). Overall, the findings substantially support our previous observations that the incentives had no undermining effect.

Table 4: Treatment effects on the shower frequency of the participants in the respective intervention phases

Independent variable	Financial group	Environmental group	General group
Single text message	−0.137	−0.155	−0.192*
	(0.123)	(0.126)	(0.116)
Outcome feedback	−0.008	−0.061	−0.028
	(0.097)	(0.091)	(0.085)
Outcome and real-time feedback (1)	+0.055	+0.031	−0.005
	(0.143)	(0.124)	(0.111)
Outcome and real-time feedback (2)	−0.114*	−0.050	−0.057
	(0.063)	(0.068)	(0.061)
Real-time feedback only	+0.076	+0.115	+0.095
	(0.117)	(0.127)	(0.111)
Observations	427	425	425
Clusters	79	78	79

Notes: The table displays the treatment effects on shower frequency (showers per day), controlling for individual- and time-fixed effects. H_0: No impact of the treatments: F-test: p value = .800. Coefficients were estimated using the within estimator. Standard errors are presented in parentheses, adjusted for clustering at the participant level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Control group observations: 483; control group clusters: 90; constant: 1.04; within R^2: 0.007.

Subgroup analyses

To derive practical implications for behavioral programs, it is important to further analyze whether the behavioral response of the individuals differ among certain subgroups. For instance, it may be that the financial incentives undermine the motivation of individuals with strong environmental attitudes, leading to higher consumption levels, while inducing less environmentally friendly individuals to engage in conservation behavior. Similarly, it may be that the trade-off among the economic effects of the financial incentives and its controlling nature depends on the wealth of participants: Specifically, it could lead to desirable effects for less wealthy individuals while resulting in adverse effect for more affluent individuals. To shed light on such mechanisms, we run three additional analyses. In this context, we no longer differentiate between the two intervention phases evaluating different intensities of consumption feedback (i.e., Phases 2 to 3b), as such a differentiation provides no additional insights. We do, however, still take potential behavioral differences after the withdrawal of the financial incentives into account.

Environmental attitudes and conservation behavior

To analyze whether participants' environmental attitudes translate (differently) into conservation behavior across the treatment groups, we integrate a variable from the first questionnaire into the regression model. Specifically, as a proxy of environmental attitude, we use again the five-point Likert scale question "I act environmentally responsible, even if this is associated with much higher costs and efforts" as a binary interaction term z_i (a value of 1 indicates an environmental attitude higher than the median, the default value is 0). Moreover, we control for linear interactions with the baseline energy use by including the interaction term b_i in the model (i.e, as mean centered variable). Mathematically, we estimate the following equation:

$$
\begin{aligned}
y_{it} = \alpha_i + IN_{it}^{phase\,1} \times & \begin{pmatrix} \beta_0 + \beta_1 T_i^{Financssial} + \beta_2 T_i^{Environmental} + \beta_3 T_i^{General}) \\ + z_i \times (\beta_4 T_i^{Financial} + \beta_5 T_i^{Environmental} + \beta_6 T_i^{General}) \\ + b_i \times (\beta_7 T_i^{Financial} + \beta_8 T_i^{Environmental} + \beta_9 T_i^{General}) \end{pmatrix} \\
+ IN_{it}^{phases\,2-3b} \times & \begin{pmatrix} \gamma_0 + \gamma_1 T_i^{Financial} + \gamma_2 T_i^{Environmental} + \gamma_3 T_i^{General}) \\ + z_i \times (\gamma_4 T_i^{Financial} + \gamma_5 T_i^{Environmental} + \gamma_6 T_i^{General}) \\ + b_i \times (\gamma_7 T_i^{Financial} + \gamma_8 T_i^{Environmental} + \gamma_9 T_i^{General}) \end{pmatrix} \\
+ IN_{it}^{phase\,4} \times & \begin{pmatrix} \delta_0 + \delta_1 T_i^{Financial} + \delta_2 T_i^{Environmental} + \delta_3 T_i^{General}) \\ + z_i \times (\delta_4 T_i^{Financial} + \delta_5 T_i^{Environmental} + \delta_6 T_i^{General}) \\ + b_i \times (\delta_7 T_i^{Financial} + \delta_8 T_i^{Environmental} + \delta_9 T_i^{General}) \end{pmatrix} \\
+ d_{it} + \varepsilon_{it} &
\end{aligned} \tag{2}
$$

Table 5 reports the corresponding treatment effects. When we look at the interaction coefficients between the interventions and the environmental attitude, we do not find that these point estimates are precisely estimated. At the same time, the analysis nevertheless documents an important finding. Controlling for the baseline consumption and the environmental attitude, the previously observed main effects are robust, even when adjusting the *p* values of the associated main treatment effects through the false discovery rate (see Benjamini and Hochberg, 1995). When we repeat this analysis by integrating the environmental attitude variable as a linear interaction term into the corresponding equation (i.e, as mean centered variable), the results are quite similar. In other words, the interaction variables are far away from statistical significance levels, while the main effects are still relatively precisely estimated.* Overall, we thus reject hypothesis 6a and 6a', which predicted heterogeneous effects depending on moral values.

Financial status and conservation behavior

As the last part of our subgroup analysis, we conduct two subgroup analyses to investigate whether less wealthy and more affluent individuals have different behavioral responses to

* For transparency reasons, Table A3 reports the main treatment effects and the coefficients of their interaction with the baseline energy use. The baseline energy use is centered, and therefore it represents the treatment effects of deviating from the mean baseline energy use of the sample.

Table 5: Treatment effects on energy use per shower in conjunction with the proxy for intrinsic motivation

Independent variable	Financial group	Environmental group	General group	Temporal effect
Single text message	−0.229*	−0.105	+0.062	+0.002
	(0.119)	(0.165)	(0.140)	(0.103)
Single text message × Env. attitude	+0.002	+0.094	−0.124	–
	(0.159)	(0.196)	(0.198)	
Feedback$^{phases\,2\text{–}3b}$	−0.464***	−0.586***	−0.480***	+0.143
	(0.179)	(0.154)	(0.136)	(0.122)
Feedback$^{phases\,2\text{–}3b}$ × Env. attitude	−0.134	+0.050	+0.088	–
	(0.215)	(0.190)	(0.248)	
Feedback$^{phase\,4}$	−0.465***	−0.432**	−0.210	+0.003
	(0.166)	(0.216)	(0.182)	(0.145)
Feedback$^{phase\,4}$ × Env. attitude	−0.076	−0.130	+0.842	–
	(0.260)	(0.300)	(0.738)	
Observations	5,011	4,955	5,189	
Clusters	78	78	79	

Notes: The table displays the treatment effects on energy use per shower (in kWh), controlling for baseline energy use (coefficients not reported), individual effects, and time-fixed effects. Standard errors are presented in parentheses, adjusted for clustering at the participant level. Coefficients were estimated using the within estimator. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Please note that two participants did not answer the corresponding survey question and, thus, were excluded for this analysis. Control group observations: 5,525; control group clusters: 89; constant: 2.80; within R^2: 0.019.

the financial incentives. To this end, we use questionnaire data about participants' financial situations (from the second questionnaire). Before reporting potential differences in their behavioral responses, it is important to mention that the questionnaire indicates that our sample is not representative: In comparing our study's participants with a large German social survey of students from 2016 (see Apolinarski et al., 2018; Middendorff et al., 2017), individuals in our sample are more likely to have a job (present study = 48.0%, social survey = 17.2%) and to have less money left at the end of each month (Median_{Study} = €51 to €100, $\text{Median}_{SocialSurvey}$ = €101 to €200) than individuals in the social survey with similar living conditions (e.g., living in single apartments). This might suggest that the impact of the incentives could be diminished when extrapolating our findings to the (more affluent) wider student population.

