Computers in Human Behavior 113 (2020) 106504

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

Assisting mental accounting using smartphones: Increasing the salience of credit card transactions helps consumer reduce their spending

Johannes Huebner *, Elgar Fleisch, Alexander Ilic

Swiss Federal Institute of Technology in Zurich, Chair of Information Systems, Weinbergstrasse 56-58, 8006, Zurich, Switzerland

ARTICLE INFO

Keywords:
Credit card transaction salience
Exceptional purchases
Transaction rehearsal
Spending feedback
Financial literacy
Field experiment

ABSTRACT

Credit card-related over-spending represents an important issue for consumers. Over-spending arises in parts from reduced payment transparency compared to cash and other payment methods. Additionally, week-by-week credit card spending exhibits high variance even on an intrapersonal level, which makes it hard to intuitively learn from prior transactions and control one's spending. As mobile-mediated information systems have been proven effective in delivering behavior change interventions, this study investigates the efficacy of using a novel smartphone application that increases the salience of credit card transactions to help consumers control their cashless payments better and ultimately spend less. We implemented a goal-setting feature and provided weekly goal attainment feedback highlighting ordinary, exceptional, or both types of purchases. This work was conducted as a field experiment, studying a large sample of credit card consumers in the wild over several months, which yielded a significant reduction in spending with unobtrusive interventions. It further highlights the importance of including exceptional purchases in households' spending budgets and discusses how people adjusted their consumption to lower their expenditures.

1. Introduction

A shift from cash towards cashless payment methods is happening in many economies around the world, where consumers increasingly use payment methods such as credit cards, mobile payments or other digital payment forms instead of cash (Capgemini, 2019). For example, between 2010 and 2014, the number of cashless payments increased by 34%, with payment cards experiencing the biggest increase; the extent to which cash has already been replaced still varies significantly across countries, but the trend is undeniable on a global level (Batiz-Lazo, 2016; Pratz, Bloos, Engebretsen, & Gawinecki, 2013; Raconteur, 2016).

This trend is by no means a result of chance. Cashless forms of payment offer greater convenience, greater liquidity (in the case of credit cards), less liability in case of theft, as well as the ability to track expenses after the fact. This post-hoc transparency lets consumers meaningfully maintain digital budgets, and could thereby inform future consumption behavior. However, it stands to question whether current online and mobile banking systems provide sufficiently salient feedback to effectively support consumers' future decision-making–especially considering that many consumers maintain multiple banking relationships and therefore arguably lack a complete overview over their financial lives (Kaye, McCuistion, Gulotta, & Shamma, 2014; Reville, 2019).

In addition, research suggests that the benefits of cashless payments come at a price, *e.g.* consumers spend more money on the same items, focus more on product properties rather than the associated cost, and are more likely to indulge in treats and luxury items when using credit cards instead of cash (Chatterjee & Rose, 2012; Norvilitis et al., 2006; Prelec & Simester, 2001; Raghubir & Srivastava, 2008; Soman, 2003). Additionally, i) an increase in total outstanding credit card debt, ii) an estimated one third of US households carrying forward at least some credit card debt from month to month, and iii) stagnating median income levels in combination suggest that credit cards are increasingly used as an alternative source of income, a pattern that exposes households to financial hardships, *e.g.* in case of job loss (Hodson, Dwyer, & Neilson, 2014; NFCC, 2016; ProQuest, 2017; US Census Bureau, 2017).

When choosing a form of payment, consumers therefore face a trade-off between the convenience and post-hoc-transparency of cashless payments on one side, and greater financial discipline when using cash on the other side. Still, we do not suggest giving up digital forms of payment. Instead, similar to how it was necessary for pedestrians to rehearse new behaviors when automobiles emerged, *e.g.* ensure no car is

* Corresponding author.
E-mail addresses: johannes.huebner@autoidlabs.ch (J. Huebner), efleisch@ethz.ch (E. Fleisch), alexander.ilic@autoidlabs.ch (A. Ilic).

https://doi.org/10.1016/j.chb.2020.106504
Received 2 September 2019; Received in revised form 20 July 2020; Accepted 23 July 2020
Available online 29 July 2020
0747-5632/

approaching when crossing the street (Thompson, 2014), we argue that consumers should also be guided in the responsible use of digital payment methods – to help them navigate the challenges credit cards incur while enjoying their benefits.

We address this trade-off by increasing the salience of cashless payments through personalized feedback interventions, thus helping people gain better control over their credit card spending. The main research question guiding this work was: **To what extent are smartphone-mediated feedback interventions effective in reducing credit card spending?**

Thanks to a collaboration with a credit card issuer, this research leveraged a sizeable group of consumers ($N > 1'000$), who downloaded a mobile app specifically developed for the purpose of this study, which was linked to their credit cards. In doing so, this paper makes multiple important contributions: First, we offer an empirically tested, technology-mediated and therefore scalable strategy for how consumers can be assisted in reducing their spending with credit cards. The solution offered in this paper would be easily transferable to other payment methods such as debit cards or mobile payments, and it could be generalized into a blueprint for other behavioral challenges in the financial domain. Second, we discuss the role of *exceptional* transactions and how they should be incorporated in spending feedback for effective budgeting. Third, the intersection between mental accounting and technological advances especially in the financial industry are currently underexplored (Zhang & Sussman, 2018). This work provides insights how technology and mental accounting can be leveraged in combination to help people overcome their cognitive biases and assist decision-making. To the best of our knowledge, it is one of the first studies doing so. Fourth, this paper quantifies to what extent the "credit card premium" can be counter-balanced through comparatively simple behavioral interventions.

2. Related work

2.1. The effect of payment methods on salience of spending behavior: pain of payment

Prior research has found that the (planned) choice of payment method affects our consumption behavior. As a wide-spread low-salience payment method, especially credit cards have been subject to many studies, which typically concluded that people exhibit a strong tendency to spend more for the same items than they would have when using cash. In fact, the presence of credit card scheme logos alone can prime people to spend more and make spending decisions faster (Feinberg, 1986). For example, in one study, researchers conducted an auction for sports tickets and found that respondents who were told they had to pay by credit card on average bid 64% more than those who were told they would have to pay in cash (Prelec & Simester, 2001). In another study, researchers asked consumers in the parking lot of a grocery store for their receipts and payment method after they completed their purchase, and found that those paying by credit card paid on average 32% more than cash users, which was especially due to higher spending on flexible items such as treats and luxuries (Soman, 2003). Similar findings have been replicated in further research (Raghubir & Srivastava, 2008; Runnemark; Hedman, & Xiao, 2015).

Behavioral economists attribute these differences in behavior when people pay with credit cards instead of cash to varying intensities of a psychological pain of payment, which depends on the transparency of the used payment form (Prelec & Loewenstein, 1998; Raghubir & Srivastava, 2008; Soman, 2003). The very fact that money is spent, as well as the amount, is more obvious when using banknotes and coins, whereas swiping a credit card or tapping a prepaid public transportation card against a card reader introduces an abstraction of money, and is thus less transparent (Raghubir & Srivastava, 2008; Soman, 2003).

Additionally, credit cards are special compared to other payment cards, because the actual wealth depletion occurs with a considerable delay after the purchase event (or acquisition phase)–often one to two months–, and also because purchases are aggregated into a monthly bill (Soman & Lam, 2002). This circumvents some of our mental accounting safeguards and makes it challenging for people to intuitively learn from purchases and adapt their behavior. In particular, the category depletion effect describes that future spending decisions in a given category (such as groceries) are affected by prior spending in the same category. However, delaying the wealth depletion effect significantly weakens this effect, thus causing consumers to spend more (Gourville & Soman, 1998; Soman & Lam, 2002).

What remains underexplored is how consumers can be aided in using credit cards and other digital payment forms responsibly, even though practitioners and researchers have envisioned some first possible remedies. For example, US-based design firm New Deal Design have developed a concept for a tangible payment wallet dubbed Scrib, which would require its users to perform a number of swipes on the wallet's surface to make a payment, each of which would simulate giving away a digital banknote (New Deal Design, 2016). Researchers at MIT have developed a proverbial wallet, which is connected to bank accounts or credit cards and communicates key metrics through haptic feedback (Kestner, Leithinger, Jung, & Petersen, 2009). However, also simpler methods such as push notifications and reminders have been shown to be effective in helping consumers favorably adjust their financial behavior (*e.g.* saving money or reducing expenses, see *e.g.* Benartzi, 2017; Karlan, McConnell, Mullainathan, & Zinman, 2010).

2.2. Mental accounting of expenses and the role of exceptional transactions

In addition to the aforementioned challenges arising from the use of low-salience payment methods, consumers are further influenced by mental biases leading to suboptimal decision-making (Tversky & Kahneman, 1975). It has been established that consumers define per-category budgets either in a formal budgeting system or mentally, and they track purchases against the corresponding budget (Heath & Soll, 1996; Soman & Lam, 2002). For this, consumers first need to notice a particular purchase, and then assign it to a category (Heath & Soll, 1996). Completely overlooking a transaction becomes increasingly difficult thanks to cashless payments that leave a digital trace, which can be looked up any time in most online or mobile banking systems.

Nonetheless, consumers usually do not have a complete picture of their financial lives readily available when needed, causing them to take decisions under incomplete information. This phenomenon has received academic attention from multiple perspectives. For example, consumer behavior researchers have investigated how incomplete information about a set of products impacts their purchase decisions (*e.g.* Kivetz & Simonson, 2000). In the realm of personal financial management, even those consumers that put a formal money management system in place may have blind spots due to them possessing multiple bank accounts and payment methods (Kaye et al., 2014). Transactions between self-owned accounts (e.g. transfers to accounts jointly owned by spouses), saving and investment transactions, and transactions that are reimbursed – sometimes with considerable delay – (e.g. business expenses, (partial) refunds for returned items of an online purchase) add to the challenge of keeping an accurate picture of one's household spending. These issues aggravate for consumers who do not have formal budgeting systems in place.