To check whether there are heterogeneous treatment effects, we examine whether individuals respond differently to the incentives based on their remaining budgets at the end of a month. For this analysis, we consider all the participants who have answered the corresponding question and belong either to the control group or to the financial group. We first investigate whether the key characteristics are equal among these subsamples to check for potential biases. A chi-squared test reveals no statistically significant effect of the group membership on the budget left at the end of the month ($X^2 = 0.6$; p value = .966). We subsequently split both the control group and the financial group into two subgroups based on their remaining monthly budget (median split: less than €51 left; more or equal to €51 left). This allows us to estimate the behavior change of the financial subgroups as compared to the respective control subgroup with the same budget level (e.g., Financial$_{<€51}$ vs. Control$_{<€51}$)*. Next, we estimate the behavioral effects as follows:

$$\begin{aligned} y_{it} = {} & \alpha_i + IN_{it}^{phase\,1} \times (\beta_0 + \beta_1 T_i^{Control_{\geq €51}} + \beta_2 T_i^{Financial_{<€51}} + \beta_3 T_i^{Financial_{\geq €51}}) \\ & + IN_{it}^{phases\,2-3b} \times (\gamma_0 + \gamma_1 T_i^{Control_{\geq €51}} + \gamma_2 T_i^{Financial_{<€51}} + \gamma_3 T_i^{Financial_{\geq €51}}) \\ & + IN_{it}^{phase\,4} \times (\delta_0 + \delta_1 T_i^{Control_{\geq €51}} + \delta_2 T_i^{Financial_{<€51}} + \delta_3 T_i^{Financial_{\geq €51}}) \\ & + d_{it} + \varepsilon_{it} \end{aligned} \tag{3}$$

where $T_i^{Control \geq €51}$, $T_i^{Financial < €51}$, and $T_i^{Financial \geq €51}$ are binary indicators that take the value of 1 if the participant belongs to the respective subgroup. Table 6 reports the impact of the behavioral interventions on the financial subgroups in relation to their control group. Here, column (1) displays the behavioral change of the financial subgroup$_{<€51}$ in comparison to the control subgroup$_{<€51}$, and column (2) displays the behavioral change for the subgroups with a remaining budget of equal to or higher than €51. The point estimates suggest that the impact of the financial incentives, in the presence and absence of feedback, is different across both budget levels. To formally investigate whether the behavioral change is different for both financial subgroups in comparison with the respective control group, we construct a linear hypothesis test on the corresponding model coefficients (H_0: $\beta_2 = \beta_3 - \beta_1 \wedge \gamma_2 = \gamma_3 - \gamma_1$). The test fails to reject the null hypothesis that the financial subgroups respond equally to the financial incentives ($F(2, 4977) = 1.41$; p value = .245). In other words, the results indicate that less wealthy individuals respond to the financial incentives to the same extent as relatively affluent individuals. When we repeat this analysis on a second question, namely how often the participants are short of cash at the end of the month, we do not find different subgroup-specific effects ($F(2, 5634) = 2.26$; p value = .104). In view of both analyses, we thus reject hypothesis 6b, which predicted different behavioral responses depending on an individual's financial situation.

To summarize, the analyses suggest that (i) financial incentives do not undermine the intrinsic motivation of individuals with strong environmental attitudes, (ii) the financial incentives have the same effect regardless of the financial situation of the participants, and (iii) the appeals have the same effect regardless of participants' environmental attitudes.

---

* Randomization checks indicate that the subgroups are equal in key characteristics.

Table 6: Subgroup analysis: Median split of the control group and the financial group by the amount of the remaining budget at the end of a month

Independent variable	Financial$_{<€51}$ (1)	Financial$_{\geq €51}$ (2)
Single message	−0.092	−0.467**
	(0.261)	(0.200)
Feedback$^{phases\,2-3b}$	−0.464*	−0.805**
	(0.244)	(0.329)
Feedback$^{phase\,4}$	−0.333	−0.994***
	(0.311)	(0.300)
Treatment group observations	1,305	1,573
Treatment group clusters	19	24

Notes: The table displays the treatment effects on energy use per shower (in kWh) per subgroup, controlling for individual- and time-fixed effects. The temporal effects are not reported. Standard errors are presented in parentheses, adjusted for clustering at the participant level. Coefficients were estimated using the within estimator. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Control group clusters$_{<€51}$: 22; control group clusters$_{\geq €51}$: 23; constant: 2.73; within R^2: 0.008.

Discussion and conclusions

Implications

A key objective of this paper was to empirically evaluate the behavioral response to financial incentives, environmental appeals, and consumption feedback side by side. To this end, we have conducted a randomized controlled field trial in the environmental domain that targeted a resource-intensive everyday activity (showering). In this setting, we have tested hypotheses derived from theories with conflicting predictions to advance our theoretical understanding. Based on the results, we derive five main implications that are highly relevant for both theory and practice.

First, our study provides empirical evidence that financial incentives can encourage considerable resource conservation as appeals aimed at activating recipients' moral values had no measurable effect on the participants' target behavior. While the latter seems surprising to some extent, these results are consistent with observations from macro- and microeconomic studies. Specifically, political campaigns and studies that have used appeals to promote pro-environmental behavior report modest changes over time at best (see, e.g., Dolnicar et al., 2017; Luyben, 1982). With this in mind, our study design deliberately made the appeals as salient and context-specific as we could. Nevertheless, the results suggest that such appeals on their own are insufficient to induce substantial resource savings. Small financial incentives, on the other hand, are clearly effective.

Second, our study investigated the impact of feedback (i) when additional cues for pro-environmental behavior are given (i.e., financial incentives, appeals) and (ii) when they are later withdrawn. When such additional cues are in place, feedback appears to be the dominant component that facilitates behavior change. In particular, financial incentives neither unleash additional resource savings nor do they dampen savings as compared to the treatment groups that received feedback but no financial incentives, providing some evidence against predictions of both standard economic theory and motivation crowding theory. To the best of our knowledge, this is the first experiment in the environmental domain that delivers evidence on the non-additivity of feedback information and additional cues, a theoretical derivation of feedback intervention theory (Kluger and DeNisi, 1996).

Third, our study generates additional important evidence against motivation crowding theory in the environmental domain by considering crucial confounds that could otherwise mask adverse effects (e.g., intrinsic motivation at baseline, financial status). Our results from the subgroup analyses refute the undermining effects of financial incentives. More specifically, neither differences in intrinsic motivation (our proxy variable was the environmental attitude) nor differences in individuals' financial standing substantially moderated the effects of the financial incentives. In doing so, our study corroborates the findings of related research which did not find evidence of adverse effects in their heterogeneity analyses (see, e.g., Mizobuchi and Takeuchi, 2013; Silvi and Rosa, 2017).

Fourth, our study points to a general limitation of environmental programs, as a higher intensity of consumption feedback (i.e., the combined provision of real-time feedback and outcome feedback) did not produce substantially higher conservation effects. In view of this, our findings are consistent with resource allocation theory (Kanfer and Ackerman, 1989): Resource conservation is a task that does not require substantial learning, as people are more and more aware of the environmental degradation. Therefore, appeals might have little conservation effects. Instead, it requires interventions that personally affect individuals, such as by promising financial benefits or providing feedback on personal conservation performance, and thereby direct their attention to the target activity. After an initial allocation of attention from individuals, further interventions on the same target face diminishing returns. This rationale explains why financial incentives alone were effective but, when combined with consumption feedback, did not enhance conservation effects. Clearly, for substantially higher financial incentives than in our study, it is still plausible that financial incentives foster higher resource savings. More specifically, the additional utility of resource conservation could become at some point so high that individuals pay more attention to reaping the financial incentives. Likewise, an alternative explanation to the dominance of feedback is that higher financial incentives may be needed to compensate individuals for the inherent disutility of further reducing consumption in the shower, which might make some aspects inconvenient (e.g., to thoroughly rinse out the shampoo of the hairs). Yet, as the focus of our study is more oriented toward testing the effects of economically feasible financial incentives (i.e., incentives as high as the monetary savings from resource conservation) in the absence and presence of feedback, we cannot say at what level of financial incentives this will be the case. Nevertheless, our study additionally demonstrates that motivational cues, such as financial incentives or environmental appeals, might be important for sustaining conservation efforts over time. Both motivational cues had a sustained positive post-treatment effect, as the treatment group that had previously received no motivational cue (i.e., feedback, but no financial incentive/no appeal) stopped conserving resources during this timeframe with only real-time feedback. While this is somewhat surprising, it is conceivable that the motivational cues led individuals to more fully internalize the relevance of the target behavior (showering), in contrast to those who did not receive such cues. This potential mechanism is also corroborated by research on habit formation, which emphasizes that changes in beliefs and interpretations may be a crucial pathway to creating enduring behavioral effects (Frey and Rogers, 2014).

Fifth, our study provides important insights into the effectiveness of environmental campaigns at a more general level. While we cannot refute environmental appeals to have some small effect on resource conservation, it is important to note that interventions require large effects to meaningfully address pressing problems such as climate change. Clearly, achieving a cost-effective conservation effect on a resource-intensive activity is valuable, however, every little does not help. Climate change is a pressing issue, with very limited time left to achieve the goals of the Paris Agreement, requiring substantial changes also from the behavioral side (Höhne et al., 2020; Rogelj et al., 2022). The good news, however, is that financial incentives may not be necessary to trigger behavior change in everyday contexts. In particular, the provision of outcome feedback induced conservation effects that were on average more than 89% higher than mere financial incentives. As digital technologies can capture more and more environmental activities in a scalable way, our study highlights the potential of feedback interventions for behavior change.

Limitations and suggestions for future research

Despite our best efforts, this study is subject to several limitations. First of all, like the vast majority of studies, we studied behavioral responses for one target behavior (i.e., showering) from one sample, which provides limited evidence how the observed effects translate to other scenarios (e.g., different target behaviors, different samples; see Andor et al., 2022; Gravert and Collentine, 2021). Yet, unlike prior studies, we provided a more exhaustive overview of the effects and the interplay of interventions. We still believe that our findings on behavioral mechanisms might generalize to other scenarios as they are consistent with those of related studies (i.e., regarding the effects of small financial incentives, environmental appeals, and consumption feedback). However, it is conceivable that the effects are quite different when the tested interventions feature other conditions or elements. For instance, in cases in which one's financial rewards for pro-environmental actions will be visible to others or when consumption feedback is combined with descriptive norms, additional behavioral mechanisms might come into play (see, e.g., Handgraaf et al., 2013; Sudarshan, 2017). Second, despite the potential of behavioral interventions in changing everyday behaviors, it should be kept in mind that there might be a notable limit on their overall effects. More specifically, when individuals receive behavioral interventions on multiple activities in parallel, they may become less sensitive to such interventions (Tiefenbeck et al., 2018), which could reduce the overall impact of behavioral interventions. To avoid this, an approach could be to focus on specific policy-relevant activities first. Once a desired habit regarding the target behavior has formed (e.g., resource conservation), it might be valuable to start targeting the next activity for behavior change. We, however, cannot provide substantial evidence on habit formation, as our experiment lacks a post-treatment phase (i.e., a phase without real-time feedback). Still, for the same target behavior, Byrne et al. (2023) provide important insights into habit formation by studying the behavioral responses to alternating periods with and without real-time feedback.

Third, as with every study that empirically evaluates the impact of incentive schemes, the behavioral effects might depend considerably on the frequency of the payout (e.g., monthly, bimonthly; McClure et al., 2004), the amount of compensation (Gneezy and Rustichini, 2000; Pokorny, 2008), and the incentive type (e.g., contingent vs. non-contingent rewards; Deci et al., 1999). With this in mind, we deliberately designed the incentive scheme to strongly undermine motivation (i.e., performance-contingent rewards that were announced in advance; Deci et al., 1999) while simultaneously considering typical constraints for its scalable application in the real world (e.g., financial rewards as high as the monetary savings would be from resource conservation). Yet, we may still underestimate the undermining effect of financial incentives for two reasons: (i) the duration of the experiment and (ii) the relatively late remuneration may have made the financial incentives and their withdrawal less salient and explicit than in related studies from the field of psychology (see, e.g., Esteves-Sorenson and Broce, 2022). At the same time, however, the timeframe used in our experiment is probably more realistic for the practical implementation in other real-world applications. Fourth, even though 51 (64.6%) of the participants in the financial group visited the web page that explained how the remuneration would be paid out, only 7 of them (13.7%) provided us with their bank details that were necessary to receive the payments. Two explanations for this could be that (i) potential recipients simply forgot to register their bank details, or (ii) were not interested enough in reaping the financial incentives. In light of this, the financial incentives may have operated largely as an attentional effect, as discussed by Schwartz et al. (2019), which we cannot demonstrate. These incentives might have given individuals a tangible reason to engage in conservation, rather than functioning as an economic incentive in itself.

In addition to addressing these limitations, our study highlights various avenues for future research. For instance, while our focus was on environmental behavior, future research could identify whether our observation on the non-additivity of feedback and financial incentives in promoting behavior change can be generalized to other domains (e.g., to promote health behavior). Additionally, future work could also investigate the interventions over even longer periods of time in order to uncover long-term effects. In a similar vein, it is conceivable that at some point in time, new, stable habits form so that financial incentives and/or feedback can be discontinued while the efforts for desirable behavior remain. Moreover, apart from the direct behavioral effects, there is evidence in the environmental domain that campaigns that appeal to moral values could lead to more positive side effects (i.e., pro-environmental behavior in other activities) than campaigns that appeal to self-interests (Steinhorst et al., 2015). Future research could thus extend our study design by capturing such potential side effects, thereby shedding more light on the overall behavioral response to these interventions.

Acknowledgments

This work has been funded in part by the Bavarian State Ministry of Education, Science and the Arts within the framework "TechnologieAllianzOberfranken (TAO)."

References

Allcott, H. (2011). "Social Norms and Energy Conservation." *Journal of Public Economics* 95 (9-10), pp. 1082–1095. DOI: 10.1016/j.jpubeco.2011.03.003.

Andor, M. A., Gerster, A., and Peters, J. (2022). "Information Campaigns for Residential Energy Conservation." *European Economic Review* 144, Article nr. 104094. DOI: 10.1016/j.euroecorev.2022.104094.

Apesteguia, J., Funk, P., and Iriberri, N. (2013). "Promoting Rule Compliance in Daily-Life: Evidence from a Randomized Field Experiment in the Public Libraries of Barcelona." *European Economic Review* 64, pp. 266–284. DOI: 10.1016/j.euroecorev.2013.08.010.

Apolinarski, B., Becker, K., Bornkessel, P., Brandt, T., Heißenberg, S., Middendorff, E., Naumann, H., and Poskowsky, J. (2018). *21st Social Survey (2016). Data Collection: 2016. Version: 1.0.0. Data Package Access Way: SUF: Download.* Edited by F. Baillet and A. Weber. Hannover, Germany: FDZ-DZHW. DOI: 10.21249/DZHW:ssy21:1.0.0.