In either case, the way that consumers categorize a purchase is essential, since it influences the decision-making for future purchases: Consumers first define what they believe is a reasonable budget for a given category within a given timeframe, and given their financial situation. Even if they fail to do so explicitly, consumers often have at least a vague frame of reference from their prior expenditure and are constrained by their disposable income. Then, whenever they are faced with a spending decision, instead of analyzing one's entire financial situation–which would incur immense cognitive effort–, they take a shortcut and evaluate a purchase against the per-category budget (Zhang & Sussman, 2018). This decomposition step transforms everyday spending decisions from complex puzzles into manageable problems, and thus constitutes an effective money management technique that will directly influence consumers' spending (Heath & Soll, 1996; Ülkümen, Thomas, & Morwitz, 2008). In fact, this technique is so effective that *e.g.* people with self-efficacy issues may even cheat on themselves by mentally categorizing individual purchases into categories with some remaining budget. For instance, they may categorize a dinner in a restaurant into the category "Food" rather than perhaps "Entertainment", rationalizing the decision with a focus on the nutritional value the food had, rather than on the pleasure drawn from having a meal in the atmosphere of a restaurant (Heath & Soll, 1996).

Especially for purchase categories with frequent and comparatively non-discretionary transactions, per-category budgets become well-rehearsed, prior in-category spending will deplete the available budget, and decrease the likelihood of further spending in said category until the accounting period ends and the budgets are replenished (*e.g.* Heath & Soll, 1996; Soman, 2001, 2003; Soman & Lam, 2002; Soster, Gershoff, & Bearden, 2014). Examples for this include categories like groceries, phone bills, insurance, rent or transportation. There are many factors that influence the size of the aforementioned category depletion effect, such as a transaction's dollar amount (Heath & Soll, 1996), the temporal distance between prior spending and the pending decision (Gourville & Soman, 1998), rehearsal intensity (and thus salience) of a prior purchase, as well as the immediacy of the corresponding payment (*e.g.* with a delay after the purchase) (Soman, 2001), and relative timing of acquisition and payment (*e.g.* payment at the time of acquisition, or post-payment) (Soman & Lam, 2002).

In contrast, the prediction and management of less frequently occurring or unusual, exceptional transactions is more challenging. Sussman and Alter posited that consumers often consider such exceptional transactions in an overly narrow category with no or little prior spending (narrow choice bracketing), thus sidelining the category depletion effect and the resulting budgetary discipline. Additionally, consumers tend to underestimate the frequency with which such exceptional transactions occur (Sussman & Alter, 2012). Examples for such purchases may be gifts or electronic gadgets–instead of thinking of purchases in these broad categories, consumers may even assign individual categories per purchase, such as "smartphone" or "valentine's day gift for my spouse". Since the purchase of a smartphone and a valentine's day gift seem unrelated, they are also unconnected when attempting to estimate the frequency of exceptional transactions. In other words, consumers fail to make the connection between these seemingly unrelated, rare events (*e.g.* once per year) and therefore underestimate the overall frequency of such transactions. Because of this, people also tend to grant themselves greater financial slack and ultimately overspend on such exceptional transactions (Soman & Lam, 2002; Sussman & Alter, 2012).

The distinction between ordinary and exceptional spending contains no value judgement, *i.e.* one category is not necessarily better or worse than the other; instead, we simply refer to their relative frequency (Bhattacharjee & Mogilner, 2013; Sussman & Alter, 2012). However, when it comes to helping consumers achieve greater budgetary control and prevent over-spending, a focus should be set on purchases that are considered exceptional.

2.3. Strategies for supporting behavior changes

Any solution to the issues outlined above would have to be unobtrusive enough to uphold the convenience of digital payments, and yet be powerful enough to counterbalance the premium people pay when using credit cards. In this context, the work of Richard Thaler comes to mind: Nudges are designed to leave consumers in full control over their decisions, while setting the context or decision parameters such that beneficial outcomes are promoted (Thaler & Sunstein, 2008). This can be achieved with a number of detailed strategies, nine of which were summarized by Dolan et al. (2012) with the mnemonic MINDSPACE: For instance, the initial "M" represents the *Messenger* strategy, which is based on the fact that how we perceive and react to information greatly depends on factors such as the messenger's authority, expertise, and emotional connection messenger and recipient share – an important insight that can be leveraged to nudge behavior. The remaining eight strategies are dubbed *Incentives, Norms, Defaults, Salience, Priming, Affect, Commitments, Ego* (Dolan et al., 2012). Some of these strategies may be problematic to apply in the given context of financial transactions, but there is one strategy that particularly stands out in light of the above discussion: Why not tackle the source of the issue that behavioral economists have identified as the source of the phenomenon that we spend more when using credit cards – decreased pain of payment (Prelec & Loewenstein, 1998; Raghubir & Srivastava, 2008; Soman, 2003) – heads-on and increase the salience of credit card transactions? This appears to be a promising and unobtrusive intervention strategy to help people reduce their spending.

The phenomena described in the above sections may vary between individuals, hence the efficacy of any intervention geared towards helping consumers reduce spending may also vary. For example, low-salience payment methods are particularly challenging for compulsive shoppers and may lead to more serious financial issues than for non-affected consumers (Lo & Harvey, 2011).

In the context of the study at hand, the concept of frugality deserves special consideration, for which multiple definitions highlighting different perspectives exist (Lastovicka, Bettencourt, Hughner, & Kuntze, 1999). We follow the definition brought forward by Lastovicka et al. (1999): "Frugality is a unidimensional consumer lifestyle trait characterized by the degree to which consumers are both restrained in acquiring and in resourcefully using economic goods and services to achieve longer-term goals". Highly frugal, "tightwad" consumers are expected to generally pay greater attention to their spending and experience greater pain of payment than "spendthrifts" (Berman, Tran, Lynch Jr, & Zauberman, 2016; Rick, Cryder, & Loewenstein, 2008). Consequently, high-frugality consumers may already have great budgeting systems in place, and have optimized their spending to such an extent that exposing them to any sort of intervention aimed at reducing their spending even further will likely be ineffective – the opposite is expected for their low-frugality counterparts.

In addition, when it comes to behavior change interventions, existing routines play a central role. As habit formation theories posit, new routines may be challenging to establish, but it can also be hard to break with pre-existing routines. Returning to the context of this study, capturing financial routine is not a trivial task: Research on the econometric modeling of household spending provides evidence for the diversity of spending across households depending on factors such as income (*e.g.* Chai, Rohde, & Silber, 2015), and it has outlined the challenges associated with accurately modeling spending for individuals or households (Lawson, 2013). Especially for such households that have some disposable income to allocate at their discretion, the total intra-household spending may also vary considerably over time, depending on the modeled time period. For instance, while in most cases, the total yearly spending of any given household will likely only exhibit little variance due to income constraints, the same household's monthly, weekly, and daily spending is likely to exhibit larger amounts of variance, because some purchases are only made in certain (longer) intervals, *e.g.* yearly insurance premiums, monthly subscriptions for communication services, or booking a trip. Moreover, some households generally exhibit far more stable spending patterns than others, *e.g.* by frequenting the same grocery stores with only slight variations in shopping lists, by spending similar amounts on hobbies, hardly changing the frequency of bar and restaurant visits, etc. Such stable spending patterns may originate from financial necessity – low-income households have been shown to allocate most of their available income to food (see *e.g.* Chai et al., 2015; Lawson, 2013), from particularly stable life situations, or simply be habits formed over time.

3. Hypothesis development

As illustrated in the above section, researchers have identified several potential sources of over-spending, such as i) the mere use of digital payment forms, and ii) failing mental accounting safeguards for exceptional transactions. With this work, we thus aim at helping consumers use digital payment methods, and especially credit cards, with greater budgetary control through increased transaction salience, while also considering the different types of transactions – ordinary, exceptional, or both categories of purchases. While many salience-increasing strategies are conceivable (*e.g.* Kestner et al., 2009; New Deal Design, 2016), we first queried the available literature and concluded that requiring participants to actively review and categorize every transaction would be a good fit. It has long been acknowledged that performing semantic tasks, such as categorizing transactions, leads to better recall compared to simpler tasks, *e.g.* simply showing people the amount of a transaction, or asking them to re-iterate the amounts themselves (Craik & Lockhart, 1972). Indeed, the accurate recall of a particular transaction is a prerequisite for the category depletion effect to work. This is especially critical for credit card payments, where the wealth depletion event usually occurs with considerable delay after the purchase event (decoupling). As the work of Soman (2001) has shown, prompting consumers to rehearse a transaction can improve the recall of past expenses and thus lead to reduced future spending. However, it is conceivable that the rehearsal effect of transactions will lead to an immediate spending adjustment only of ordinary spending due to the category depletion effect (Soman, 2001). However, for exceptional purchases, the rehearsal of individual transactions is not expected to have any effect on future spending, since consumers mentally budget them too narrowly, hence no available budget gets depleted, and future spending happens unrestricted by the extraordinary purchase. Therefore, we argue that the mere rehearsal of individual transactions will not be enough for helping consumers sustainably reduce their spending, because the review of individual transactions is too fine-granular of a task for people to draw conclusions about their overall spendinsg across time periods. To examine whether this is indeed the case, we define our first hypothesis as follows:

H1. Merely *reviewing* and performing a simple semantic task regarding each transaction will not suffice to help consumers reduce their spending, compared to those who do not perform this task.

The role of immediacy and aggregation level when providing feedback have been studied in domains such as energy consumption, where highly disaggregated real-time feedback effectively impacts behavior (Tiefenbeck, 2014). This makes sense because consumers are often unaware of their energy consumption, hence real-time feedback enables them to change their behavior effectively. In contrast, consumers know how much money they spend in the moment they do, but may lack an appropriate feedback system regarding their spending over time, which would enable behavioral changes. Even though monthly credit card bills provide some feedback, they are highly aggregated and arrive with considerable delay, making it difficult for consumers to link individual actions (purchases) to outcomes (total spending). We thus expect that both the rehearsal of transactions as well as providing reflective information using evidence of more than a single purchase data point will be necessary to achieve a spending reduction.

As a result, increasing the salience of transactions made with a low-salience payment method like credit cards by asking consumers to review and categorize every transaction, while explicitly highlighting both ordinary and exceptional spending in a weekly goal attainment feedback is expected to reduce customers' overall spending. We therefore formulate the following hypothesis:

H2a. Compared to consumers in a low-salience setting (control), consumers who are prompted to classify each transaction as either ordinary or exceptional and receive weekly feedback highlighting both categories (Treatment Group 1, or TG1 in short) will reduce their spending.

However, it is also conceivable that people reduce their spending when the salience of transactions is increased in the same way, while particularly highlighting those transactions marked as *ordinary* in the weekly feedback, even if *exceptional* purchases do not count towards the spending goal. After all, this mimics the budgeting techniques many people apply, by defining budgets for individual categories (*e.g.* groceries) and regularly comparing expenses against these planned budgets; expenses that do not fit into any of the categories will likely be consolidated as "unforeseen expenses" or be subtracted from a financial buffer (Hilgert, Hogarth, & Beverly, 2003). Regardless of whether people apply such techniques mentally or in a formalized system, if they were completely dysfunctional, they would likely be replaced by a better alternative. In addition, even though we may be prone to underestimating their frequency, it is conceivable that due to their outstanding nature we already pay more attention to such *exceptional* purchases especially when prompted to systematically review each and every transaction, *i.e.* we more carefully research alternatives and we do a better job in including them in our spending budgets.

We therefore hypothesize that particularly highlighting those transactions marked as *ordinary*, will also result in a reduction of overall spending.

H2b. Compared to consumers in a low-salience setting (control), consumers who are prompted to classify each transaction as either ordinary or exceptional and *receive* weekly feedback highlighting the **ordinary category** (Treatment Group 2, or TG2 in short) will reduce their spending.

Prior work moreover posited that consumers are particular prone to underestimating the frequency and total volume of *exceptional* transactions (Soman & Lam, 2002; Sussman & Alter, 2012), which causes them to grant themselves greater financial slack for and overspend on those *exceptional* purchases, *e.g.* they rationalize that it is ok for them to spend a little more on a new smartphone, since they only buy one every X time units. We therefore hypothesize that particularly highlighting the frequency and volume of transactions marked as *exceptional* will make people aware of their bias and result in a reduction in overall spending, too.

H2c. Compared to consumers in a low-salience setting (control), consumers who are prompted to classify each transaction as either ordinary or exceptional and receive weekly feedback highlighting the **exceptional category** (Treatment Group 3, or TG3 in short) will reduce their spending.

As discussed in the previous chapter, there are additional variables that may impact the effect of such interventions. First, the level of consumers' frugality will likely moderate the effect, since high-frugality consumers are expected to already be proficient at managing their financial lives, and the interventions outlined above are unlikely to reduce their spending even further. However, the same interventions are expected to work very well for less frugal customers.

H3a. . Consumers' level of frugality moderates the effect of salience-increasing interventions on spending reduction.

Finally, the previous chapter outlined that some consumers exhibit much more stable spending patterns than others, which may moderate the effectiveness of our interventions. While modeling the stability of spending patterns itself has been subject to academic research (*e.g.* Chai et al., 2015; Lawson, 2013), we utilize a rather simple construct for this purpose – the variance in total weekly spending prior to the start of the study. We thus hypothesize that:

H3b. . The variance in consumers' baseline spending moderates the effect of salience-increasing interventions on spending reduction.

Fig. 1 summarizes our research model with the above hypotheses.

4. Research design

While researchers of consumer behavior often choose surveys or controlled lab settings, we set out to test our hypotheses in the wild, and over a sufficient period of time, to gain a rich picture of the efficacy of our interventions under natural conditions. The study took place over a period of three months (two-week warmup period plus 13 weekly feedback cycles), with all participants beginning the study within a one-week period in early September, which put the study in a somewhat regular phase of the year, *i.e.* avoiding major holiday seasons.

The goal of this study was to test technology-mediated, scalable mechanisms to impact consumers' financial behavior that could be rolled out to a large audience by financial service providers. Similarly, behavioral interventions have been successfully trialed in field experiments in domains such as health (see *e.g.* Zhao, Freeman, & Li, 2016 for a review of studies), ii) reducing energy consumption (*e.g.* Loock, Staake, & Thiesse, 2013), or water consumption (Tiefenbeck, 2014) and also in the financial domain (*e.g.* Akbaş, Ariely, Robalino, & Weber, 2016; Karlan et al., 2010; Taubinsky & Rees-Jones, 2017) and therefore provide confidence that such interventions are also applicable to the given context.

4.1. Construct operationalization

We set out to design a between-subject, randomized controlled trial field experiment using mobile apps as a channel to deliver interventions. The first step was to operationalize the theoretical constructs outlined in the previous section, the result of which is summarized in Table 1 including the underlying rationale for our design choices. As a dependent variable, we chose the overall credit card spending in a given week, and normalized it by the individual mean baseline spending (over 52 weeks), *i.e.* 1.0 would indicate no change, 1.5 would indicate a 50% increase in spending, and so on. This also means that the results presented in this paper are not influenced by whether participants stayed within their self-defined spending goals. For the remainder of this paper, with "reduction in spending", we refer to a participant spending less than their individual baseline spending period prior to the experiment start.

4.2. Research setting and data collection

For the purpose of our study, we cooperated with the largest Swiss credit card issuer to be able to study consumers' real-world credit card transactions. Our partner sent email invites to a random set of 31′000 eligible customers who had a standard or gold credit card and used it at least on a somewhat regular basis in the previous year, *i.e.* for at least three weekly transactions on average. They received information on the scope and purpose of the study, and had the chance to win a laptop, a tablet or a smartphone if they chose to partake. The lottery was independent from users' spending behavior, with prizes being raffled out amongst all users who downloaded the app and completed the onboarding (see below), which was clearly stated in the invite.

Table 1
Constructs and operationalization.

Construct	Operationalization	Rationale
Salience	Prompted consumers to review every transaction and perform a simple semantic task (binary categorization). After the warmup period, participants additionally received weekly feedback focused on either or both of the categories	Compared to cash transactions, credit card payments are less salient, which can lead to people spending at least 15–30% more, as has been demonstrated in a number of studies (*e.g.* Prelec & Simester, 2001; Raghubir & Srivastava, 2008; Soman, 2003). Additionally, exceptional transactions can aggravate this phenomenon, since people underestimate the frequency with which they occur and may thus grant themselves greater financial slack. The warmup period (without feedback) tested whether the categorization feature alone would sufficiently increase the salience of credit card transactions to achieve a reduction in spending
Frugality	Elicited by consumers in an initial survey in the app using three items measured on a five-point Likert scale from (Lastovicka et al., 1999)	Highly frugal individuals are likely to already have established well-functioning systems to control their spending. Hence, any intervention is unlikely to result in an additional spending reduction, whereas their less frugal counterparts are expected to show greater effects
(Variance in) Baseline spending	Retrieved total weekly spending in CHF for a total of 52 baseline weeks, through a private API by the card issuer after a consumer opted into the study (as part of the study onboarding process)	Very stable, established spending patterns may be harder to adjust and thus result in smaller effects than for individuals with greater variance in baseline spending
Reduction in spending	Retrieved individual transactions through a private API by the card issuer throughout the study period, and measured overall spending	The variable that ultimately matters for people's financial well-being is their total spending. While the interventions highlight particular categories of spending (*e.g.* exceptional transactions), consumers may also choose to adjust their spending elsewhere (*e.g.* ordinary transactions) to reach their overall goal

The invitees had to download a mobile app (available for iOS and Android, in English, German, French, and Italian), which was developed specifically for the purpose of this study, and then copy an individual user token from their invite into the app to link their credit card. They

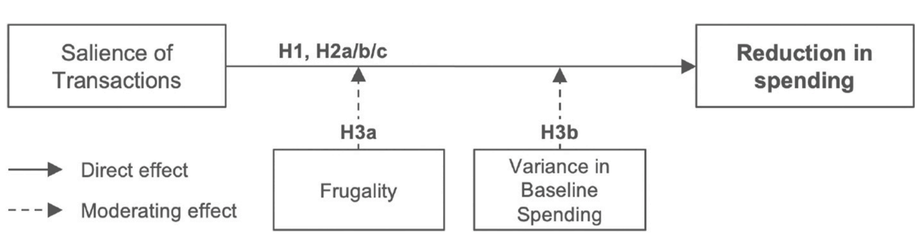

Fig. 1. Research model.

were then randomly assigned to one of the four experiment groups (see below), received instructions on how to use the app, and were asked to fill out an initial survey comprised of 19 questions that covered demographic questions, economic attitudes, the estimated share of wallet, and financial behaviors. Then, participants were asked to define a weekly spending goal, which by default was set to a reduction in spending by 15% compared to their individual baseline (see Fig. 2) – a default that we deemed attainable, but neither too loose nor too strict (Erez & Zidon, 1984). The spending goal could also be adjusted throughout the study. Since the efficacy of goal-setting had been demonstrated in prior work across various domains (*e.g.* Loock et al., 2013; Tiefenbeck, 2014; Ülkümen et al., 2008), we incorporated it in our study for all groups (including control) without further testing its effects. In the financial domain, the importance of goal-setting to achieve financial outcomes is echoed by researchers and practitioners alike, e.g. in the form of defining and adhering to spending budgets (maximum spending goals) or setting long-term goals such as retirement (Dew & Xiao, 2011; Ülkümen & Cheema, 2011). Therefore, goal-setting and goal attainment feedback were considered appropriate vehicles to deliver salience-manipulating interventions to consumers.