Arellano, M. (1987). "Practitioners' Corner: Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics* 49 (4), pp. 431–434. DOI: 10.1111/j.1468-0084.1987.mp49004006.x.

Ariely, D., Bracha, A., and Meier, S. (2009). "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review* 99 (1), pp. 544–555. DOI: 10.1257/aer.99.1.544.

Asensio, O. I. and Delmas, M. A. (2015). "Nonprice Incentives and Energy Conservation." *Proceedings of the National Academy of Sciences* 112 (6), E510–E515. DOI: 10.1073/pnas.1401880112.

Azarova, V., Cohen, J. J., Kollmann, A., and Reichl, J. (2020). "Reducing Household Electricity Consumption during Evening Peak Demand Times: Evidence from a Field Experiment." *Energy Policy* 144, Article nr. 111657. DOI: 10.1016/j.enpol.2020.111657.

Bénabou, R. and Tirole, J. (2006). "Incentives and Prosocial Behavior." *The American Economic Review* 96 (5), pp. 1652–1678. DOI: 10.1257/aer.96.5.1652.

Benjamini, Y. and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1), pp. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Berger, S., Bärtsch, C., Schmidt, C., Christandl, F., and Wyss, A. M. (2018). "When Utilitarian Claims Backfire: Advertising Content and the Uptake of Insects as Food." *Frontiers in Nutrition* 5, Article nr. 88. DOI: 10.3389/fnut.2018.00088.

Bertrand, A., Mastrucci, A., Schüler, N., Aggoune, R., and Maréchal, F. (2017). "Characterisation of Domestic Hot Water End-Uses for Integrated Urban Thermal Energy Assessment and Optimisation." *Applied Energy* 186 (Part 2), pp. 152–166. DOI: 10.1016/j.apenergy.2016.02.107.

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2020). "You've Got Mail: A Randomized Field Experiment on Tax Evasion." *Management Science* 66 (7), pp. 2801–2819. DOI: 10.1287/mnsc.2019.3390.

Breusch, T. S. (1978). "Testing for Autocorrelation in Dynamic Linear Models." *Australian Economic Papers* 17 (31), pp. 334–355. DOI: 10.1111/j.1467-8454.1978.tb00635.x.

Breusch, T. S. and Pagan, A. R. (1979). "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* 47 (5), pp. 1287–1294. DOI: 10.2307/1911963.

Brülisauer, M., Goette, L., Jiang, Z., Schmitz, J., and Schubert, R. (2020). "Appliance-Specific Feedback and Social Comparisons: Evidence from a Field Experiment on Energy Conservation." *Energy Policy* 145, Article nr. 111742. DOI: 10.1016/j.enpol.2020.111742.

Burton-Jones, A., Butler, B. S., Scott, S. V., and Xu, S. X. (2021). "Next-Generation Information Systems Theorizing: A Call to Action." *MIS Quarterly* 45 (1), pp. 301–314. DOI: 10.25300/MISQ/2021/15434.

Byrne, D. P., Goette, L., Martin, L. A., Jones, A., Miles, A., Schob, S., Staake, T., and Tiefenbeck, V. (2023). "How Nudges Create Habits: Theory and Evidence from a Field Experiment." Working Paper. DOI: 10.2139/ssrn.3974371.

Chadi, A., Mechtel, M., and Mertins, V. (2022). "Smartphone Bans and Workplace Performance." *Experimental Economics* 25 (1), pp. 287–317. DOI: 10.1007/s10683-021-09715-w.

Charness, G. and Gneezy, U. (2009). "Incentives to Exercise." *Econometrica* 77 (3), pp. 909–931. DOI: 10.3982/ECTA7416.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Hillsdale, NJ, USA: Lawrence Erlbaum Associates. ISBN: 978-0-203-77158-7. DOI: 10.4324/9780203771587.

Deci, E. L. (1971). "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology* 18 (1), pp. 105–115. DOI: 10.1037/h0030644.

Deci, E. L. (1972). "Intrinsic Motivation, Extrinsic Reinforcement, and Inequity." *Journal of Personality and Social Psychology* 22 (1), pp. 113–120. DOI: 10.1037/h0032355.

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin* 125 (6), pp. 627–668. DOI: 10.1037/0033-2909.125.6.627.

De Sousa, C. O. M. and Dias Fouto, N. M. M. (2019). "Economic Incentives for Water Consumption Reduction: Case Study of the City of São Paulo, Brazil." *Water Policy* 21 (6), pp. 1266–1278. DOI: 10.2166/wp.2019.022.

Dolnicar, S., Knezevic Cvelbar, L., and Grün, B. (2017). "Do Pro-Environmental Appeals Trigger Pro-Environmental Behavior in Hotel Guests?" *Journal of Travel Research* 56 (8), pp. 988–997. DOI: 10.1177/0047287516678089.

Esteves-Sorenson, C. and Broce, R. (2022). "Do Monetary Incentives Undermine Performance on Intrinsically Enjoyable Tasks? A Field Test." *The Review of Economics and Statistics* 104 (1), pp. 67–84. DOI: 10.1162/rest_a_00947.

European Statistical Office (2023). *Energy Consumption in Households*. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households#Energy_products_used_in_the_residential_sector (visited on 09/15/2023).

Fang, X., Goette, L., Rockenbach, B., Sutter, M., Tiefenbeck, V., Schoeb, S., and Staake, T. (2023). "Complementarities in Behavioral Interventions: Evidence from a Field Experiment on Energy Conservation." *Journal of Public Economics* 228, Article nr. 105028. DOI: 10.1016/j.jpubeco.2023.105028.

Frey, E. and Rogers, T. (2014). "Persistence: How Treatment Effects Persist After Interventions Stop." *Policy Insights from the Behavioral and Brain Sciences* 1 (1), pp. 172–179. DOI: 10.1177/2372732214550405.

Geller, E. S. (1981). "Evaluating Energy Conservation Programs: Is Verbal Report Enough?" *Journal of Consumer Research* 8 (3), pp. 331–335. DOI: 10.1086/208872.

Gerster, A., Andor, M. A., and Goette, L. (2020). "Disaggregate Consumption Feedback and Energy Conservation." Working Paper. URL: https://ssrn.com/abstract=3638059.

Gneezy, U. and Rustichini, A. (2000). "Pay Enough or Don't Pay at All." *The Quarterly Journal of Economics* 115 (3), pp. 791–810. DOI: 10.1162/003355300554917.

Gravert, C. and Collentine, L. O. (2021). "When Nudges Aren't Enough: Norms, Incentives and Habit Formation in Public Transport Usage." *Journal of Economic Behavior & Organization* 190, pp. 1–14. DOI: 10.1016/j.jebo.2021.07.012.

Grieder, M., Bärenbold, R., Schmitz, J., and Schubert, R. (2021). "The Behavioral Effects of Carbon Taxes – Experimental Evidence." Working Paper. DOI: 10.2139/ssrn.3628516.

Günther, M., Kacperski, C., and Krems, J. F. (2020). "Can Electric Vehicle Drivers Be Persuaded to Eco-Drive? A Field Study of Feedback, Gamification and Financial Rewards in

Germany." *Energy Research & Social Science* 63, Article nr. 101407. DOI: 10.1016/j.erss.2019.101407.

Handgraaf, M. J., Van Lidth de Jeude, M. A., and Appelt, K. C. (2013). "Public Praise vs. Private Pay: Effects of Rewards on Energy Conservation in the Workplace." *Ecological Economics* 86, pp. 86–92. DOI: 10.1016/j.ecolecon.2012.11.008.

Heyman, J. and Ariely, D. (2004). "Effort for Payment: A Tale of Two Markets." *Psychological Science* 15 (11), pp. 787–793. DOI: 10.1111/j.0956-7976.2004.00757.x.