Our study manipulated transaction salience in three levels – from low (control group) to high: During the warmup period, the study design was identical for participants across all treatment groups (TG1, TG2, TG3). Participants in the treatment groups needed to review and perform a semantic task (transaction categorization in a binary schema as either *ordinary* or *exceptional*) throughout the study, while the control group did not have this categorization feature at any point in time. During the warmup period, we thus had two groups of participants – those in the *low* salience condition without the categorization task (control group), and those in a *medium* salience condition with the categorization task (treatment groups). When new credit card transactions were made, the treatment groups received push notifications prompting them to categorize said transactions – or, in case no new transactions were processed on a particular day, reminders were sent, *e.g.* "Please categorize your new transaction at SuperMerchant over CHF 79.50.

As well as X previous transactions" (sent at most once per day). Also, while the app provided guidelines as to how the classification scheme was intended to be used, it was up to the participants to pick whichever category they deemed most appropriate for each transaction, hence each participant may have arguably had a slightly different mental representation of the categories. In fact, this was necessary, since there are many types of purchases whose classification is ambiguous: Consider, for instance, visiting a cinema or zoo: For some consumers, these may be regular activities, for others they may be rarely occurring exceptions. After the warmup period, we introduced different variants of weekly goal attainment feedback interventions implemented as push notifications and inside the app (see Fig. 3), which were either focused on *ordinary (TG2)*, *exceptional (TG3)*, or *both* types (TG1) of purchases (all *high salience* conditions), whereas a control group received more general goal attainment feedback (*low salience*). The end of the study was marked by a final survey, which elicited feedback regarding the mobile application, perceived control over participants' spending, as well as changes in their credit card usage habits.

The statistics software R in version 3.5.1 was used to perform all the statistical tests, such as ANOVA and follow-up Tukey HSD tests, Kruskal-Wallis and Mann-Whitney tests.

5. Results and discussion

5.1. Data description and demographics

In total, 1′099 people downloaded the app and completed the onboarding process (opt-in, instructions in app, initial survey, definition of weekly spending goal). Over the period of three months, the study participants made a total of 125′846 credit card transactions amounting to CHF 5.96 million, and had a median spending of CHF 234 per week

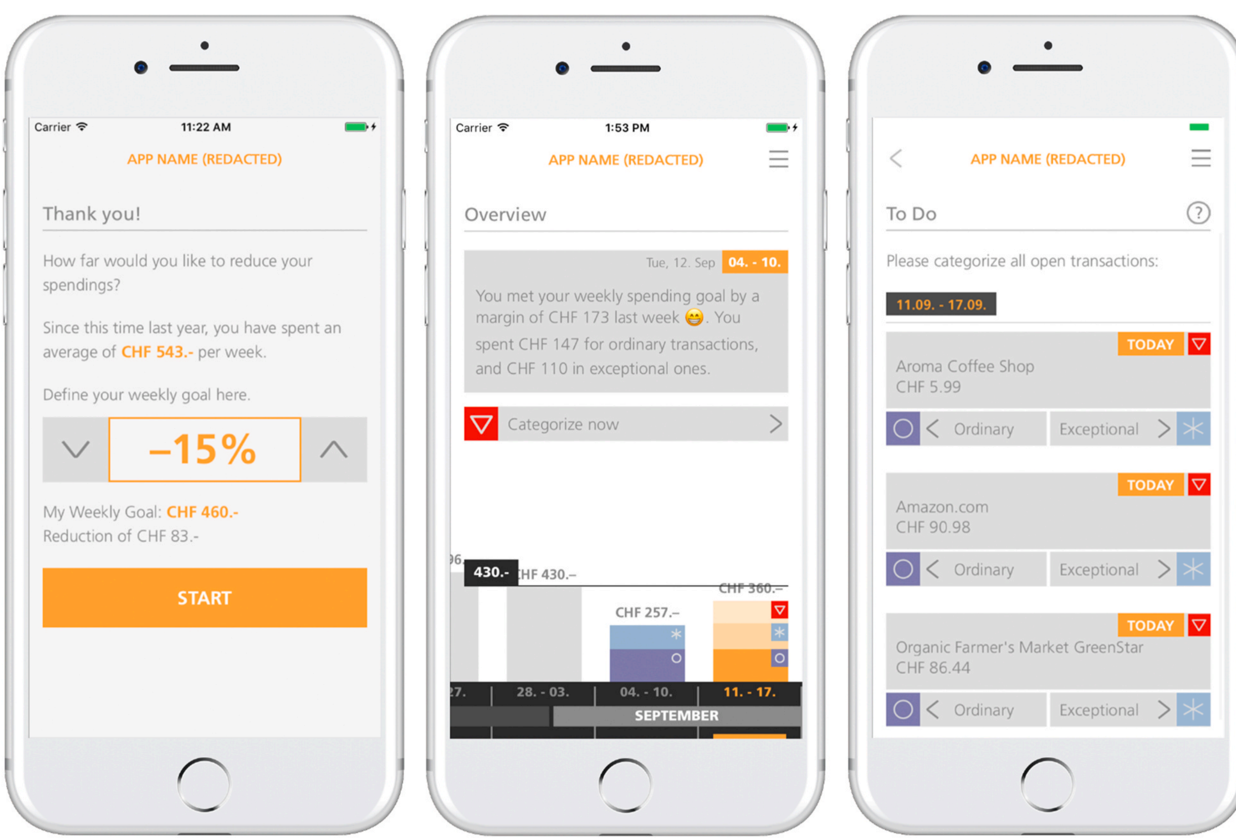

Fig. 2. Main screens of the app. i) Goal setting, ii) Dashboard with weekly spending breakdown, iii) categorization screen of individual credit card transactions.

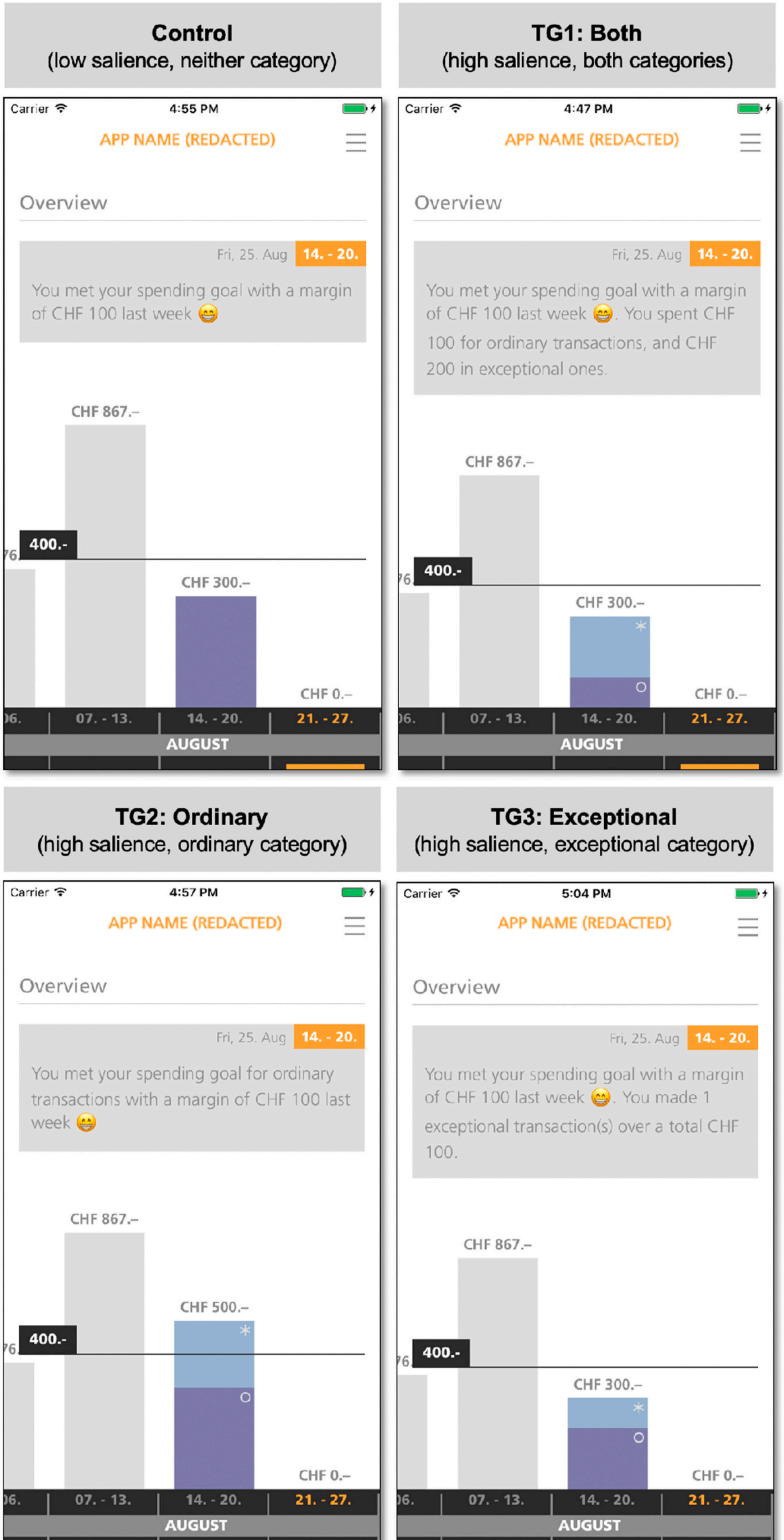

Fig. 3. Condition-specific goal attainment feedback in case the spending goals were attained. In case spending goals were exceeded, the feedback was also phrased in an encouraging way. Note that in TG2, transactions marked as exceptional did not count towards the spending goal, and only ordinary transactions were included in the goal, which is different from the other three groups where all transactions were included in the goal. This design choice mimics the way people often intuitively budget, *i.e.* including ordinary, regular transactions in their budget while granting themselves "exceptions", which was considered a more natural setup for this group than still counting exceptional transactions towards the goal. All groups received condition-specific instructions that clearly stated, which transactions counted towards the goal. The dependent variable we report in this paper (reduction of overall spending) was, of course, equal across all groups.

(mean = 391 CHF, see Fig. 4). Those participants who had the categorization feature enabled (*i.e.* all treatment groups), manually classified a total of 43′493 transactions throughout the study. Note that a sizeable share of 36.2% ($n_{transactions} = 15'733$) of categorized transactions, and 61.3% (CHF 1.21 million) of the CHF volume of categorized transactions was classified as *exceptional*. Remarkably, even though TG2 participants could have easily reached all their goals by categorizing all transactions as *exceptional* (see caption of Fig. 3), the categorization behavior was actually quite identical across all treatment groups. The tremendous share of *exceptional* transactions can, to some extent, be explained by the instructions given in the app, and by the fact that the dataset only included credit card transactions, which cannot be used for some highly ordinary payments such as rent or insurance premiums. Still, this simple descriptive data point clearly confirms what prior research has found: Such purchases, which people consider *exceptional*, actually happen with far greater frequency and volume than the term would imply (Sussman & Alter, 2012). Not fully considering such purchases in a spending budget can thus create considerable blind spots and render the entire budgeting process futile.

At the beginning of the study, the participants were an average of 39.3 years old (SD = 12.2 years), with 22.4% being females, and they had a weekly baseline spending of M = 437.56 CHF (SD = 304.48 CHF). In terms of age, baseline spending, and prior customer relationship length, our sample was highly representative of the population of 31′000 invitees. Compared to the overall Swiss population, our participants' median age of 38.0 years was 4.0 years younger, which is not a big gap considering the digital nature of our study. There was, however, a gender bias ($p < .001$***), as a result of i) the pool of invitees already exhibiting a gender bias (34.6% females), and ii) the prizes raffled out in the invitation (tech gadgets) arguably being more appealing to the male demographic, which intensified the bias. However, there is no reason to believe that the interventions would have gender-specific effects. In addition, finance apps generally still have more male than female users at this point in time, which may have been another reason for the disproportionately high share of males amongst the sample.

As Fig. 5 displays the distribution of household incomes (self-reported in the initial survey) and credit card limits. Note that in the Swiss market, the vast majority of credit card customers pay back the monthly bill in full every month, and do not get close to reaching their monthly credit card limit. Carrying over credit card debt into the following month is often not even allowed from a contractual point of view. Card limits do therefore not impose such a crucial limitation on consumption as may be the case in other markets. Instead, credit cards are used in way very comparable to debit cards.

Because the study required active participation over three months, it was not surprising that only a minority of the initial participants completed the study. In total, 259 people completed the study including the final survey. In addition, we observed that a number of people actually only used their cards rather sporadically, *e.g.* when making online purchases. Of course, our study aimed at helping people reduce their spending by implementing interventions regarding their primary means of payment, and not just one that is frequently complemented by

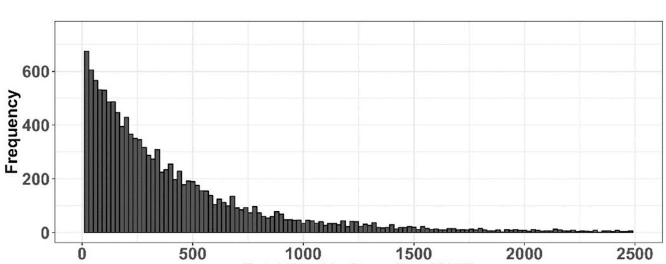

Fig. 4. Distribution of total weekly spending throughout the study, after removing weeks with spending >2′500 CHF (n = 14′287 weekly budgets).

other means of payment such as other cards or cash. Therefore, the analysis in the following was conducted for those 98 people who used the linked credit card as their primary payment method for online and offline purchases. This was achieved by excluding those people whose mean number of weekly transactions was smaller than the average across the sample (8.9 transactions per week). A recent, independent study of the Swiss payment market found that people made an average of 11.2 financial transactions per week across all payment methods, and including cash payments (Swiss National Bank, 2017). The included individuals also frequently used their cards in grocery stores (22.1% of their transactions). Therefore, for our sample of 98 individuals who used the linked credit card for at least 8.9 weekly transactions, we arguably had a near-complete picture regarding their financial transactions. Table 2 summarizes key demographics of the users included in the analysis.

Checks for differential attrition did not reveal any significant differences between the 1′099 initial study participants and those 98 individuals included in the analysis, in terms of age ($p = .308$), gender ($p = .988$), or the available economic attitudes frugality ($p = .462$), impulsiveness ($p = .679$), and self-efficacy ($p = .912$, all of which were elicited in the initial survey). The same was true between the 1′099 initial participants and those 259 individuals who completed the study. Unsurprisingly, the 98 analyzed participants who used their credit card on a regular basis during the study period also had a higher baseline spending (M = 490.84 CHF, SD = 235.36 CHF) than the total sample of 1′099 participants (M = 432.35 CHF, 310.02 CHF), which also included sporadic users ($t(132) = 2.28$, $p = .025$*).

5.2. Hypothesis tests

5.2.1. The effect of salience-increasing interventions on credit card spending

The dependent variable of interest, weekly spending, was not normally distributed, as a simple visual inspection of the distribution made evident, and which was also reflected in the Shapiro-Wilk ($W = 0.620$, $p < .001$***) and Kolmogorov-Smirnov ($D = 0.493$, $p < .001$***) normality tests returning significant deviations from a normal distribution. Since the normality assumptions of standard parametric tests (*i. e.* analysis of variance with follow-up pairwise Tukey HSD tests) were not fulfilled in this context, we report the results of their non-parametric equivalents (*i.e.* Kruskal-Wallis rank-sum tests and follow-up pairwise Wilcoxon rank sum tests), which do not make the normality assumption. While we report both means and medians throughout this section, we offer interpretations based on the medians, which are both the better measure of central tendency in skewed distributions, and in our case also usually more conservative. Parametric tests using a log-transformed dependent variable return very similar results to those reported below.

Across the sample, which consisted of n = 1′274 weekly budgets, the median normalized spending was 0.769 (mean = 1.087, sd = 1.211). The dependent variable was normalized using the mean (not median) baseline spending. As the mean normalized spending of 1.087 indicates, during the three-month study period, all participants spent on average 8.7% more than they did in the preceding 52 weeks. This increase is most likely an effect of seasonality: as described above, there were no major holiday seasons like summer holidays in the study period, and people may have started buying Christmas presents or trips for the end-of-the year holidays during the study period. However, as the detailed experiment results in Table 3 illustrate, the true effect of seasonality was actually stronger than the above number suggests: The Control group actually spent 16.7% more than they did in the year before.

The global Kruskal-Wallis test revealed significant global differences ($p = .026$*), and post-hoc Wilcoxon tests confirmed that TG1 exhibited a significantly lower normalized spending than the control group: The median for the Control group was 0.789, while that of TG1 was 0.658, a statistically and practically significant 13.1 percentage points reduction in spending ($p = .008$**). The means of the two groups draw an even more dramatic picture, with TG1 being the only group apparently

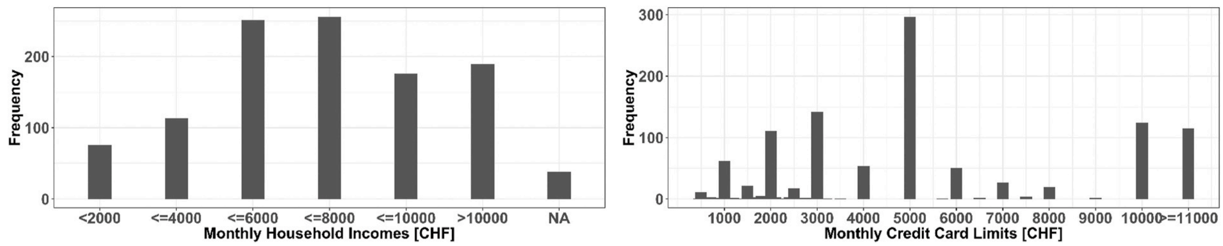

Fig. 5. Distribution of study participants' household income and credit card limits (N = 1′099 participants).

Table 2
Sample description of participants included in the analysis. N=Sample size, M = Mean weekly baseline spending in CHF, SD=Standard deviation in CHF.

		Weekly Baseline Spending [CHF]	
Group	Users (N)	M	SD
Control	16	453.76	233.11
TG1 – Both categories	28	473.34	242.19
TG2 – Ordinary	23	476.96	223.74
TG3 – Exceptional	31	536.07	243.22
All	**98**	**490.84**	**235.36**

Table 3
Spending behavior across groups throughout the study period. TG1 reduced their spending by 13.1 percentage points compared to the Control group. Note that the difference in location parameters (location shift) does not estimate the difference in medians, but rather the median of the difference between a sample from one group and a sample from the other.

Group	Wilcoxon rank sum test (Group - Control)		Normalized spending		
	Statistic and p-value	Location Shift (95% CI)	Median	Mean	SD
Control n = 208	–	–	0.789	1.167	1.229
TG1 – Both categories n = 364	W = 42′918 **p = .008****	.134 [0.034–0.230]	**0.658**	0.950	0.902
TG2 – Ordinary n = 299	W = 31′822 p = .655	.024 [−0.08 to 0.129]	0.828	1.117	1.349
TG3 – Exceptional n = 403	W = 43′453 p = .456	.042 [−0.07 to 0.146]	0.794	1.151	1.327
Kruskal-Wallis $\chi2$	χ^2 (3) = 9.302 p = 0.026*				

fighting the seasonal uptake in spending, and achieving a difference of 21.7 percentage points between TG1 and the Control group. Since the distribution of our dependent variable is quite heavily right-skewed (skewness = 4.45), medians are a better measure of the distribution's central tendency than the means. Nonetheless, while TG1 participants still had some weeks with spending volumes way above their average, the fact that there is such a stark difference in means across groups suggests that the extreme weeks of TG1 participants were less pronounced than those of the Control group.

Our interventions in TG1, with weekly feedback that highlighted both ordinary and exceptional spending, was thus highly effective in helping participants reduce their spending. At the same time, the interventions delivered to participants in TG2 and TG3 were not effective and thus did not yield any statistically significant differences from the Control group. Our first hypothesis, H2a, is thus supported, whereas H2b and H2c are rejected.