Hoffmann, C. and Thommes, K. (2020). "Can Digital Feedback Increase Employee Performance and Energy Efficiency in Firms? Evidence from a Field Experiment." *Journal of Economic Behavior & Organization* 180, pp. 49–65. DOI: 10.1016/j.jebo.2020.09.034.

Höhne, N., den Elzen, M., Rogelj, J., Metz, B., Fransen, T., Kuramochi, T., Olhoff, A., Alcamo, J., Winkler, H., Fu, S., Schaeffer, M., Schaeffer, R., Peters, G. P., Maxwell, S., and Dubash, N. K. (2020). "Emissions: World Has Four Times the Work or One-Third of the Time." *Nature* 579 (7797), pp. 25–28. DOI: 10.1038/d41586-020-00571-x.

Kanfer, R. and Ackerman, P. L. (1989). "Motivation and Cognitive Abilities: An Integrative/Aptitude-Treatment Interaction Approach to Skill Acquisition." *Journal of Applied Psychology* 74 (4), pp. 657–690. DOI: 10.1037/0021-9010.74.4.657.

Kluger, A. N. and DeNisi, A. (1996). "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory." *Psychological Bulletin* 119 (2), pp. 254–284. DOI: 10.1037/0033-2909.119.2.254.

Kramer, J.-N., Tinschert, P., Scholz, U., Fleisch, E., and Kowatsch, T. (2019). "A Cluster-Randomized Trial on Small Incentives to Promote Physical Activity." *American Journal of Preventive Medicine* 56 (2), e45–e54. DOI: 10.1016/j.amepre.2018.09.018.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). "Multiple Hypothesis Testing in Experimental Economics." *Experimental Economics* 22 (4), pp. 773–793. DOI: 10.1007/s10683-018-09597-5.

Luyben, P. D. (1982). "Prompting Thermostat Setting Behavior: Public Response to a Presidential Appeal for Conservation." *Environment and Behavior* 14 (1), pp. 113–128. DOI: 10.1177/0013916582141007.

McClure, S. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2004). "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." *Science* 306 (5695), pp. 503–507. DOI: 10.1126/science.1100907.

Middendorff, E., Apolinarski, B., Becker, K., Bornkessel, P., Brandt, T., Heißenberg, S., and Poskowsky, J. (2017). *Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2016: 21. Sozialerhebung des Deutschen Studentenwerks*. Berlin, Germany: Bundesministerium für Bildung und Forschung. URL: https://www.dzhw.eu/pdf/sozialerhebung/21/Soz21_hauptbericht_barrierefrei.pdf.

Mizobuchi, K. and Takeuchi, K. (2013). "The Influences of Financial and Non-Financial Factors on Energy-Saving Behaviour: A Field Experiment in Japan." *Energy Policy* 63, pp. 775–787. DOI: 10.1016/j.enpol.2013.08.064.

O'Keefe, D. J. and Jensen, J. D. (2007). "The Relative Persuasiveness of Gain-Framed Loss-Framed Messages for Encouraging Disease Prevention Behaviors: A Meta-Analytic Review." *Journal of Health Communication* 12 (7), pp. 623–644. DOI: 10.1080/10810730701615198.

Odyssee-Mure (2021). *Energy Efficiency Trends for Households in the EU – Sectoral Profile - Households*. URL: https://www.odyssee-mure.eu/publications/efficiency-by-sector/household/household-eu.pdf (visited on 07/19/2023).

Pokorny, K. (2008). "Pay—but Do Not Pay Too Much: An Experimental Study on the Impact of Incentives." *Journal of Economic Behavior & Organization* 66 (2), pp. 251–264. DOI: 10.1016/j.jebo.2006.03.007.

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., and Vilariño, M. V. (2022). "Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development." In: *Global Warming of 1.5°C: An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty.* Edited by V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield. Cambridge, UK: Cambridge University Press, pp. 93–174. ISBN: 978-1-00-915794-0. DOI: 10.1017/9781009157940.

Royal, A. and Rustamov, G. (2018). "Do Small Pecuniary Incentives Motivate Residential Peak Energy Reductions? Experimental Evidence." *Applied Economics* 50 (57), pp. 6193–6202. DOI: 10.1080/00036846.2018.1489508.

Schall, D. L. and Mohnen, A. (2017). "Incentivizing Energy-Efficient Behavior at Work: An Empirical Investigation Using a Natural Field Experiment on Eco-Driving." *Applied Energy* 185 (Part 2), pp. 1757–1768. DOI: 10.1016/j.apenergy.2015.10.163.

Schwartz, D., Fischhoff, B., Krishnamurti, T., and Sowell, F. (2013). "The Hawthorne Effect and Energy Awareness." *Proceedings of the National Academy of Sciences* 110 (38), pp. 15242–15246. DOI: 10.1073/pnas.1301687110.

Schwartz, D., Milfont, T. L., and Hilton, D. (2019). "The Interplay between Intrinsic Motivation, Financial Incentives and Nudges in Sustainable Consumption." In: *A Research Agenda for Economic Psychology.* Edited by K. Gangl and E. Kirchler. Elgar Research Agendas. Cheltenham, UK: Edward Elgar Publishing, pp. 87–103. ISBN: 978-1-78811-606-0. DOI: 10.4337/9781788116060.

Silvi, M. and Rosa, E. P. (2017). "Pro-Environmental Behavior: On the Interplay of Intrinsic Motivations and External Conditions." Working Paper. URL: https://ideas.repec.org/p/uab/wprdea/wpdea1703.html.

Steinhorst, J., Klöckner, C. A., and Matthies, E. (2015). "Saving Electricity – For the Money or the Environment? Risks of Limiting Pro-Environmental Spillover When Using Monetary

Framing." *Journal of Environmental Psychology* 43, pp. 125–135. DOI: 10.1016/j.jenvp.2015.05.012.

Sudarshan, A. (2017). "Nudges in the Marketplace: The Response of Household Electricity Consumption to Information and Monetary Incentives." *Journal of Economic Behavior & Organization* 134, pp. 320–335. DOI: 10.1016/j.jebo.2016.12.015.

Tang, S.-H. and Hall, V. C. (1995). "The Overjustification Effect: A Meta-Analysis." *Applied Cognitive Psychology* 9 (5), pp. 365–404. DOI: 10.1002/acp.2350090502.

Tiefenbeck, V. (2014). "Behavioral Interventions to Reduce Residential Energy and Water Consumption: Impact, Mechanisms, and Side Effects." PhD thesis. Zurich, Switzerland: ETH Zurich. DOI: 10.3929/ethz-a-010263933.

Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., and Staake, T. (2018). "Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation." *Management Science* 64 (3), pp. 1458–1476. DOI: 10.1287/mnsc.2016.2646.

Wang, B., Deng, N., Zhao, W., and Wang, Z. (2022). "Residential Power Demand Side Management Optimization Based on Fine-Grained Mixed Frequency Data." *Annals of Operations Research* 316 (1), pp. 603–622. DOI: 10.1007/s10479-021-04119-8.

Weber, S., Puddu, S., and Pacheco, D. (2017). "Move It! How an Electric Contest Motivates Households to Shift Their Load Profile." *Energy Economics* 68, pp. 255–270. DOI: 10.1016/j.eneco.2017.10.010.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data, Second Edition*. 2nd Edition. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA, USA: MIT Press. ISBN: 978-0-262-23258-6.

Zizzo, D. J., Parravano, M., Nakamura, R., Forwood, S., and Suhrcke, M. (2021). "The Impact of Taxation and Signposting on Diet: An Online Field Study with Breakfast Cereals and Soft Drinks." *Experimental Economics* 24 (4), pp. 1294–1324. DOI: 10.1007/s10683-020-09698-0.

Appendix

Table A1: Treatment effects on energy use per shower when all treatment groups are collapsed

Independent variable	Treatment groups	Temporal effect
Single text message	−0.105	+0.019
	(0.104)	(0.104)
Outcome feedback	−0.461***	+0.160
	(0.141)	(0.148)
Outcome and real-time feedback (1)	−0.658***	+0.171
	(0.132)	(0.156)
Outcome and real-time feedback (2)	−0.509***	+0.186
	(0.122)	(0.163)
Real-time feedback only	−0.245	+0.045
	(0.175)	(0.189)
Observations	15,242	
Clusters	236	

Notes: The table displays the treatment effects on energy use (in kWh) when the treatment groups are collapsed, controlling for individual- and time-fixed effects. Standard errors are presented in parentheses, adjusted for clustering at the participant level. Coefficients were estimated using the within estimator. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Control group observations: 5,588; control group clusters: 90; constant: 2.76; within R^2: 0.003.

Table A2: Treatment effects on shower duration (in minutes)

Independent variable	Financial group	Environmental group	General group	Temporal effect
Single text message	-0.448^{**}	-0.163	-0.123	$+0.058$
	(0.225)	(0.255)	(0.242)	(0.179)
Outcome feedback	-0.888^{**}	-1.140^{***}	-1.110^{***}	$+0.449^{*}$
	(0.360)	(0.337)	(0.279)	(0.253)
Outcome and real-time feedback (1)	-1.200^{***}	-0.978^{***}	-1.050^{***}	$+0.266$
	(0.342)	(0.305)	(0.271)	(0.255)
Outcome and real-time feedback (2)	-1.040^{***}	-0.695^{***}	-0.660^{***}	$+0.322$
	(0.299)	(0.265)	(0.255)	(0.275)
Real-time feedback only	-0.755^{**}	-0.624^{*}	-0.025	$+0.053$
	(0.335)	(0.373)	(0.394)	(0.340)
Observations	5,098	4,955	5,189	
Clusters	79	78	79	

Notes: The table displays the treatment effects on shower duration (in minutes), controlling for individual- and time-fixed effects. Standard errors are presented in parentheses, adjusted for clustering at the participant level. Coefficients were estimated using the within estimator. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Control group observations: 5,588; control group clusters: 90; constant: 5.97; within R^2: 0.004.

Table A3: Treatment effects on energy use per shower and their interaction with the baseline consumption

Independent variable	Financial group	Environmental group	General group	Temporal effect
Single text message	-0.235^{**}	-0.068	$+0.011$	-0.004
	(0.110)	(0.133)	(0.125)	(0.102)
Single text message × Baseline	-0.124^{***}	$+0.029$	-0.121^{*}	–
	(0.044)	(0.064)	(0.066)	
Feedback$^{phases\,2-3b}$	-0.511^{***}	-0.568^{***}	-0.446^{***}	$+0.130$
	(0.133)	(0.126)	(0.130)	(0.122)
Feedback$^{phases\,2-3b}$ × Baseline	-0.252	-0.242^{***}	-0.213^{***}	–
	(0.154)	(0.054)	(0.076)	
Feedback$^{phase\,4}$	-0.486^{***}	-0.480^{***}	$+0.178$	-0.012
	(0.144)	(0.171)	(0.316)	(0.144)
Feedback$^{phase\,4}$ × Baseline	-0.376^{***}	-0.458^{***}	$+0.128$	–
	(0.128)	(0.075)	(0.207)	
Observations	5,098	4,955	5,189	
Clusters	79	78	79	

Notes: The table displays the treatment effects on energy use per shower (in kWh) and their interaction with the centered baseline energy use (mean of 0), controlling for individual effects and time-fixed effects. Standard errors are presented in parentheses, adjusted for clustering at the participant level. Coefficients were estimated using the within estimator. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Control group observations: 5,588; control group clusters: 90; constant: 2.80; within R^2: 0.018.

Paper VI

Leveraging social norms to encourage online learning: Empirical evidence from a blended learning course

Sebastian A. Günther
University of Bamberg

Tobias Veihelmann
University of Bamberg

Thorsten Staake
University of Bamberg

Proceedings of the 41st International Conference on Information Systems, Paper nr. 1205, Hyderabad, India, 2020, https://aisel.aisnet.org/icis2020/digital_learning_env/digital_learning_env/1/

Paper VII

The impact of social norms on students' online learning behavior: Insights from two randomized controlled trials

Sebastian A. Günther
University of Bamberg

Proceedings of the 11th International Conference on Learning Analytics & Knowledge, pp. 12–21, Irvine, CA, USA, 2021, https://doi.org/10.1145/3448139.3448141

Paper VIII

A feedback component that leverages counterfactual explanations for smart learning support

Sebastian A. Günther
University of Bamberg

Felix Haag
University of Bamberg

Konstantin Hopf
University of Bamberg

Philipp Handschuh
Leibniz Institute for Educational Trajectories

Maria Klose
Leibniz Institute for Educational Trajectories

Thorsten Staake
University of Bamberg

In: Digitale Kulturen der Lehre entwickeln – Rahmenbedingungen, Konzepte und Werkzeuge. Edited by L. Mrohs, J. Franz, D. Herrmann, K. Lindner, and T. Staake. Perspektiven der Hochschuldidaktik. Wiesbaden, Germany: Springer VS, forthcoming, https://link.springer.com/book/9783658433789

Introduction

Higher education is a demanding environment that poses new challenges for the learning behavior of students. In contrast to school, in which students' learning is structured and supported in more detail by their teachers (Vosniadou, 2020), university students have to actively plan, monitor, and control their learning behaviors largely on their own to achieve their goals. Yet, many students are struggling in this process, potentially explaining severe dropout rates from study programs (see, e.g., Heublein et al., 2022). The resulting negative consequences accumulate among individuals, leading to major challenges also on a societal level (e.g., loss of time and funds).

With the growing role of digital learning environments in higher education, more and more learning data from students is automatically captured that can be leveraged to actively support them in their learning process. For instance, log data can serve to identify students at risk of dropping out of a course (Foster and Siddle, 2020), allowing instructors to proactively support them. In a similar vein, digital learning environments offer scholars and practitioners novel opportunities to implement a wide range of behavioral interventions to automatically support students' online learning. Recent examples are components that provide feedback on performance (Leung et al., 2023), that help students to monitor their learning progress (Yoon et al., 2021), or that illustrate the online learning time spent (Günther, 2021).

Despite the potential of behavioral interventions within digital learning environments, existing research in this area has neglected that students' personality and learning strategies are inherently different. More precisely, self-regulated learning theory (see, e.g., Pintrich, 2004) implies that students employ different learning strategies and therefore might need personalized guidance for their learning. However, such personalized interventions require high efforts for human instructors (Hogan and Pressley, 1997). Thus, personalized interventions are so far hardly scalable for a wide range of courses.

We argue that digital learning environments can empower such personalized guidance at great scale for university courses using online content: Through combining vast amounts of user activity data with students' course performance data (from previous runs of a course), the learning platform can identify patterns which learning actions have been influential to master a course. Ultimately, when these patterns are deployed in a digital learning environment, a corresponding feedback component can provide current students with personalized instructions on how to improve their online learning and potentially so their course performance.

Against this backdrop, we developed a feedback component that leverages the potential of digital learning environments, which we present in this paper. Specifically, we will briefly summarize our technical approach to initialize the feedback component, show its feedback design, and shed light on our experimental approach to test its effects on learners. The paper concludes with a brief description of the anticipated contributions of the study and the planned next steps.