This effect persists even when including the sporadic credit card users in our sample, who also completed the study, albeit in weakened form in terms of statistical significance and effect size. Across all 3′367 weekly budgets in the dataset, the global Kruskal-Wallis test confirms the presence of significant global differences (χ^2 (3) = 7.877, p = 0.049*), with the difference between TG1 and the Control group being marginally significant (p = .100) with a 6.0 percentage point difference in medians ($Median_{Control}$ = 0.676, $Median_{TG1}$ = 0.616). The other two groups are not significantly different from the Control, either ($Median_{TG2}$ = 0.640, $Median_{TG2}$ = 0.697, $p_{TG2\text{-}Control}$=.548, $p_{TG3\text{-}Control}$=.563).

It appears that for consumers to adjust their spending, they really required a complete picture of their spending, including both *ordinary* and *exceptional* transactions, whereas focusing on either category alone did not suffice. Based on the reasoning brought forward by prior work that states that consumers particularly underestimate the frequency and volume of *exceptional* transactions, it was conceivable that simply highlighting those exceptions, and thus debiasing consumers (as trialed with TG3), would be effective. However, as this study has shown, the lack of additional information regarding *ordinary* spending renders the feedback incomplete and thus makes it harder to learn from. Finally, TG2 mimics the naïve budgeting techniques many consumers apply by focusing on regular, ordinary spending while leaving room for exceptions, *i.e.* not budgeting for them explicitly. Consumers do so when they mentally fail to assign a broader category to an irregular, unusual transaction. Because of this, they are not affected by mental accounting effects such as depleting per-category budgets that would render them more price-sensitive for future spending in said category. Consumers may however also do so when setting up an explicit budgeting system, *e.g.* by planning only for regular, recurring purchases, and/or by leaving them too much slack for "miscellaneous" transactions that do not fit into any other category. The fact that TG2 did not yield a significant reduction in spending is alarming, since the data suggests that following such a simple, intuitive budgeting approach may not yield superior results than merely checking one's overall weekly spending, even though this needs more exploration in future research.

The effect of our interventions is comparable in size to what behavioral interventions have yielded in similar field studies in the financial domain. For example, Karlan et al. implemented reminders to nudge people towards putting money into a savings account, achieving a 6% increase in savings (11% when coupled with financial incentives) (Karlan, McConnell, Mullainathan, & Zinman, 2016). Benartzi found that users of a mobile app that aggregated multiple bank accounts and provided a spending analysis by categories reduced spent 15.7% less than their peers without the same app, even though they theoretically would have had the same information available on a website (Benartzi, 2017).

In the study at hand, we tried to design around possible selection effects by designing a meaningful control group within the app, against which we could compare the efficacy of our interventions. Aside from such behavioral interventions like the above-mentioned ones, an even more relevant frame of reference may be the premium that people pay when switching from cash to lower-salience payment methods like credit cards. As discussed before, a number of studies have investigated this very effect and found credit card premiums in the order of magnitude of 10% to more than 100% (*e.g.* Prelec & Simester, 2001; Raghubir & Srivastava, 2008; Soman, 2003). To the best of our knowledge, though, all studies that found very high credit card premiums prompted participants to estimate their willingness to pay, or conducted situated studies (*e.g.* using auctions for sports tickets) without serious financial consequences for participants. While we do not put in question the findings of these studies with remarkable and relevant results, the same participants would probably not consistently pay twice as much for most of their purchases over longer periods of time, even if they had the financial capacity to do so. We believe for the true long-term credit card premium to be in the range of 15%–30%, as studies investigating the differential effect of payment methods when purchasing everyday items and groceries have found (Raghubir & Srivastava, 2008; Soman, 2003). In light of this frame of reference, the interventions tested in the study at hand are not only statistically significant, but also practically relevant, since they effectively counter-balance a large portion of the credit card premium.

5.2.2. Hedonic and utilitarian spending

Further, we wanted to investigate which type of consumption was affected most by our interventions. Consumer research broadly classifies purchases into two categories, hedonic and utilitarian ones (*e.g.* Alba & Williams, 2013; Strahilevitz & Myers, 1998). Hedonic consumption is defined as being mainly motivated by the desire for sensual pleasure and fun. In contrast, the main motivator behind utilitarian purchases are meeting basic needs or accomplishing a particular, functional task (Strahilevitz & Myers, 1998). Therefore, the latter category was expected to be less easily adjustable than hedonic purchases. While many consumers may still have room for reducing their spending even in utilitarian categories, *e.g.* by buying no-name instead of brand clothing, we expected that similar to what others have found (e.g. Benartzi, 2017), discretionary, hedonic categories would be the primary source of spending reduction. To investigate this, our research team first set out to classify the top 100 merchant category codes[a] by volume as either *utilitarian, hedonic,* or *inconclusive*. Two raters of our research team independently coded the codes following the same category definitions, and iterated until consensus was reached. Despite this consensus in the coding process, the categorization is not perfect: Some category codes are entirely ambiguous and were thus marked as *inconclusive*. Others are only be approximately correct, *e.g.* the category code "supermarkets" was categorized as *utilitarian*, which we believe to be accurate for most items of most baskets, but buying supplies for a party in a supermarket would hardly be a utilitarian purchase. Other examples of utilitarian categories include clothing, transportation, or pharmacies. On the other hand, some categories such as jewelry stores, are less ambiguous. Other examples for hedonic categories include luxury clothing, movie theaters, and travel agencies.

With this in mind, we analyzed whether primarily *utilitarian* or *hedonic* categories were adjusted. We found that TG1 participants did indeed decrease their share of spending in hedonic categories. This is true both in absolute terms as well as when normalized analogously to the main dependent variable, by the mean individual baseline spending in the same hedonic categories. A Kruskal-Wallis test (χ^2 (3) = 9.035, p = .029*, $\text{median}_{\text{Control}}$=.543, $\text{median}_{\text{TG1}}$=.379, $\text{median}_{\text{TG2}}$=.414, $\text{median}_{\text{TG3}}$=.435), and follow-up pairwise Wilcoxon tests reveal significant differences between Control and TG1 (p = .022*, other pairwise comparisons not significant), indicating that TG1 participants adjusted their hedonic spending compared to their prior spending in the same categories, which stands in contrast to the behavior in the other groups. Contrary to this, there was absolutely no significant difference in *utilitarian* spending across the four groups (Kruskal-Wallis χ^2 (3)=.166, p = .983, $\text{median}_{\text{Control}}$ = 77.3, $\text{median}_{\text{TG1}}$ = 80.7, $\text{median}_{\text{TG2}}$ = 85.0, $\text{median}_{\text{TG3}}$ = 79.1). Normalizing this variable by the individual baseline spending in utilitarian categories yielded very similar results. Importantly, this also shows that TG1 participants did not substitute their credit cards with other forms of payment, since this would have also reflected in a considerable decrease in utilitarian purchases.

5.2.3. Moderating variables

Furthermore, we found that two variables significantly moderated the main effect: frugality and the variance of baseline spending. To test the corresponding hypotheses, we conducted ANOVA tests with the log-transformed normalized spending as dependent variable.

We found that the interventions worked best for individuals of medium frugality, and had less of an effect for low- and high-frugality individuals, as can be seen in Fig. 6 (F (7) = 25.6, p = .003**). Note that the graph includes only few high-frugality individuals in the Control group, which may be the reason for the ascending slope. The fact that high-frugality participants did not react to our interventions makes intuitive sense, since they are more likely to already have had effective money management in place, thus rendering any additional measures obsolete. The fact that the interventions also had less of an effect amongst low-frugality suggests that at least some level of alignment with an individual's attitude is required for interventions to be effective, similar to how goal-setting is effective only if people accept a given goal (Erez & Zidon, 1984). An alternative interpretation is that low-frugality individuals may have been particularly thrilled about having a digital companion, and delegated the controlling of their spending to the app to a greater degree than others. In conclusion, frugality appears to non-linearly moderate the effect of salience-increasing interventions on spending reduction, with medium-frugality individuals being the most responsive.

The regularity of participants' baseline was also found to be a significant moderator (F (6) = 22.9, p = .004**). The interventions had the largest effect on individuals with high variances in their baseline spending, which in combination with the comparatively low standard deviation amongst TG1 participants suggests that the designed interventions may have helped consumers stabilize their spending behavior over time. There are only few participants in the Control group with medium baseline variance, hence the true slope may be less steep. Nonetheless, this moderating effect is aligned with TG1 participants primarily adjusting their hedonic, exceptional transactions, which contribute a lot of variance in their weekly spending. Our data thus supports hypotheses H3a and H3b.

5.2.4. The categorization feature in isolation did not impact spending

In addition, we checked whether merely categorizing transactions in the binary scheme we offered already led to a spending reduction within a two-week warmup period. During those two weeks, there were only two groups, categorizers (TG1, TG2, TG3) and non-categorizers (Control), all without goal attainment feedback.

No significant differences could be found between these two groups (W = 16′602, p = .884), indicating that this highly disaggregated semantic task alone, i.e. the rehearsal of an individual transaction, was not sufficient to nudge credit card users towards spending less. Instead, both the transaction-level semantic task and the aggregated feedback were necessary for participants to reduce their spending. In other domains, such as energy or water usage, people do not usually have great transparency over their real-time consumption, which is why highly disaggregated feedback alone can yield behavior changes. In contrast, people are already comparatively aware of their spending at the point in time when a transaction occurs, but a certain level of aggregation of one's consumption appears to be necessary to provide meaningful behavioral guidelines. Therefore, hypothesis H1 is supported.

[a] ISO 18245 defines a list of merchant category codes (MCC), which were made available to us together with the other transaction data through an API by the credit card issuer. For example, there is one MCC (5411) for supermarkets, others for gas stations, airlines, restaurants, etc.