Research design

For our feedback component, we have instantiated machine learning (ML) models that learn relationships between students' digital learning actions and their overall course performance from past runs of a corresponding course. These models, each of them is solely used for feedback provision in a specific week of the course, are embedded into the learning platform. For each course participant, a week-specific ML model predicts a participant's performance in the final exam of the course, based on the participant's past behavior (log data, time tracking data) and their characteristics (socio-demographic backgrounds yielded from the registration page). To subsequently provide personalized feedback, we employ counterfactual explanation methods, which are a recent technical innovation in the field of explainable ML. Counterfactual explainers estimate how model input parameters (i.e., features) need to change in order to achieve a desired model outcome. Embedded in our feedback component, the explainer method infers what additional actions a learner has to perform (that is the change in input parameters) to improve their exam performance (that is the ML model's output), which Figure 1 illustrates. The feedback component displays the obtained changes in input parameters as actions for exam improvement. By contrast, the ML model's predicted exam performance and the potential for exam improvement are not displayed (we treat these just as internal technical metrics).

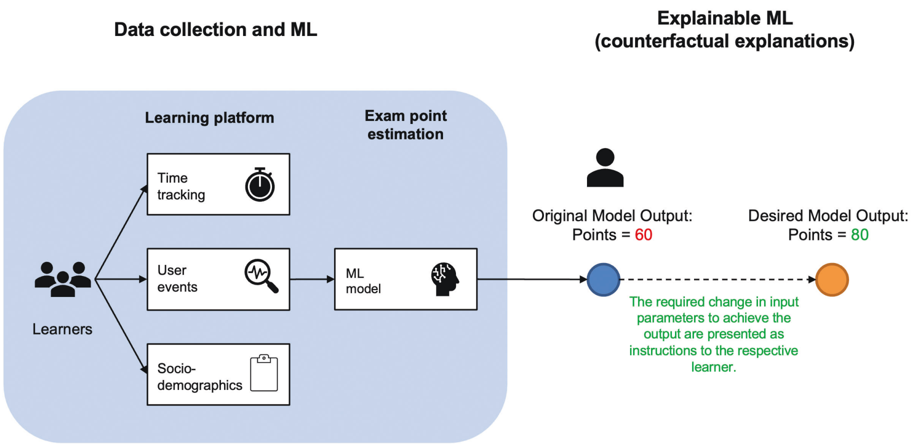


Figure 1: Technical approach

The displayed actions for learners are based on learning strategies and are presented as instructions for learning behavior in the digital learning environment. To put that into context, instructions such as "Watch the video of lecture 3 again" or "Do the quizzes of Lecture 1" should encourage students to catch up with the learning content, monitor their knowledge, or deepen their understanding of specific topics of the lecture. Figure 2 displays the feedback component with examples of personalized actions for learners to improve their performance.

The component is embedded into the main course page of the associated digital learning environment (i.e., open edX) so that it is salient to the learners.

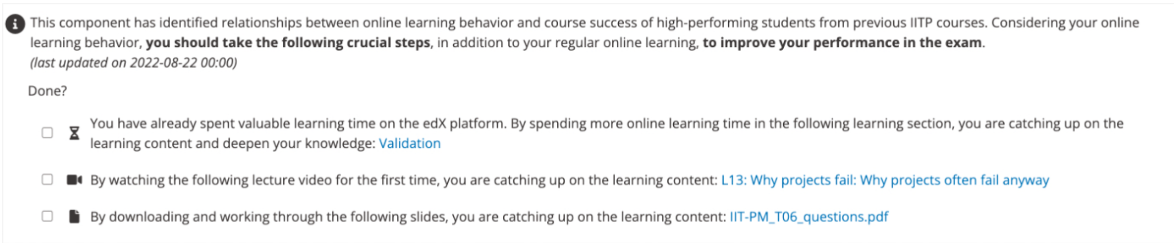


Figure 2: Feedback based on counterfactual explanations

To evaluate the effects of the feedback component, we are running two experimental studies: One in a bachelor level course (summer semester 2022) and one in a master level course (winter semester 2022/2023). Each study follows a difference-in-differences experimental design. More precisely, we provide a group of students with feedback after a baseline phase, while the control group does not receive any feedback. We hypothesize that our feedback component will have desirable effects on students' course success (for an overview of learning techniques, see Dunlosky et al., 2013; and for the effectiveness of feedback, see Hattie and Timperley, 2007). We investigate course success in terms of exam-taking rate and points in the exam. To better understand the effects of the feedback intervention, we additionally conduct a pre- and post-survey that capture influential psychological constructs on online learning. These constructs relate for example to students' self-regulated learning skills, their procrastination behavior, and feedback acceptance. Additional analyses using these constructs will allow us to understand which students implement the instructions from the feedback component and benefit most from them in terms of study behavior, exam participation, and exam grade. Figure 3 shows the experimental setup for each of the two studies.

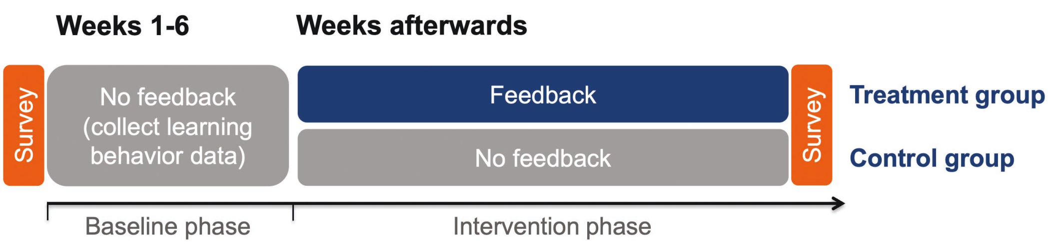


Figure 3: Experimental setup

Conclusion

In this paper, we have presented a novel component that applies counterfactual explanations for providing personalized feedback. Even though we still have to statistically evaluate the effects of this feedback component on learners, its underlying technical approach is promising. The approach has the potential to unite learner characteristics (e.g., self-regulation skills, susceptibility to procrastination, etc.) and behavioral data to derive personalized guidance for learners on how to improve educational success. In doing so, the feedback component, operating in a digital sphere, can potentially promise more benefits than those that can be expected from a human instructor: Providing personalized feedback at scale. Given the relevance of learning strategies on academic success in higher education (Broadbent and Poon, 2015), this paper encourages practitioners and scholars to consider such scalable approaches for empowering personalized learning support.

Acknowledgments

This work was funded in part by "Stiftung Innovation in der Hochschullehre" within the project "Digitale Kulturen der Lehre entwickeln (DiKuLe)."

References

Broadbent, J. and Poon, W. (2015). "Self-Regulated Learning Strategies & Academic Achievement in Online Higher Education Learning Environments: A Systematic Review." *The Internet and Higher Education* 27, pp. 1–13. DOI: 10.1016/j.iheduc.2015.04.007.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., and Willingham, D. T. (2013). "Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology." *Psychological Science in the Public Interest* 14 (1), pp. 4–58. DOI: 10.1177/1529100612453266.

Foster, E. and Siddle, R. (2020). "The Effectiveness of Learning Analytics for Identifying At-Risk Students in Higher Education." *Assessment & Evaluation in Higher Education* 45 (6), pp. 842–854. DOI: 10.1080/02602938.2019.1682118.

Günther, S. A. (2021). "The Impact of Social Norms on Students' Online Learning Behavior: Insights from Two Randomized Controlled Trials." *Proceedings of the 11th International Conference on Learning Analytics & Knowledge*. Irvine, CA, USA. DOI: 10.1145/3448139.3448141.