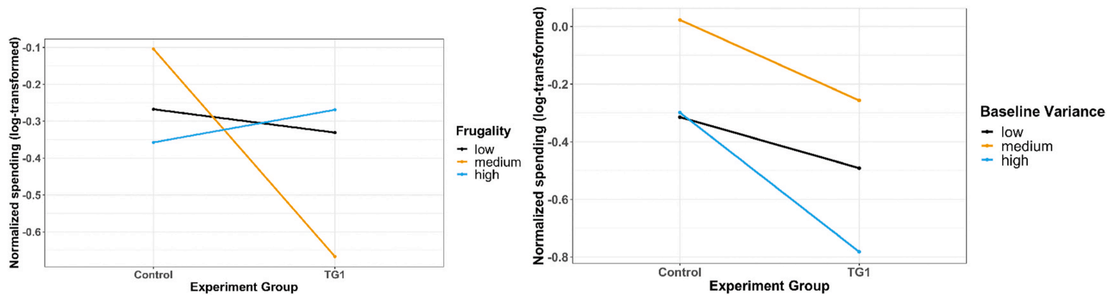

Fig. 6. Checks for moderating effects: Both frugality and the variance of baseline spending were found to be significant moderators.

6. Conclusion

6.1. Findings and contributions

This work is one of the first to illustrate and empirically test how technology-mediated interventions can impact consumer spending. We found that by increasing the credit card transaction salience while explicitly highlighting both *ordinary* purchases as well as those that are considered *exceptional*, consumers reduced their spending compared to a control group. We increased salience by prompting participants to perform a simple semantic task for every transaction, and providing weekly spending feedback. Literature suggests that the frequency and volume of *exceptional* transactions are often underestimated, which makes them particularly hard to manage and learn from (Sussman & Alter, 2012). Interestingly, we found that providing spending feedback that particularly highlighted the frequency and volume of *exceptional* transactions in a given week (TG3) did not help participants reduce their spending. Instead, more holistic feedback including both ordinary and exceptional purchases was necessary. Similarly, asking consumers to review and categorize individual transactions in isolation, *i.e.* without weekly spending feedback, also did not have an effect.

This work thus illustrates how technology can be used to counterbalance the decreased pain of payment that cashless transactions exhibit and thereby help consumers control their spending better. This is crucial when considering the global shift towards cashless economies: Scandinavian countries and China are famous for the almost complete absence of cash payments, and other countries are likely to follow suit. This development requires better control mechanisms for consumers to counterbalance the approximately 15%–30% people tend to spend more with cashless payment methods (see *e.g.* Prelec & Simester, 2001; Raghubir & Srivastava, 2008).

With this work, we devised and empirically tested a strategy for how consumer decision-making can be assisted, based on prior research on consumer behavior and behavioral economics. This work thus contributes to the body of academic knowledge by providing a blueprint for how technology can be used to nudge consumer behavior even in a domain such as personal finance, with large individual differences in behavior and well-rehearsed behavioral patterns that may seem hard to break without highly disruptive interventions. We tested our interventions on credit cards, but we have no reason to believe that they would not work for other cashless payment methods. To the contrary, in an ideal world, such interventions would cover not just one, but all payment forms available to an individual. For the purpose of our study, we addressed this issue of focusing on those customers that used their credit card as main form of payment, *i.e.* those of whom an almost complete view of financial transactions was available. Contrary to many other studies in this domain, this work was not conducted in the lab (*e.g.* Chatterjee & Rose, 2012; Raghubir & Srivastava, 2008), but demonstrated how smartphones can be used to deliver behavioral interventions in a scalable manner. In addition, to the best of our knowledge, this is one of the first studies exploring how technology can be leveraged to overcome some of the cognitive biases consumers are subject to in the financial domain – an intersection of two literature streams that would be worth exploring further (Zhang & Sussman, 2018).

Finally, this work discussed the so-called credit card premium, and offered rare empirical insights into the extent to which it can be overcome by using unobtrusive behavioral interventions. We explicitly decided for a field experiment lasting several months in order to be able to paint a rich picture of consumer behavior, were able to test multiple intervention strategies and discussed the role of goal-setting, moderating variables such as frugality, and investigated how people decreased their spending. We were also able to reproduce prior research concluding that *exceptional* transactions are indeed anything but that, and thus need to be treated as first-class citizens in any household budget (Sussman & Alter, 2012).

6.2. Practical implications

Consumer banking generally shifts towards greater automation and removing friction in financial transactions, which becomes obvious when considering the recent years' developments in self-service banking, in particular online and mobile banking. While this shift certainly makes sense from the perspective of financial service providers and merchants, a more differentiated picture arises from the perspective of consumers. While added convenience in banking and payments appears highly desirable, it does come at a cost (*e.g.* Prelec & Loewenstein, 1998; Raghubir & Srivastava, 2008; Soman, 2003). Modern online and mobile banking systems routinely provide consumers with an overview of per-category spending, using machine learning techniques to automatically attribute every purchase to a particular category. While this undoubtedly is a desirable, transparency-increasing feature to have, it may give consumers a false sense of control over their spending. As this study has shown, simply providing an aggregated overview of spending does not suffice when it comes to reducing one's spending.

Financial services providers should focus not only on removing friction and increasing credit card revenues as much as possible. Consumers will likely come to expect offerings that help them more effectively manage their financial lives. Hence, behavioral support systems like the one developed for the purpose of this study may also be a wise item to include in the digital roadmaps of financial service providers.

From the perspective of consumers, an important implication of this study is that reaching financial short-term goals will much likelier work if *both ordinary* and *exceptional* transactions are explicitly accounted for. Across the participants of this study, 36.2% of all transactions and 61.3% of the volume were marked as *exceptional*. While we acknowledge that this number overestimates the true figure (*e.g.* because some high-value, ordinary transactions such as accommodation and insurances are not paid via credit card), it is still a lot higher than the terminology implies. Still, providing feedback focused on explicitly debiasing people regarding their underestimation of *exceptional* transactions was not effective in helping them decrease their overall spending, arguably because important information on the remainder of their spending, the ordinary spending, was not as readily available.

It is also crucial for households to understand their financial needs, spending patterns, and parameters that can be (easily) adjusted in times of financial distress. Therefore, being able to effectively manage spending across all payment forms, including *exceptional* purchases, is of great importance. Additionally, by doing so, healthy financial spending habits can be formed, which is especially important in retirement or phases with decreased income – and this is true for households across all income classes. In fact, comparatively income-rich households may not feel the consequences of over-spending immediately and be able to compensate for it more easily by either reducing their saving rate, taking a loan, or by slightly reducing their consumption down the road, *e.g.* by traveling a bit less expensively or less often. However, during retirement, even well-off households may then have to acquire greater budgetary discipline and may have to break with spending patterns established over decades.

6.3. Limitations and future work

This study was conducted as a large-scale field experiment in cooperation with a credit card issuer, who kindly offered us the opportunity to conduct research with their customers. While this setup provided us with unique access to credit card transaction data, it also incurred some limitations. First, due to the field experiment setup, not all variables were fully under our control. For example, it is possible that some participants switched to cash or other payment methods for parts of their transactions in order to reach their self-defined goals. However, all experiment groups would be equally affected by Hawthorne effects so that our results are not expected to be biased due to this. In addition, we did not find that participants reduced their utilitarian purchases such as supermarket transactions, which suggests that they continued to use their cards on a regular basis. In future research, it might be worth estimating the share of wallet both prior to and during the experiment, *e.g.* using an approach based on the work by Chen & Steckel (2012). Moreover, even though we were able to work with a quite unique dataset and carefully designed the study to avoid biases, we would have been able to capture the intricacies of the financial behavior of the Swiss population even better with a larger sample size. Therefore, repeating the same study with more participants would be an intriguing endeavor.

In addition, varying the timing of feedback interventions would be an interesting area of future research. While our research design and technical constraints only allowed for ex-post feedback, we would be particularly interested in exploring the efficacy of predictive interventions, *e.g.* by informing consumers of their prior grocery spending when they enter a supermarket, or even a few hours before weekly grocery trips are likely to occur.

Finally, there is a lot of potential to further personalize the feedback interventions to achieve an overall reduction in spending. Machine learning techniques or simple customer personas could be used to further inform the design of more fine-granular feedback, and the notion of pain of payment should be understood as a tool used to optimize consumers' utility by systematically increasing the pain of payment for day-to-day expenditures somebody wishes to control better, while decreasing the pain of payment for purchases one seeks to enjoy (Ariely, 2013). After all, simply minimizing consumption is usually not a recipe for happiness. Different streams of research in behavioral economics and happiness research suggest that happiness can be drawn even from enjoying mundane, ordinary experiences (*e.g.* Quoidbach, Dunn, Petrides, & Mikolajczak, 2010). At the same time, collecting highly memorable experiences can increase utility, which is especially true for young people (Bhattacharjee & Mogilner, 2013; Zauberman, Ratner, & Kim, 2008), and it appears that money can buy happiness if it is spent in ways that fit our personality (Dunn, Gilbert, & Wilson, 2011; Matz, Gladstone, & Stillwell, 2016). Future work should thus explore higher-precision interventions that take individual preferences into account and specifically address transactions that one would like to avoid or reduce in terms of frequency or amount.

CRediT authorship contribution statement

Johannes Huebner: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Elgar Fleisch:** Supervision, Resources, Writing - review & editing, Funding acquisition. **Alexander Ilic:** Conceptualization, Methodology, Validation, Writing - review & editing, Project administration.

Acknowledgements

We thank our project partner, Viseca Card Services SA, and particularly Mr. Tony Weber, Mr. Herbert Bucheli, Mr. Bedrija Hamza, Mr. Daniel Bernard, Ms. Raquel Fandos, Ms. Diana Liviero and Ms. Michaela Simeon, for the unique opportunity to conduct research with existing credit card customers, and for their continued support in designing, developing, and publishing the study and mobile application.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2020.106504.

References

Akbaş, M., Ariely, D., Robalino, D. A., & Weber, M. (2016). *How to help the poor to save a bit: Evidence from a field experiment in Kenya*. IZA Discussion Paper.
Alba, J. W., & Williams, E. F. (2013). Pleasure principles: A review of research on hedonic consumption. *Journal of Consumer Psychology, 23*(1), 2–18.
Ariely, D. (2013). *The pain of paying*. Retrieved from http://danariely.com/2013/02/05/the-pain-of-paying/.
Batiz-Lazo, B. (2016). *Are we heading towards a cashless economy?*. Retrieved from https://www.weforum.org/agenda/2016/01/are-we-heading-towards-a-cashless-economy/.
Benartzi, S. (2017). *How digital tools and behavioral economics will save retirement*. Harvard Business Review. Retrieved from https://hbr.org/2017/12/how-digital-tools-and-behavioral-economics-will-save-retirement.
Berman, J. Z., Tran, A. T. K., Lynch, J. G., Jr., & Zauberman, G. (2016). Expense neglect in forecasting personal finances. *Journal of Marketing Research, 53*(4), 535–550.
Bhattacharjee, A., & Mogilner, C. (2013). Happiness from ordinary and extraordinary experiences. *Journal of Consumer Research, 41*(1).
Capgemini. (2019). *World payments report 2019*. Retrieved from https://worldpaymentsreport.com/.
Chai, A., Rohde, N., & Silber, J. (2015). Measuring the diversity of household spending patterns. *Journal of Economic Surveys, 29*(3), 423–440.
Chatterjee, P., & Rose, R. L. (2012). Do payment mechanisms change the way consumers perceive products? *Journal of Consumer Research, 38*(6), 1129–1139.
Chen, Y., & Steckel, J. H. (2012). Modeling credit card share of wallet: Solving the incomplete information problem. *Journal of Marketing Research, 49*(5).
Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684.
Dew, J., & Xiao, J. J. (2011). The financial management behavior scale: Development and validation. *Journal of Financial Counseling and Planning, 22*(1), 43.
Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology, 33*(1), 264–277.
Dunn, E. W., Gilbert, D. T., & Wilson, T. D. (2011). If money doesn't make you happy, then you probably aren't spending it right. *Journal of Consumer Psychology, 21*(2), 115–125.
Erez, M., & Zidon, I. (1984). Effect of goal acceptance on the relationship of goal difficulty to performance. *Journal of Applied Psychology, 69*(1), 69.
Feinberg, R. A. (1986). Credit cards as spending facilitating stimuli: A conditioning interpretation. *Journal of Consumer Research, 13*(3), 348–356.
Gourville, J. T., & Soman, D. (1998). Payment depreciation: The behavioral effects of temporally separating payments from consumption. *Journal of Consumer Research, 25*(2), 160–174.
Heath, C., & Soll, J. B. (1996). Mental budgeting and consumer decisions. *Journal of Consumer Research, 23*(1), 40–52.
Hilgert, M. A., Hogarth, J. M., & Beverly, S. G. (2003). Household financial management: The connection between knowledge and behavior. *Federal Reserve Bulletin, 89*, 309.
Hodson, R., Dwyer, R. E., & Neilson, L. A. (2014). Credit card blues: The middle class and the hidden costs of easy credit. *The Sociological Quarterly, 55*(2), 315–340.
Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science, 62*(12).
Kaye, J., McCuistion, M., Gulotta, R., & Shamma, D. (2014). Money talks: Tracking personal finances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 521–530).

Kestner, J., Leithinger, D., Jung, J., & Petersen, M. (2009). Proverbial wallet: Tangible interface for financial awareness. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (pp. 55–56).

Kivetz, R., & Simonson, I. (2000). The effects of incomplete information on consumer choice. *Journal of Marketing Research, 37*(4), 427–448.

Lastovicka, J. L., Bettencourt, L. A., Hughner, R. S., & Kuntze, R. J. (1999). Lifestyle of the tight and frugal: Theory and measurement. *Journal of Consumer Research, 26*(1), 85–98.

Lawson, T. (2013). Modelling household spending using a random assignment scheme. *International Journal of Microsimulation, 6*(2), 56–75.

Lo, H.-Y., & Harvey, N. (2011). Shopping without pain: Compulsive buying and the effects of credit card availability in Europe and the Far East. *Journal of Economic Psychology, 32*(1), 79–92.

Loock, C.-M., Staake, T., & Thiesse, F. (2013). Motivating energy-efficient behavior with green is: An investigation of goal setting and the role of defaults. *MIS Quarterly, 37*(4).

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2016). Money buys happiness when spending fits our personality. *Psychological Science, 27*(5), 715–725.

New Deal Design. (2016). *Scrip - the tangible future of digital cash*. Retrieved from http://newdealdesign.com/work/scrip.

NFCC. (2016). *The 2016 consumer financial literacy survey*. Retrieved from https://nfcc.org/wp-content/uploads/2016/04/NFCC_BECU_2016-FLS_datasheet-with-key-findings_041516.pdf.

Norvilitis, J. M., Merwin, M. M., Osberg, T. M., Roehling, P. V., Young, P., & Kamas, M. M. (2006). Personality factors, money attitudes, financial knowledge, and credit-card debt in college students. *Journal of Applied Social Psychology, 36*(6), 1395–1413.

Pratz, A., Bloos, J.-W., Engebretsen, O., & Gawinecki, M. (2013). *European payment strategy report - winning the growth challenge in payments*. Retrieved from https://www.atkearney.com/documents/10192/1448080/Winning+the+Growth+Challenge+in+Payments.pdf/b9da93a5-9687-419e-b166-0b25daf585ff.

Prelec, D., & Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. *Marketing Science, 17*(1), 4–28.

Prelec, D., & Simester, D. (2001). Always leave home without it: A further investigation of the credit-card effect on willingness to pay. *Marketing Letters, 12*(1), 5–12.

ProQuest. (2017). Credit card debt outstanding in the United States from 2000 to 2015, by type of credit card. In *Statistical abstract of the United States 2017* (p. 762). Retrieved from https://www.statista.com/statistics/245393/credit-card-debt-outstanding-in-the-united-states-by-credit-card-type/.

Quoidbach, J., Dunn, E. W., Petrides, K. V., & Mikolajczak, M. (2010). Money giveth, money taketh away: The dual effect of wealth on happiness. *Psychological Science, 21*(6), 759–763.

Raconteur. (2016). *Future of payments*. Retrieved from https://www.raconteur.net/future-of-payments-2016.

Raghubir, P., & Srivastava, J. (2008). Monopoly money: The effect of payment coupling and form on spending behavior. *Applied Journal of Experimental Psychology, 14*(3), 213.

Reville, P. (2019). *ATM banking: It's not just about cash withdrawal anymore*. Retrieved from https://www.mercatoradvisorygroup.com/Reports/ATM-Banking-It's-Not-Just-About-Cash-Withdrawal-Anymore/.

Rick, S. I., Cryder, C. E., & Loewenstein, G. (2008). Tightwads and spendthrifts. *Journal of Consumer Research, 34*(6), 767–782.

Runnemark, E., Hedman, J., & Xiao, X. (2015). Do consumers pay more using debit cards than cash? *Electronic Commerce Research and Applications, 14*(5), 285–291.

Soman, D. (2001). Effects of payment mechanism on spending behavior: The role of rehearsal and immediacy of payments. *Journal of Consumer Research, 27*(4), 460–474.

Soman, D. (2003). The effect of payment transparency on consumption: Quasi-experiments from the field. *Marketing Letters, 14*(3), 173–183.

Soman, D., & Lam, V. M. W. (2002). The effects of prior spending on future spending decisions: The role of acquisition liabilities and payments. *Marketing Letters, 13*(4), 359–372.

Soster, R. L., Gershoff, A. D., & Bearden, W. O. (2014). The bottom dollar effect: The influence of spending to zero on pain of payment and satisfaction. *Journal of Consumer Research, 41*(3), 656–677.

Strahilevitz, M., & Myers, J. G. (1998). Donations to charity as purchase Incentives : How well they work may depend on what you are trying to sell. *Journal of Consumer Research, 24*(4), 434–446.

Sussman, A. B., & Alter, A. L. (2012). The exception is the rule: Underestimating and overspending on exceptional expenses. *Journal of Consumer Research, 39*(4), 800–814.

Swiss National Bank. (2017). *Survey on payment methods*. Retrieved from https://www.snb.ch/en/mmr/reference/paytrans_survey_report_2017/source/paytrans_survey_report_2017.en.pdf.

Taubinsky, D., & Rees-Jones, A. (2017). Attention variation and welfare: Theory and evidence from a tax salience experiment. *The Review of Economic Studies, 85*(4), 2462–2496.

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy, 19*(4), 356–360.

Thompson, C. (2014). *When pedestrians ruled the streets*. Smithsonian. Retrieved from https://www.smithsonianmag.com/innovation/when-pedestrians-ruled-streets-180953396/.

Tiefenbeck, V. (2014). *Behavioral interventions to reduce residential energy and water consumption*. Retrieved from http://e-citations.ethbib.ethz.ch/view/pub:143277.

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making* (pp. 141–162). Springer.

Ülkümen, G., & Cheema, A. (2011). Framing goals to influence personal savings: The role of specificity and construal level. *Journal of Marketing Research, 48*(6), 958–969.

Ülkümen, G., Thomas, M., & Morwitz, V. G. (2008). Will I spend more in 12 months or a year? The effect of ease of estimation and confidence on budget estimates. *Journal of Consumer Research, 35*(2).

US Census Bureau. (2017). *Income and poverty in the United States: 2016*. Retrieved from https://www.census.gov/library/publications/2017/demo/p60-259.html.

Zauberman, G., Ratner, R. K., & Kim, B. K. (2008). Memories as assets: Strategic memory protection in choice over time. *Journal of Consumer Research, 35*(5), 715–728.

Zhang, C. Y., & Sussman, A. B. (2018). Perspectives on mental accounting: An exploration of budgeting and investing. *Financial Planning Review, 1*(1–2), Article e1011.

Zhao, J., Freeman, B., & Li, M. (2016). Can mobile phone apps influence people's health behavior change? An evidence review. *Journal of Medical Internet Research, 18*(11).