Hattie, J. and Timperley, H. (2007). "The Power of Feedback." *Review of Educational Research* 77 (1), pp. 81–112. DOI: 10.3102/003465430298487.

Heublein, U., Hutzsch, C., and Schmelzer, R. (2022). *Die Entwicklung der Studienabbruchquoten in Deutschland*. Hannover, Germany: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung. DOI: 10.34878/2022.05.dzhw_brief.

Hogan, K. and Pressley, M. (1997). *Scaffolding Student Learning: Instructional Approaches & Issues.* Advances in Teaching and Learning. Cambridge, MA, USA: Brookline Books. ISBN: 1-57129-036-2.

Leung, A. C. M., Santhanam, R., Kwok, R. C.-W., and Yue, W. T. (2023). "Could Gamification Designs Enhance Online Learning Through Personalization? Lessons from a Field Experiment." *Information Systems Research* 34 (1), pp. 27–49. DOI: 10.1287/isre.2022.1123.

Pintrich, P. R. (2004). "A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students." *Educational Psychology Review* 16 (4), pp. 385–407. DOI: 10.1007/s10648-004-0006-x.

Vosniadou, S. (2020). "Bridging Secondary and Higher Education – The Importance of Self-regulated Learning." *European Review* 28 (S1), pp. S94–S103. DOI: 10.1017/S1062798720000939.

Yoon, M., Hill, J., and Kim, D. (2021). "Designing Supports for Promoting Self-Regulated Learning in the Flipped Classroom." *Journal of Computing in Higher Education* 33 (2), pp. 398–418. DOI: 10.1007/s12528-021-09269-z.

Paper IX

The behavioral mechanisms behind feedback – A preliminary model for quantifying cause-effect relationships

Carlo Stingl
University of Bamberg

Sebastian A. Günther
University of Bamberg

Thorsten Staake
University of Bamberg

Proceedings of the 30th European Conference on Information Systems, Paper nr. 1122, Timișoara, Romania, 2022, https://aisel.aisnet.org/ecis2022_rip/7/

Appendix

Publications

Journal articles (Peer reviewed)

Dangis, G., Terho, K., Graichen, J., Günther, S. A., Rosio, R., Salanterä, S., Staake, T., Stingl, C., and Pakarinen, A. (2023). "Hand Hygiene of Kindergarten Children—Understanding the Effect of Live Feedback on Handwashing Behaviour, Self-Efficacy, and Motivation of Young Children: Protocol for a Multi-Arm Cluster Randomized Controlled Trial." *PLOS ONE* 18 (1), Article nr. e0280686. DOI: 10.1371/journal.pone.0280686.

Weigert, A., Hopf, K., Günther, S. A., and Staake, T. (2022). "Heat Pump Inspections Result in Large Energy Savings When a Pre-Selection of Households Is Performed: A Promising Use Case of Smart Meter Data." *Energy Policy* 169, Article nr. 113156. DOI: 10.1016/j.enpol.2022.113156.

Günther, S. A., Staake, T., Schöb, S., and Tiefenbeck, V. (2020). "The Behavioral Response to a Corporate Carbon Offset Program: A Field Experiment on Adverse Effects and Mitigation Strategies." *Global Environmental Change* 64, Article nr. 102123. DOI: 10.1016/j.gloenvcha.2020.102123.

Hauser, M., Günther, S. A., Flath, C. M., and Thiesse, F. (2019). "Towards Digital Transformation in Fashion Retailing: A Design-Oriented IS Research Study of Automated Checkout Systems." *Business & Information Systems Engineering* 61 (1), pp. 51–66. DOI: 10.1007/s12599-018-0566-9.

Schöb, S., Günther, S. A., Regensburger, K., and Staake, T. (2018). "NIWM: Non-Intrusive Water Monitoring to Uncover Heat Energy Use in Households." *Computer Science - Research and Development* 33 (1-2), pp. 127–133. DOI: 10.1007/s00450-017-0353-8.

Conference articles (Peer reviewed)

Günther, S. A., Haag, F., Hopf, K., Handschuh, P., Klose, M., and Staake, T. (forthcoming). "A Feedback Component That Leverages Counterfactual Explanations for Smart Learning Support." In: *Digitale Kulturen der Lehre entwickeln – Rahmenbedingungen, Konzepte und Werkzeuge*. Edited by L. Mrohs, J. Franz, D. Herrmann, K. Lindner, and T. Staake. Perspektiven der Hochschuldidaktik. Wiesbaden, Germany: Springer VS.

Haag, F., Günther, S. A., Hopf, K., Handschuh, P., Klose, M., and Staake, T. (2023). "Addressing Learners' Heterogeneity in Higher Education: An Explainable AI-Based Feedback Artifact for Digital Learning Environments." *Proceedings of the 18th International Conference on Wirtschaftsinformatik*. Paderborn, Germany.

Graichen, J., Stingl, C., Günther, S. A., Staake, T., Pakarinen, A., Rosio, R., Salanterä, S., and Terho, K. (2022). "No Longer Without a Reward: Do Digital Rewards Crowd Out Intrinsic Motivation of Young Children?" *Proceedings of the 43rd International Conference on Information Systems*. Copenhagen, Denmark.

Michels, L., Ochmann, J., Günther, S. A., Laumer, S., and Tiefenbeck, V. (2022). "Empowering Consumers to Make Environmentally Sustainable Online Shopping Decisions: A Digital Nudging Approach." *Proceedings of the 55th Hawaii International Conference on System Sciences*. Maui, HI, USA. DOI: 10.24251/hicss.2022.574.

Stingl, C., Günther, S. A., and Staake, T. (2022). "The Behavioral Mechanisms Behind Feedback – A Preliminary Model for Quantifying Cause-Effect Relationships." *Proceedings of the 30th European Conference on Information Systems*. Timișoara, Romania.

Günther, S. A. (2021). "The Impact of Social Norms on Students' Online Learning Behavior: Insights from Two Randomized Controlled Trials." *Proceedings of the 11th International Conference on Learning Analytics & Knowledge*. Irvine, CA, USA. DOI: 10.1145/3448139.3448141.

Stingl, C., Günther, S. A., and Staake, T. (2021). "A Feedback Information System for Improving Hand Hygiene on a Personal and Organizational Level." *Proceedings of the 16th International Conference on Wirtschaftsinformatik*. Duisburg, Germany.

Günther, S. A., Stingl, C., Coroamă, V. C., Schöb, S., and Staake, T. (2020). "Empowering Personalized Feedback on Hot Water Usage: A Field Study with Shower Meters." *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. Brno, Czech Republic. DOI: 10.1145/3341105.3374053.

Günther, S. A., Veihelmann, T., and Staake, T. (2020). "Leveraging Social Norms to Encourage Online Learning: Empirical Evidence from a Blended Learning Course." *Proceedings of the 41st International Conference on Information Systems*. Hyderabad, India.

Hauser, M., Günther, S., Flath, C. M., and Thiesse, F. (2017a). "Leveraging RFID Data Analytics for the Design of an Automated Checkout System." *Proceedings of the 13th International Conference on Wirtschaftsinformatik*. St. Gallen, Switzerland.

Hauser, M., Günther, S. A., Flath, C. M., and Thiesse, F. (2017b). "Designing Pervasive Information Systems: A Fashion Retail Case Study." *Proceedings of the 38th International Conference on Information Systems*. Seoul, South Korea.

Manuscripts

Günther, S. A., Schöb, S., Coroamă, V. C., Tiefenbeck, V., Mattern, F., and Staake, T. (under review). "All Eyes on Consumption Feedback: A Randomized Controlled Trial on the Interplay of Financial Incentives, Environmental Appeals, and Consumption Feedback." Working Paper. Submitted to Experimental Economics.